BAYESIAN STATISTICS FOR FISHERY STOCK ASSESSMENT AND MANAGEMENT

By

Paul G. Kinas

B. Sc. (Oceonology) Universidade do Rio Grande

M. Sc. (Statistics) Universidade de São Paulo

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF STATISTICS

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

1993

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Statistics

The University of British Columbia

2075 Wesbrook Place

Vancouver, Canada

V6T 1W5

Date: *12 October 1993*

Abstract

This work is about the use of Bayesian statistics in fishery stock assessment and management. Multidimensional posterior distributions replace classical parameter estimation in surplus-production and delay-difference models. The maximization of expected utilities replaces the estimation of optimal policies.

Adaptive importance sampling is used to obtain approximations for posterior distributions. The importance function is given as a finite mixture of heavy-tailed Student distributions. The performance of the method is tested in five case-studies, two of which use data simulation. Real data refer to Skeena river salmon (*Oncorhynchus nerka*), Orange Roughy (*Hoplostethus atlanticus*) and Pacific cod (*Gadus macrocephalus*).

The results show that the technique successfully approximates posterior distributions in higher dimensions even if such distributions are multimodal. When comparing models in terms of their performance as management tools, simpler and less realistic models can do better than more sophisticated alternatives. The Bayesian approach also sheds new light on the controversy about the Orange Roughy fishery.

Table of Contents

List of Tables

List of Figures

Acknowledgement

Chapter 1

Introduction

1.1 Motivation

This work is about the practical implementation of Bayesian statistical analysis for quantitative fisheries stock assessment and management. The basis of the study is the recognition that fisheries management is decision making in the presence of uncertainty. Different possible actions have to be compared according to measures of their consequences. Such analysis typically seeks a compromise between biological, economical and social components — see Section 1.2 — while traditional assessment concentrates on the biology of the stock. The Bayesian approach described in Section 1.3 expresses uncertainty as probabilities and uses utility theory to measure the consequences of different actions. Lindley [31] and Berger [2] give convincing arguments endorsing the claim that coherent and rational decision making in the presence of uncertainty must correspond to a Bayesian analysis. Hilborn and Walters [25] emphasize the importance of a reorientation of thinking within fishery sciences. I believe that the Bayesian paradigm to statistics has the ingredients to support such a philosophical change: (i) the consistent expression of uncertainties in the appropriate language of probability, and (ii) a coherent framework for decision making in the presence of uncertainties. Technical difficulties have hindered a wider use of the Bayesian approach in practical fishery problems. The *adaptive importance sampling* procedure described in Chapter 2 is designed to overcome some of these obstacles. In Chapter 3 I propose ways in which such procedure can be implemented for *biomass dynamic models*. I further illustrate the procedure for a variety of case-studies in Chapter 4 and make general comments and draw conclusions in Chapter 5.

In fisheries sciences it is assumed that management actions are taken upon recommendations

of scientific assessment of the resource and its potential production. But criteria for real-world decisions need to include considerations of a political, social, economic, and cultural nature as well. Quantitative stock assessment is too often seen as synonymous with management, while in fact assessment and management serve distinct purposes: the former should produce useful *advice* which can effectively *assist* decision making which is responsibility of the latter. The conventional approach of producing a set of statistical point estimates has proved to be misleading for at least two reasons: (*i*) it ignores uncertainties about models and estimation and (*ii*) it does not tell *how* the results are to be used. The scientists in charge of assessment were expected to recommend such quantities as total fishing effort and catch quotas. This procedure improperly placed the judgment about risk-taking into the hands of technical advisers. In contrast, when producing an answer in form of a (posterior) probability distribution, the Bayesian approach gives a complete description of current beliefs and uncertainties. It becomes the responsibility of the manager to formulate an utility function which measures the consequences (biological, economical, social) of different actions. The assessment expert is no longer the decision maker, but an analyst of uncertainty. A clear indication of the appeal of Bayesian statistics in fisheries, is the considerable growth of the use of such methods in current fisheries literature (Ludwig and Walters [36], Mangel and Clark [40], Clark [6] [7], Walters [68], Charles [5], Fried and Hilborn [19], Deriso and Parma [13], Nakamura et al. [43], Hilborn and Walters [25], Zeh et al. [80], Geiger and Koenings [20]).

Fundamental characteristics of fisheries sciences are the lack of reliable controls and difficulty to obtain replicates. The possibility of performing planned experiments is severely limited. Studies of the resource, using data produced during the history of its exploitation, are mostly observational. In contrast, statistical science originally dealt with the collection of data and later moved to the analysis of (replicated) data obtained under controlled conditions. Classical *frequentist* probabilities involve the concept of a long sequence of repetitions for a given situation (the probability of $\frac{1}{2}$ that a fair coin comes up heads builds on such rationale). This (frequentist) view of probability is central to classical statistics. In fishery sciences, however, where often

there are no replicates, the uncertainties about the effects of an action have a very different meaning. For example, consider the statement: "There is a probability of $\frac{1}{2}$ that the stock will collapse within five years if action **A** is taken today". The frequentist viewpoint does not apply here. The theory of *subjective* probability has been created to enable one to talk about probabilities in such situations. A comprehensive review of some systems describing subjective probabilities is found in Press [51]. According to Berger [2] (p.75), "such systems show that the complexity of probability theory is needed to quantify uncertainty (i.e., nothing simpler will do), and at the same time indicate that it is not necessary to go beyond the language of probability". I believe that the classical statistical approach, with its ideas of repetition and relative frequency is inadequate to cope with the reality of fishery sciences in most cases, and that the Bayesian paradigm, founded on the concept of subjective probability, has the potential to provide better support of scientific procedures in this field. Given the uninformative nature of most fishery data, the practitioner knows that all bits of additional information can be of great relevance and should therefore be used whenever possible. A *fair* way of proceeding is to explicitly unfold available beliefs and insights in the form of probability distributions. By formally stating such probabilities, uncertainties become open to public scrutiny and evaluation.

Starfield and Bleloch [65], by pointing out that "people's perception about models and modelling can be very different depending on the kind of problems they usually face and the types of tools they commonly use", provide a very general classification which can be helpful to summarize the difficulties typically faced by fisheries scientists. Their classification is according to two factors:

Factor 1 the level of understanding (and control) of the problem: low, high.

Factor 2 the quality and/or quantity of available data: low, high.

Most problems in engineering and physical sciences, for example, are close to a *high-high* classification for Factors 1 and 2 and consequently benefit most from the developments in classical statistics. Many problems in the nonphysical sciences, particularly fisheries, are closer to a *low-low* classification and are therefore unsuitable for a traditional approach.

Within this general picture of fishery science and given the argumentation above, the modelling approach proposed here has two central features:

1. The sequential nature of system dynamics and data acquisition.

2. The Bayesian paradigm to perform analysis and assist the decision making process in the presence of uncertainties.

The next section discusses the different levels of uncertainty encountered in fisheries and the tradeoff between information acquisition and monetary returns, when management actions are chosen. In Section 1.3 I describe the Bayesian approach. Finally, Section 1.4 gives a summary of the contents in the remaining chapters.

1.2 Uncertainties, Models and Learning

The necessity for explicit recognition of uncertainties was used in the previous section as a central argument in support to a reorientation in thinking among fishery scientists and managers. The Bayesian approach, suggested as an adequate framework in which to perform the statistical analysis, will be described in Section 1.3. Here I will comment on three types of uncertainty in fisheries models: (*i*) model structure, (*ii*) parameter estimation and (*iii*) prediction; and I will discuss how each one of them affects assessment and/or management.

Model Structure.

This basic uncertainty is central to modelling in general and is not a feature confined to problems in fisheries. It is concerned with the formulation of problems so that relevant properties are correctly captured. While there is no specific way to address this complex issue, a variety of books can provide helpful advice: Starfield and Bleloch [65], Walters [68], Clark [7], Hilborn and Walters [25].

Linhart and Zucchini [32] propose one possible strategy to confront this difficulty. They

introduce the concept of an *Operating Model* which should be the nearest possible representation of the "true situation". Ideally, it is on the operating model (or a family of alternative possibilities) that interpretation, prediction, and decision should be based. Since these models usually have many unknown parameters, considerations on the behaviour of estimates play an important limiting role in practice and will often require the use of simpler *Estimation Models*. This last consideration leads to the second level of uncertainty: the estimation of parameters.

Parameter Estimation

The lack of good controls, often caused by management constraints, result in weak contrast in fishery data (Ludwig at al.[35]). When combined with (often large) observation errors, such data tend to be non-informative. Parameter estimation under such circumstances becomes worse (increasingly uncertain) as the dimensionality of the problem grows, a phenomenon known to statisticians as "overfitting".

The uncertainties associated with parameter estimation are accuracy (bias and variance) and confounding. The estimation may be done by minimizing least squares or maximizing the likelihood. Using simulation studies, it has been shown that for typical fisheries data, simpler models produce better policy estimates than more realistic highly parameterized ones (Ludwig and Walters [37] [38]). For further applications of the operating model concept in fisheries see Fournier and Doonan [16], Ludwig et al. [39] and Fournier and Warburton [17].

A drawback in all these procedures is the use of a single best estimate upon which management recommendations are based. It is common that various other parameter combinations fit the data nearly as well. Different combinations, however, might imply different management actions. Such legitimate and important uncertainty simply vanishes from the problem when conclusions are given as point estimates.

Prediction

As already mentioned in Section 1.1, fishery science depends heavily on observations rather than controlled experiments. This limits the range of available data (because conventional management tends to seek stability) and affects the reliability of predictions. It is generally true that moving outside the range of observed data and controls is a very uncertain undertaking. For fishery models in particular, it usually involves extrapolation for non-linear systems. For example, one can only know how a fish stock responds to increased fishing pressure by trying it out. But in many instances this form of testing might translate into short-term losses (or unacceptable risks) for which there are no guaranties of future compensating returns. A compromise between short-term performance and long-term learning rates needs to be established. An action (control) which is good under performance considerations, perhaps by maximizing profit, is often incapable of producing useful information about the system (i.e., it results in non-informative data). This principle has become known as the "dual effect of control". Performance oriented strategies usually seek stability of the system, while better learning rates are associated with instability and strong contrast in the data. Actively adaptive (or "dual control") strategies are designed to establish some balance between learning and short-term performance. Walters [68] (Chapter 9) gives an extensive description of these methods, which involve stochastic dynamic programming. While acknowledging the importance of the described conflict and the relevance of actively adaptive strategies to management of fisheries, the present work will *not* further elaborate on the issue. But, as a final remark, I mention three features which are important for the development of actively adaptive strategies and are of particular interest here: (1) The identification of alternative hypothesis about stock dynamics, (2) the development of performance criteria for comparing adaptive policy options, and (3) the formal comparison of different options. In Section 1.3 it will become clear that the Bayesian approach can adequately deal with all three demands. In this sense, a support for active adaptive strategies is a support for the Bayesian approach as well.

In Section 1.1 I have pointed out the time-series nature of models and data acquisition in

fisheries. The adaptation of traditional time series methods (Box and Jenkins [3]) has been proposed in the past (Schnute [56], [58]) but has not proved to be successful in describing fisheries stock dynamics (Hilborn and Walters [25]). Some classical *autoregressive moving average model* ARMA(p,q) say, can be good to describe past data but is not of much help in guiding management actions for the future. The literature in time-series and dynamic system methodology (Brockwell and Davies [4], Aoki [1]) is mainly suited to situations in which a large amount of good quality data and a good understanding (and control) of the system dynamics is available.

Pella and Tomlinson [48] introduced a sequential fitting procedure which is unrelated to the traditional time-series methodology but has proved very useful in fisheries. They assumed a known underlying deterministic (but unobservable) model describing the biomass dynamic, together with an observation model establishing the link between observed catches and biomass. An estimate for the parameters of the biomass model was then obtained by minimizing least-squares of the differences between the observed and the predicted time-series of catches (observation errors). The operating model approach, described earlier in this section, is more general due to the inclusion of two features: (1) system noise (Schnute [58] discusses how the assumed noise structure can fundamentally affect the results of an assessment) and (2) more elaborated goodness-of-fit criteria allowing for the incorporation of auxiliary information. I shall give more details about this approach in Chapter 3 (Section 3.3) where I extend these concepts into Bayesian inference.

1.3 The Bayesian Method

In Sections 1.1 and 1.2, I presented a series of features of fishery science and argued in favour of a reorientation of thinking about the distinct role of fishery assessment and management. I further claimed that the Bayesian paradigm of statistical modelling is the correct framework in which to operate. Fishery management is ultimately a problem of decision making in the presence of uncertainties and the Bayesian theory consists of an axiomatic system developed to provide a rational and coherent way of reaching such decisions.

Here I shall describe the central ideas which characterize the Bayesian method. As introductory books to the subject I suggest Lindley [31] and Lee [29]. More technical treatments are Berger [2], Press [51] and, in the context of dynamic models, West and Harrison [77].

There are four basic elements:

1. The unknown quantity θ which denotes the uncertain element in the decision process and is called the *state of nature*. The set of all possible states of nature will be symbolized by Θ. For definiteness and clarity of exposition, let Θ be a subset of the real line $\Re$. A probability density $\pi(\theta)$ is assumed to describe the *subjective* relative credibility associated to a particular θ.

2. A sample space of elements x and a set of probabilities $p_o(x \mid \theta)$ of x, indexed by θ. The model $p_o(x \mid \theta)$ (the index 'o' indicating *observation*) is seen as a function of θ and, conditional on a fixed observation of x, is denoted the *likelihood function* and symbolized by $l(\theta)$. This "conditional view" — which makes the sample space (i.e., the set of all possible outcomes for x) irrelevant in most cases — is known as the *likelihood principle*. This principle makes it explicit that *only* the observed x should be relevant to provide evidence about θ.

3. A set of all possible actions under consideration. I denote an individual action by a and the set of all such actions by A.

4. A numerical value measuring the consequence (utility) of an action a if θ turns out to be the true state of nature. This value, symbolized by $U(\theta, a)$ and denoted *utility function*, is assumed to be defined for all pairs $(\theta, a) \in \Theta \times A$. Such an utility function is described on a scale which is also based on probabilities (this provides for coherence).

The probability density $\pi(\theta)$ describes uncertainties about θ *prior* to the observation x; therefore called the "prior density". In light of the information provided by the observed data x, the prior beliefs about θ are updated using *Bayes Theorem*, which says:

$$p(\theta \mid x) \propto p_o(x \mid \theta)\pi(\theta)$$

The constant of proportionality is a normalizing constant to make $p(\theta \mid x)$ a proper probability density so that it integrates to 1 over the set Θ.

From a Bayesian viewpoint, the conditional density $p(\theta \mid x)$ gives a complete description of θ *after* observation of the data x; it is denoted the "posterior density". Since $p(\theta \mid x)$ says it all, there is no need to go further in terms of estimation. Anything less, however, implies loss of information.

Any action should have its value described on an utility scale. In the light of current beliefs, fully expressed in the posterior $p(\theta \mid x)$, a best action must be chosen. It is an implication of Savage's [55] work that, given a coherent comparison of utilities, the best action is one that maximizes the expected (posterior) utility $E[U(\cdot, a)]$. In the continuous case and for a given a this expectation is obtained as the integral

$$E[U(\cdot, a)] = \int U(\theta, a)p(\theta \mid x)d\theta$$

Additional arguments in favour of the maximization of expected utility are given by Lindley [31] (p. 57).

Some remarks are needed here:

- For simplicity, I have assumed Θ to be a subset of the real line $\Re$ in the above description, and consequently $\pi(\theta)$ and $p(\theta \mid x)$ are one-dimensional probability densities. But, I can have Θ as finite or countable ($n \to \infty$) sets, in which case a discrete probability distribution is used for both, prior and posterior. Also, $\theta = (\theta_1, \cdots, \theta_p)'$ can be a p-dimensional vector so that Θ becomes a subset of $\Re^p$ and multivariate probability densities would apply. In fact, the latter (multivariate) will be the case of interest in the remaining chapters.

- Bayes' Theorem is central to the approach (hence, the name) in that in establishes a systematic way of handling (and modifying) subjective beliefs about θ in the presence of observation x.

- In Bayesian inference it is common to consider *loss functions* instead of utilities, so that minimization of expected losses replaces maximization of expected utilities in determining an optimal action. The "squared-error loss", which defines

$$L(\theta, a) = (\theta - a)^2$$

  is a typical example. If $A = \Theta = \Re$, it is easy to show that the posterior expectation $E[\theta]$ is the value for a which minimizes the posterior expected square-error loss $E[(\theta - a)^2]$.

  In general it is adequate to think simply of the relation $L(\theta, a) = -U(\theta, a)$; that is, think of loss as being a negative utility.

There would be little resistance to the use of prior distributions if θ were a random variable subjected to frequentist interpretation. The subjectivistic concept of probability and the use of prior distributions causes some discomfort even among advocates and users of Bayesian analysis for fisheries. To give a taste of the controversy involving the former, I quote Lindley [31](p.20):

> "Some have regarded this (subjective probability) as a disadvantage. In fact, the subjective nature of probability is its strength, for it describes a real situation of a subject, or observer, contemplating a world, rather than talking about a world divorced from the observer living in it, as do some other sciences."

As to the latter, Berger [2] (Sections 3.7 and 4.7) discusses extensively the common criticism associated to prior specifications and presents possible solutions to this dilemma. I give next a short account of his discussion regarding four arguments which are likely to be raised in any critical discussion about the merits (or not) of a Bayesian analysis.

1. *Lack of Objectivity:* If an overwhelming amount of (good quality) data were available, almost any analysis would yield similar conclusion, so that prior information would be largely irrelevant. But, when such data availability is limited (as in most cases), the choice of a model and its error structure can have a comparatively much stronger impact on the answer than the choice of a prior does; and it is just as subjective.

2. *Misuse of Prior*: It is true that the use of a prior introduces an additional source for abuse
   of statistics. The best prevention against potential misuse consists in a proper education
   regarding prior information and the handling of uncertainties. In reporting results of a
   Bayesian analysis, prior, likelihood, and utility should be reported separately so the reader
   can evaluate how reasonable all subjective inputs are. It also makes it possible to try out
   alternative priors if such a change is felt to be appropriate.

3. *Robustness*: "Do small changes in prior specification cause significant changes in the final
   decision ?" This question is justified and a central issue. Studies in Bayesian robustness
   and related methodology address the topic. The construction of robust priors is sometimes
   an adequate compromise. If, however, different reasonable priors result in contradicting
   conclusions, this is an indication of scientific uncertainty of the problem rather than a
   weakness of prior specification.

4. *Data or Model Dependence of the Prior*: The Bayesian approach assumes that the infor-
   mation contained in $\pi(\theta)$ does not depend in any way on the data. This idealized view is
   hard to match in practice. This happens because a model $p_o(x \mid \theta)$ is often chosen *after*
   inspection of available data. Such choice further defines the parameters θ under consider-
   ation; hence, the circularity between data and parameters. The answer to these criticism
   is twofold. First, all analyses suffer from the circularity between data and model. Second,
   if in a given analysis the model seems reasonable as the priors do, and other reasonable
   priors give similar conclusions, then details of prior development are not central to the
   analysis; this leads back to the robustness considerations discussed previously.

The Bayesian paradigm is conceptually easy to implement. But the calculation of the
posterior $p(\theta \mid x)$ and the corresponding expected utility entail evaluation of integrals which
can become difficult, especially for multidimensional θ. To avoid such difficulties, the use of
conjugate priors has been an important concept and is explained below.

Conjugate Families and Numerical Approximations

If a class F of density functions $p_o(x \mid \theta)$ is given, and P is a restricted class of prior distributions $\pi(\theta)$, then we say that "the class P is a conjugate family for F", *if* the posterior $p(\theta \mid x)$ is again a member of P. Some examples:

1. If $(x \mid \theta)$ has Gaussian distribution with mean θ and (known) variance σ^2, then the class of Gaussian priors $\pi(\theta)$ is a conjugate family for the class of Gaussian (sample) densities.

2. If $(x \mid \theta)$ has Gaussian distribution with (known) mean θ and variance σ^2, then the class of Inverse-Gamma priors $\pi(\sigma^2)$ is a conjugate family for the class of Gaussian (sample) densities; that is, $p(\sigma^2 \mid x)$ is an Inverse-Gamma density.

3. If $(x \mid \theta)$ has Poisson distribution with intensity parameter θ (i.e. $h(x \mid \theta) = \theta^x e^{-\theta}(x!)^{-1}$), then the class of Gamma priors $\pi(\theta)$ is a conjugate family for the class of Poisson (sample) densities.

The need to look for conjugate families is a constraint on practical application. For fisheries in particular, the non-linearity of models further complicates the matter. In an attempt to free applications from both difficulties, theoretical and technological developments over the last decade have produced a variety of computationally demanding approximating methods. Smith [60] in a recent review, divides these techniques according to three different strategies:

(i) Laplace and related analytical approximations,

(ii) adaptive quadrature based on classical numerical analysis and

(iii) sample-based methods or Monte Carlo integration.

While (i) and (ii) require a great deal of expertise and become harder with growing dimensionality of θ, the sample-based methods of (iii) are suitable to a broader audience for their relative ease of implementation and practicality for high-dimensional θ.

The final output of a sample-based method consists of a random sample $\Theta = \{\theta_1, \cdots, \theta_n\}$ drawn from the posterior distribution $p(\theta \mid x)$. The essential duality between a sample and the distribution from which it is drawn is an important concept which I shall explore next.

The distribution clearly generates the sample. But, conversely, given a sample one can approximately recreate the distribution. More generally, one can use the sample Θ to explore and summarize features of $p(\theta \mid x)$ as one would proceed in exploratory data analysis on ordinary data x. Features which might be of interest are (Smith and Roberts [62]):

- examining shapes of uni-variate marginals for θ or for functions $H(\theta)$.

- producing marginal moments or quantile summaries.

- examining bivariate marginals or pairs of functions.

- examining trivariate distributions (by taking advantage of the powerful statistical packages for graphical analysis).

- connecting Θ to an utility function so that the output gives answers to a specific decision problem.

- using Θ in predictive analysis.

The Monte Carlo methods are further sub-divided into two groups. First there are the *Markov Chain* methods which include the Gibbs sampler of Geman and Geman [22]. The second group is based upon *importance sampling*. The present work uses the latter. In the next two chapters I shall explore possibilities for the use of *adaptive importance sampling* with *mixture models* (West [75]) to obtain posterior distributions for parameters in a variety of fishery models.

1.4 Summary of Contents

In Chapter 1 I have described the motivations for this work and the relevance of the Bayesian paradigm in fishery sciences (Section 1.1). Different levels of uncertainty and their impact in fishery assessment and management were examined in Section 1.2. The Bayesian method was described in Section 1.3. This last section is designed to give an overview of the contents of remaining chapters and assist a selective reader.

In Chapter 2 I describe the Monte Carlo methods used in Bayesian computation of posterior probability distributions. Central aspects of basic importance sampling are reviewed in Section 2.2 and the adaptive importance sampling proposed by West [75] is reviewed and illustrated in Section 2.3. The use of the approach for sequential analysis of dynamic models is described and illustrated in Section 2.4.

In Chapter 3, I describe a variety of models for biomass dynamics: stock-recruitment, surplus-production and delay-difference. Starting from the basic time-series fitting approach of Pella and Tomlinson [48], I incorporate the adaptive importance sampling methodology to construct an approximate posterior distribution for the underlying parameter vector. The derivation of the results are in Sections 3.2 and 3.3. Section 3.4 gives an overview of delay-difference models.

In Chapter 4 I present different case-studies to illustrate the results of Chapters 2 and 3. In Section 4.2 I construct a bivariate posterior distribution for the stock-recruitment parameters of Skeena river sockeye salmon data. A simulation model using catch and effort data is described in Section 4.3. This case also proposes an utility function and uses it to make management decisions. In Section 4.4 I examine data for the Orange Roughy fisheries in New Zealand. Different management strategies are analyzed in the context of Bayesian decision theory and the results compared to the (non-Bayesian) risk analysis of Francis [18]. The last two case-studies refer to situations for which the posterior distributions are bi-modal. In Section 4.5 I use artificial data generated from simulations and in Section 4.6 I use data from the Pacific cod fisheries off the coast of British Columbia, Canada.

The insight gained from the work, topics for further research, and final conclusions are all given in Chapter 5.

Chapter 2

Bayesian Posterior Computation

2.1 Introduction

In the previous chapter I discussed general features of fisheries stock assessment and management. I gave arguments supporting the adequacy of the Bayesian paradigm to perform statistical analysis and to guide decision making. In Section 1.3 the Bayesian method was described. While conceptually easy to implement, its application reduces mostly to calculations of integrals of the form

$$E[H(\cdot)] = \frac{\int H(\theta)f(\theta)d\theta}{\int f(\theta)d\theta} \tag{2.1}$$

where $\theta \in \Re^p$, $H(\theta)$ is a measurable function, and $f(\theta)/\int f(\theta')d\theta'$ a probability density. Different choices of $H(\theta)$ result in different aspects of practical Bayesian inference. To focus attention, assume that $f(\theta)$ is proportional to the posterior density of interest. Then,

- To calculate the posterior mean μ, I define $H(\theta) = \theta$. Further, taking $H(\theta) = (\theta - \mu)^2$ yields the corresponding posterior variance.

- If a decision has to be made between actions a_1 and a_2 according to the expectation of given utility functions $U(\theta, a_1)$ and $U(\theta, a_2)$, I define $H(\theta) = U(\theta, a_1) - U(\theta, a_2)$.

- To calculate posterior probabilities of specific sets (or hypothesis), $H(\theta)$ can be defined as an indicator function I_A for the appropriate set A. As an example, suppose I want to estimate the posterior probability of θ being smaller than a fixed value θ_0; this is obtained by defining

$$H(\theta) = I_{\theta < \theta_0}$$

15

- Within the context of fishery models in particular, the function $H(\theta)$ sometimes defines a policy (e.g. optimum fishing effort) for which an estimate is desired. The posterior expectation μ_H say, can serve this purpose (although posterior largest mode and median are also possibilities). Posterior estimation of the corresponding standard error is obtained by defining a new function $H^*(\theta) = (H(\theta) - \mu_H)^2$.

Models in fisheries are mostly non-linear in θ, which is often a vector of moderate to high dimension. As a result, the evaluation of (2.1) is in general quite complex and impractical to be solved analytically. A wide range of approximating methods have been developed to overcome this difficulty. These methods can be subdivided into two major groups. One group concentrates on *analytic approximations* involving Taylor series expansions (Lindley [30], Tierney and Kadane [67], Delampady et al. [11] [10]). Difficulties with the implementation of the analytic methods grow rapidly as the dimensionality and complexity of the models increase. The second group of approximating methods, involving *Monte Carlo integration*, is generally necessary under those circumstances. These methods are presently an active field of research; see Smith [60] for an overview.

The principle behind Monte Carlo integration is as follows. Suppose one can generate an independent identically distributed (i.i.d.) sequence of random variables $\{\theta_1, \cdots, \theta_n\}$ having common density $p(\theta) > 0$ on Θ. Then, for a well-behaved $H(\theta)$, it follows from the strong law of large numbers, that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(\theta_i) = \int_{\Theta} H(\theta)p(\theta)d\theta$$

While the principle is simple, difficulties for practical implementation consist in the ability to perform random drawings from a general density $p(\theta)$; it might be computationally very expensive (as the dimension of θ increases) or impossible for complicated forms of $p(\theta)$. In attempting to overcome these hurdles, Monte Carlo methods have followed two distinct paths:

1. *Markov Chain Methods:* These methods include the Gibbs sampler algorithm of Geman and Geman (1984) [22]. Given a p-dimensional θ, the key feature of the algorithm is to

fix the full conditional densities $p(\theta_i \mid \theta_j; \ j \neq i)$ for $i = 1, \cdots, p$, from which computationally inexpensive random draws are made. Possible limitations of the methods are twofold: (1) the need to specify full conditionals (often a difficult task in itself) within a convenient class of distributions from which sampling is easy and (2) the possibility of bad performance for distributions featuring two separate modes (with a region of small densities in between). In the latter case, and due to its Markovian nature, a chain of random generation could be "trapped" at one of the modes. Refer to Smith and Roberts [62] for a review. A related method is the substitution sampler of Gelfand and Smith [21].

2. *Importance Sampling Methods*: In its original form, as described in Geweke [23], basic importance sampling replaces the density $p(\theta)$ by a "similar" density $g(\theta)$, from which the random draws can be made efficiently. The ability to find a suitable $g(\cdot)$ is the key assumption. Adaptive sampling schemes have been developed to assist the choice of an *importance function* $g(\cdot)$ (Kloeck and van Dijk [28], Naylor and Smith [45], Oh and Berger [46]). An important variation, specially useful for high-dimensional θ, was proposed by West [75] [76]. It consists in modelling $g(\cdot)$ as a discrete mixture of distributions of some standard parametric form. By recursively updating $g(\cdot)$ a few times, the final *mixture model* seems capable of providing a satisfactory approximation to $p(\cdot)$ quite generally.

Sample-based Monte Carlo methods are becoming a central tool for practical implementation of Bayesian inference. In this work I shall focus attention on adaptive importance sampling methods with the importance function given as a finite mixture of distributions. I shall explore possibilities for its use in a variety of biomass dynamic models in the upcoming chapters. Before examining these, I use the remainder of the present chapter to describe importance sampling.

The next section introduces basic importance sampling. In Section 2.3 I present the algorithm for adaptive importance sampling using mixture models and provide an illustrative example. In Section 2.4 I describe the use of the adaptive algorithm for sequential updating of dynamic models and illustrate the procedure with another example.

2.2 Monte Carlo Integration with Basic Importance Sampling (BIS)

Let θ have a density $p(\theta)$ with support $\Theta \in \Re^p$. Assume that it is impractical to produce a sequence of i.i.d. random variables from $p(\cdot)$; but the exact evaluation of the unnormalized non-negative function $f(\theta) \propto p(\theta)$ is possible for a given θ. Suppose it is required to evaluate the expectation (with respect to $p(\cdot)$), of some given function $H(\theta)$. Monte Carlo integration with importance sampling performs this calculation as follows:

1. Specify a probability density $g(\theta)$ as an approximation for $p(\theta)$. If $E^p[\cdot]$ and $E^g[\cdot]$ denote respectively the expectations under densities $p(\cdot)$ and $g(\cdot)$, then

$$E^p[H(\cdot)] = \frac{\int H(\theta)f(\theta)d\theta}{\int f(\theta)d\theta}$$

$$= \frac{\int H(\theta)\frac{f(\theta)}{g(\theta)}g(\theta)d\theta}{\int \frac{f(\theta)}{g(\theta)}g(\theta)d\theta}$$

$$= \frac{\int H(\theta)\omega(\theta)g(\theta)d\theta}{\int \omega(\theta)g(\theta)d\theta}$$

$$= \frac{E^g[H(\cdot)\omega(\cdot)]}{E^g[\omega(\cdot)]}$$

where $\omega(\theta) = \frac{f(\theta)}{g(\theta)}$.

2. Draw a sample of size n from $g(\theta)$, denoted $\Theta = \{\theta_1, \cdots, \theta_n\}$.

3. Get the set of normalized weights $\Omega = \{\omega_1, \cdots, \omega_n\}$, where ω_j is defined as

$$\omega_j = \frac{\omega(\theta_j)}{k} = \frac{f(\theta_j)}{kg(\theta_j)}$$

with $k = \sum_{j=1}^n \omega(\theta_j)$.

4. Use Monte Carlo approximation to evaluate the expectations with respect to g. Namely,

$$E^g[H(\cdot)\omega(\cdot)] \approx \frac{1}{n}\sum_{j=1}^n H(\theta_j)\omega(\theta_j)$$

$$E^g[\omega(\cdot)] \approx \frac{1}{n} \sum_{j=1}^{n} \omega(\theta_j) = \frac{k}{n}$$

The final result is

$$E^p[H(\cdot)] \approx \sum_{j=1}^{n} H(\theta_j)\omega_j \qquad (2.2)$$

In essence, the procedure consists in replacing $p(\cdot)$ by a discrete approximation with mass function ω_j at θ_j

The description above assumes the existence of a good importance function $g(\cdot)$. Hence, the key issue is *finding* a suitable $g(\cdot)$. It should be easy to draw i.i.d. random variables from it while, at the same time, it should make the approximation in (2.2) accurate for n as small as possible. These are sometimes competing goals, and different situations might need different solutions. Berger [2] (pp.263-265) provides some guidelines for choosing an adequate $g(\cdot)$. Geweke [23] gives mild conditions under which the numerical approximation in (2.2) converges almost surely to the true value and is approximately normal distributed with mean $\mu_H = E^p[H(\cdot)]$ and variance σ^2/n, as the number of Monte Carlo replications n increases. The variance is defined as

$$\sigma^2 \propto E^p\left[(H(\cdot) - \mu_H)^2 \omega(\cdot)\right].$$

The performance of BIS clearly depends on σ^2 and n, so that a $g(\cdot)$ yielding small σ^2 is desirable. Although there is not a general rule for choosing a good $g(\cdot)$, Oh and Berger [46] give some desirable properties of a good importance function:

1. It should be easy to generate random drawings from it, a requirement already mentioned earlier.

2. It should not have sharper tails than $f(\cdot)$. The intuition being that sharper tails would result in large values of $\omega(\theta_j)$ for outlying values of θ_j. The result is increase in the variance σ^2 and instability in the estimates.

3. It should mimic the true $f(\cdot)$. Here the goal is to avoid having a large proportion of $\omega(\theta_j)$ near zero. Such values have little effect on (2.2), reducing the efficiency of the method.

A complementary way of reducing the variance of the importance sampling estimate has been the search for appropriate sampling strategies (Rubin [54]). According to Smith [60], and provided that a good procedure can be found, it is likely that importance sampling is more efficient than the Markov approach mentioned earlier.

Probability Densities and Random Samples

An important aspect of all the Monte Carlo sampling methods, and importance sampling in particular, is the duality between the density $p(\cdot)$ and a sample $\Theta = \{\theta_1, \cdots, \theta_n\}$ drawn from it. If only Θ is known, it is possible to approximately reconstruct the underlying density $p(\cdot)$ using an histogram, kernel techniques, or the empirical cumulative distribution. Switching attention to the sample Θ as a description of the posterior leads to two important observations.

First concerns Bayesian Decision Theory. There it is usually necessary to evaluate the expected utility $E^p[U(\theta, a)]$ to find the optimal action a. If Θ is an i.i.d sample from $p(\cdot)$, it seems natural to use the same set Θ (and $\Omega = \{\omega_1, \cdots, \omega_n\}$) to calculate the expected utility for all required values of a. Since the generation of Θ is often the most expensive part of the Monte Carlo process (Berger [2]), the availability of Θ and Ω is an important advantage of the sampling methods with respect to computational efficiency.

The second observation relates to resampling techniques and can be very helpful in the context of Bayesian robustness. There it is desirable to determine potential changes to the posterior when alternatives to the likelihood and/or the prior are considered. In the context of *distributions* such a study can pose formidable difficulties and requires a fresh start when a new possibility has to be investigated. Focusing attention on the available *sample* Θ, the resampling approach makes such an analysis remarkably simple (Smith [60]). Starting with the sample Θ from $p(\theta)$, suppose the goal is to produce a new sample Θ^* from a distribution proportional to some known function $f^*(\theta)$. Resample from $\Theta = \{\theta_j; j = 1, \cdots, n\}$ as follows:

1. Calculate $\omega_j = f^*(\theta_j)/p(\theta_j)$.

2. Obtain the normalized weights

$$q_j = \frac{\omega_j}{\sum_{j=1}^n \omega_j}$$

3. Draw a random sample with replacement from the discrete distribution over Θ placing mass q_j on θ_j. Denote this sample $\Theta^* = \{\theta_1^*, \theta_2^*, \cdots\}$. Then Θ^* is approximately distributed according to the density

$$f^*(\theta) / \int f^*(\theta') d\theta'$$

Here is an example. Let $\pi_1(\theta)$ and $\pi_2(\theta)$ be distinct priors to be compared and $l_1(\theta)$, $l_2(\theta)$ two likelihoods. For the posterior densities $p_i(\theta) \propto l_i(\theta)\pi_i(\theta)$ for $i = 1, 2$ the following relation holds:

$$p_2(\theta) \propto \left(\frac{l_2(\theta)\pi_2(\theta)}{l_1(\theta)\pi_1(\theta)}\right) p_1(\theta)$$

If Θ is a sample from p_1, changes from the likelihood $l_1(\theta)$ to $l_2(\theta)$ and/or from the prior $\pi_1(\theta)$ to $\pi_2(\theta)$ can be examined using the resampling algorithm described in steps 1 to 3 above. Simply define

$$v(\theta) = \frac{l_2(\theta)\pi_2(\theta)}{l_1(\theta)\pi_1(\theta)}$$

$$p(\theta) = p_1(\theta)$$

$$f^*(\theta) = v(\theta)p_1(\theta)$$

$$\omega_j = v(\theta_j)$$

2.3 Adaptive Importance Sampling (AIS) and Mixture Models

For efficiency in using the importance sampling method, we must use an appropriate importance function $g(\cdot)$ (in the sense discussed in the previous section). Finding such a function requires skills. It becomes harder as the dimension of θ increases. For wider applicability of the method, more automated ways to obtain a satisfactory $g(\cdot)$ are needed. This is the motivation behind the development of adaptive importance sampling (AIS), the subject of this section.

The AIS procedure was introduced by Kloeck and van Dijk [28]. It starts with some crude approximation for $p(\cdot)$ denoted $g_0(\cdot)$. Because many of the sample points $\theta_{0,j}$ will have negligible weights $\omega_{0,j}$ (i.e. small mass $p(\theta_{0,j})$ relative to $g_0(\theta_{0,j})$), such an importance function is very inefficient computationally, requiring huge sample size n_0, as well as statistically, converging slowly or not at all. The adaptive refinement consists in taking a small n_0 and, based on the outcome, constructing the updated $g_1(\cdot)$. A new sample from $g_1(\cdot)$ is expected to have a better match with $p(\cdot)$. The final importance function obtained by such an adaptive scheme is expected to display good mimicry to $p(\cdot)$; a desirable property according to the previous section. Different methods to perform the adaptive refinement of $g(\cdot)$ were proposed by Naylor and Smith [45], Oh and Berger [46] and West [75] [76].

I will focus on West's method: the use of a discrete mixture of distributions of some standard parametric form. Three features which justify my choice for this method are: (1) the capability of *mixture models* to emulate very irregular forms of distributions (Wolpert [79]), (2) the practicality for high-dimensional θ, and (3) the potential for automation.I use the remainder of this section to reviewing this method.

Assume θ is a p-dimensional vector with true density $p(\theta)$. Further assume that $f(\theta) \propto p(\theta)$ can be evaluated for any given θ. The adaptive procedure for obtaining a good $g(\theta)$ goes as follows:

1. Define an initial importance function $g_0(\cdot)$ and draw a stochastic sample of size n_0 from it. Denote this sample by $\Theta_0 = \{\theta_{0,j}; j = 1, \cdots, n_0\}$ and the associated weights by $\Omega_0 = \{\omega_{0,j}; j = 1, \cdots, n_0\}$. Let

$$\Gamma_0 = \{g_0(\cdot), n_0, \Theta_0, \Omega_0\}$$

represent the available quadruple after this initial step.

2. Compute the Monte Carlo estimate for the variance V_0, as the weighted sum

$$V_0 = \sum_{j=1}^{n_0} \omega_{0,j}(\theta_{0,j} - \bar{\theta}_0)^2$$

where $\bar{\theta}_0 = \sum_{j=1}^{n_0} \omega_{0,j}\theta_{0,j}$ is the Monte Carlo mean.

3. Construct a revised importance function as the following weighted mixture:

$$g_1(\theta) = \sum_{j=1}^{n_0} w_{0,j}\, d(\theta \mid \theta_{0,j}, h^2 V_0) \qquad (2.3)$$

where $d(\theta \mid m, M)$ denotes a known p-variate density function with mean m and variance M. The *bandwidth* h is a smoothing parameter as used in conventional kernel density estimation (Silverman [59]). Refer to the comments below for more details about both $d(\cdot)$ and h.

4. Draw a larger sample $n_1 > n_0$ from $g_1(\cdot)$ and replace Γ_0 by

$$\Gamma_1 = \{g_1(\cdot), n_1, \Theta_1, \Omega_1\}$$

5. Use diagnostics (see comments below) and either stop and base inference on Γ_1 or repeat steps 2 and 3 for a revised version Γ_2.

Next a few comments to clarify some aspects of this adaptive procedure:

Smoothing: The mixture model $g(\cdot)$ given in (2.3) is simply a weighted kernel estimate with kernel $d(\cdot \mid m, M)$. The theory for kernel estimation (Silverman [59]) implies that the mixture model $g(\cdot)$ approaches $p(\cdot)$ for increasing sample size n if the bandwidth decays to zero at an appropriate rate. For Gaussian kernel, a standard h — defined as a function of the sample size n, the dimension p of θ and two constants a and b — is

$$h = \frac{a}{n^{1/b}}$$

with a given by

$$a = \left(\frac{4}{1 + 2p}\right)^{1/b}$$

Silverman [59] defines $b = p+4$. West [76] defines $b = 1+4p$, instead. For one-dimensional θ's, both bandwidths are identical. If $p > 1$ (and n is fixed), West's formula gives

comparatively larger values. If multimodality is suspected, this standard bandwidth tends to oversmooth and slightly smaller values, say rh $(0.5 < r < 1)$ are preferable. West [76] also shows that, for fixed n, the function $g(\cdot)$ is always overdispersed: a desirable feature for any importance function. The constant variance V in the kernel of (2.3) can be replaced by its Monte Carlo estimate obtained from the current values of Θ and Ω. Furthermore, in any adaptive procedure, a crude $g_0(\cdot)$ need only capture gross features of $p(\cdot)$ and a small sample size n_0 will do.

Possible variations (not used here) consist in: (1) allowing for a different bandwidth at each recursion, and (2) using local variances V_j for each kernel in the mixture. In (1) the spread of $g(\cdot)$ is controlled directly: it can be increased by intervention whenever the current importance function is questionable. The use of local variances suggested in (2), can better explore the existence of local dependencies among elements of θ. Although this is a good feature, the possible gain in statistical efficiency needs to be balanced against a corresponding loss in computational tractability.

The kernel: In all applications to be considered here, I shall use a heavy-tailed p-dimensional Student distribution as the kernel $d(\cdot \mid m, M)$. According to West [76], the Student distribution is an adequate choice, provided $\theta_i \in (-\infty, +\infty)$ for $i = 1, \cdots, p$. I use the notation $T(k, m, C)$ to denote a Student distribution with k degrees of freedom, mean m and scale C[1]

Diagnostics: After each recursion in the adaptive reconstruction of (2.3), different diagnostics can be used to check for improvement in the importance function. Particularly, West refers to four possibilities:

1. Changes in (Monte Carlo) estimated expectations.

2. Configuration of points in each dimension of the parameter space Θ.

3. Frequency distribution of weights Ω.

---

[1] The scale C and variance M of a Student distribution are related by $C = \frac{k-2}{k} M$ for $k > 2$.

4. Some measure of variability of weights.

In the approach taken here, I shall use a measure of variability of weights as the standard diagnostic tool. West [76] suggests the *entropy relative to uniformity*, which he defines as

$$\frac{-\sum_{j=i}^{n} \omega_j \log \omega_j}{\log n} \tag{2.4}$$

This value is non-negative and equals unity in case of a perfect match between $g(\theta_j)$ and $p(\theta_j)$ for all $\theta_j \in \Theta$; that is, the normalized weights ω_j will all equal $1/n$. As the variability among ω_j's increases, the relative entropy becomes smaller.

The choice of adequate sample sizes n_i, bandwidth h, and number of updates for $g(\cdot)$, can be guided by one or more of these diagnostics. For instance: look for stable Monte Carlo estimates of mean and variance (criteria 1); verify if the relative entropy (2.4) increases uniformly towards 1 (criteria 4), etc.

Reduction of Mixtures

For a very large n, the mixture (2.3) might contain many redundancies, so that another mixture with a much smaller n is alike for all practical purposes (see West [76] for examples). These smaller mixture models are obtained by reducing the n-component sets $(\Theta^{(n)}, \Omega^{(n)})$ to the smaller pair $(\Theta^{(m)}, \Omega^{(m)})$ with $m < n$ (superscripts indicating cardinality of the sets) by some clustering scheme. The procedure he suggests goes as follows:

1. Sort the elements $\theta_j \in \Theta^{(n)}$ according to increasing weights $\omega_j \in \Omega^{(n)}$, such that θ_1 corresponds to the smallest weight ω_1.

2. Find the index i, such that θ_i is the "nearest neighbour" to θ_1 and calculate

$$\omega^* = \omega_1 + \omega_i$$

$$\theta^* = \frac{\omega_1 \theta_1 + \omega_i \theta_i}{\omega^*}$$

3. Remove θ_1 and θ_i from $\Theta^{(n)}$ and include θ^* instead, to get $\Theta^{(n-1)}$. Proceed similarly with ω_1, ω_i and ω^* to get $\Omega^{(n-1)}$.

4. Repeat the previous steps $(n - m)$ times, until obtaining $\Theta^{(m)}$ and $\Omega^{(m)}$.

The reduced mixture keeps the form (2.3) but has only m components, with locations $\Theta^{(m)}$ and weights $\Omega^{(m)}$. The estimate for V, needed in the kernel $d(\cdot \mid m, M)$ and calculated from the larger sets $\Theta^{(n)}$ and $\Omega^{(n)}$, is kept. But a larger bandwidth h, due to the smaller number of kernels in the "reduced" mixture, is appropriate.

To help practical implementation of AIS for the biomass models of Chapter 3, I have included a pseudo-code for this procedure as an Appendix.

2.3.1 An Example

To illustrate the AIS procedure, I shall perform a Bayesian analysis for a model first considered by McCullagh and Nelder [41] (sec. 9.3.3) and, in the Bayesian context, by Smith and Gelfand [61].

For $i = 1, 2, 3$, let $X_{i,1}$ and $X_{i,2}$ denote two binomial random variables with parameters $(n_{i,1}, p_1)$ and $(n_{i,2}, p_2)$. Assume $X_{i,1}$ and $X_{i,2}$ are conditionally independent given (p_1, p_2), and with $n_{i,1}$ and $n_{i,2}$ specified. The observed random variables are $Y_i = X_{i,1} + X_{i,2}$.

Given the observations $Y_1 = y_1$, $Y_2 = y_2$, and $Y_3 = y_3$, the likelihood for (p_1, p_2) takes the form:

$$\prod_{i=1}^{3} \sum_{j \in A(i)} \binom{n_{i,1}}{j} p_1^j (1 - p_1)^{n_{i,1}-j} \binom{n_{i,2}}{y_i - j} p_2^{y_i-j} (1 - p_2)^{n_{i,2}-y_i+j} \qquad (2.5)$$

where $A(i) = \{j; \ \max\{0, y_i - n_{i,2}\} \leq j \leq \ \min\{n_{i,1}, y_i\}\}$.

The data of McCullagh and Nelder are given in Table 2.1. Their goal was point estimation of (p_1, p_2). I shall obtain a (two-dimensional) posterior density instead.

Transforming for θ: The method requires $\theta = (\theta^{(1)}), \theta^{(2)})$ to be defined on the real plane $\Re^2$. Hence I use the *logit* transformation

$$\theta^{(i)} = \ln(p_i/(1 - p_i)), \text{ for } i = 1, 2$$

Table 2.1: Data from McCullagh and Nelder [41] (Sec.9.3.3)

i	$n_{i,1}$	$n_{i,2}$	y_i
1	5	5	7
2	6	4	5
3	4	6	6

The final results can be described for $\theta^{(i)}$'s or, as I did here, transformed back to the original parameters p_i.

The calculations: I shall guide the reader through the various steps of the AIS algorithm for this specific example.

1. Define the prior distribution $\pi(\theta)$.

   After drawing a sample of independent observations from a two-dimensional uniform distribution on the unit square $(0,1) \times (0,1)$, and applying the logit transformation to each of them, the resulting sample of values θ has a dome-shaped distribution with (approximate) mean-vector $(0,0)'$ and diagonal covariance matrix diag$(3.2, 3.2)$. Therefore I used a two-dimensional Student with 30 degrees of freedom and the given values as mean and scale as my prior for θ. The degrees of freedom were chosen arbitrarily; smaller values would result in too heavy tails inconsistent with the observed sample.

2. Define $g_0(\cdot) = \pi(\cdot)$. That is, use the prior distribution as the initial importance function.

3. Draw a sample of size $n_0 = 1500$ from $g_0(\cdot)$. This is the set

$$\Theta_0 = \{\theta_{0,j}, \ j = 1, \cdots, n_0\}.$$

   The sample size, again an arbitrary choice based on diagnostics and experimentation, is of moderate size. It is intended to capture crude features of the true posterior. In subsequent recursions the importance function becomes more refined. Therefore the sample size will gradually increase.

4. For each $\theta_{0,j} \in \Theta_0$ calculate:

   (*i*) the true posterior density $p(\theta_{0,j}) = l(\theta_{0,j})\pi(\theta_{0,j})$. Where $l(\theta)$ is the likelihood (2.5) with p_i's ($i = 1, 2$) replaced by the appropriate function in $\theta^{(i)} = \theta_{0,j}^{(i)}$

   $$\frac{e^{\theta^{(i)}}}{(1 + e^{\theta^{(i)}})}$$

   (*ii*) the value $g_0(\theta_{0,j})$. In subsequent recursions the importance function $g(\cdot)$ from which the current sample was drawn is used instead.

   (*iii*) the weight $\omega(\theta_{0,j}) = p(\theta_{0,j})/g_0(\theta_{0,j})$.

5. Get the set of normalized weights

   $$\Omega_0 = \{w_{0,j}, \; j = 1, \cdots, n_0\}$$

   where $\omega_{0,j} = \omega(\theta_{0,j})/K$ and $K = \sum_{j=1}^{n_0} \omega(\theta_{0,j})$.

6. Using Θ_0 and Ω_0 calculate the Monte Carlo estimates of the (2×2) covariance matrix V, as indicated earlier. Also calculate the relative entropy for $g_0(\cdot)$ by using Ω_0 in equation (2.4).

7. Using the procedure for reduction of mixtures, described earlier in this section, reduce Θ_0 and Ω_0 to smaller sets of $m = 500$ points. Denote these new sets $\Theta_0^{(500)}$ and $\Omega_0^{(500)}$ and use them, together with the estimate V calculated in step 6, to construct the mixture model $g_1(\cdot)$ of the form (2.3). I fixed the kernel $d(\cdot \mid m, M)$ as a Student distribution with 9 degrees of freedom (to provide for heavy tails in the importance function) and the bandwidth of Silverman [59] defined earlier.

8. Draw a new sample of size $n_1 = 2500$ from $g_1(\cdot)$ and denote it

   $$\Theta_1 = \{\theta_{1,j}, \; j = 1, \cdots, n_1\}$$

9. Repeat steps 4, 5, and 6 to obtain Ω_1, V, and the relative entropy for $g_1(\cdot)$.

10. Obtain the reduced sets $\Theta_1^{(500)}$ and $\Omega_1^{(500)}$. Construct the new mixture model $g_2(\cdot)$.

11. Draw a new sample of size $n_2 = 4000$ from $g_2(\cdot)$ and denote it

$$\Theta_2 = \{\theta_{2,j}, \ j = 1, \cdots, n_2\}$$

12. Repeat steps 4, 5 and 6 to obtain Ω_2, V and the relative entropy for $g_2(\cdot)$.

13. Obtain the reduced sets $\Theta_2^{(500)}$ and $\Omega_2^{(500)}$. Use them in the final mixture model (2.3) to generate the contour plot of Figure 2.1.

Results: The relative entropy (2.4) was 0.873 for $g_0(\cdot)$, 0.993 for $g_1(\cdot)$ and 0.997 for $g_2(\cdot)$. The major improvement is shown for the first recursion, where the entropy jumps from 0.87 to 0.99. A final entropy close to 1 means good agreement between the mixture model and the true posterior at the 4000 observations in the set Θ_2. Figure 2.1 is a contour[2] plot based on the mixture approximation for the posterior. Negative correlation and a slightly larger value for p_2, as compared to p_1, are identified. For comparison, I included the values reported by McCullagh and Nelder: the quasi-likelihood estimate (0.36,0.84) (triangle) and the likelihood estimate (0.2,1.0) (cross). Both estimates satisfy $p_1 + p_2 = 1.2$, and are on the (dotted) line which represents all such cases. I have included this line to show that the shape of the contour plot is approximately symmetric to it.

Marginal posterior probabilities for p_1 and p_2 are shown in Figure 2.2(a). Despite the large uncertainties, indicated by the flatness of the distributions, there is a clear trend towards smaller values of p_1 (full line) as compared to p_2 (dotted line). When comparing both parameters, a natural function of interest is

$$H(p_1, p_2) = p_2 - p_1$$

I calculated this difference for each of the 4000 sample points in Θ_2. A summary of the outcome is shown as the full line in Figure 2.2(b). This curve was obtained using standard density estimation (a simple histogram gives the same message); there is a tendency towards positive

---

[2]All contour plots are made using 'interp()', an S-Plus function.
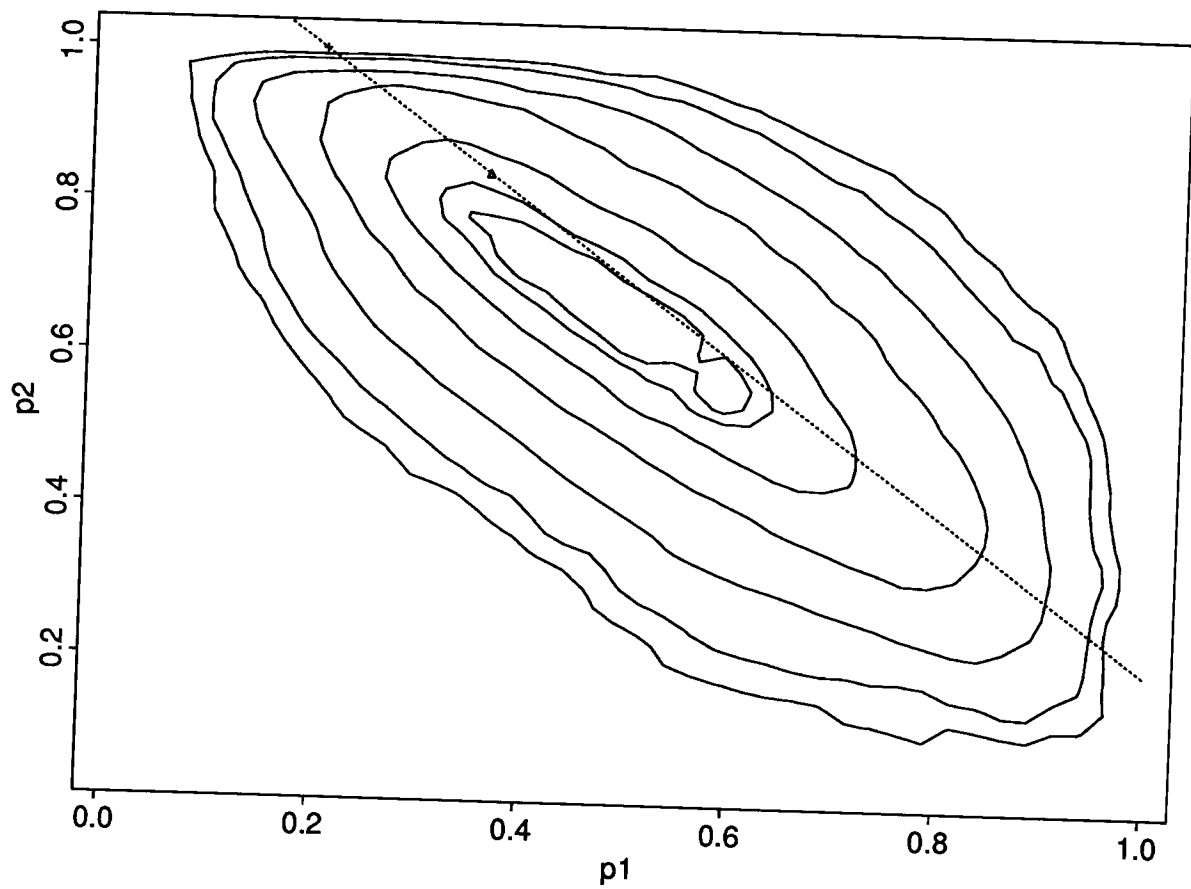
Figure 2.1: Contour plot for posterior distribution of (p_1, p_2). Contour lines are for .05, .10, .25, .50, .75, .90, .95 of largest observed density.

(a)

(b)
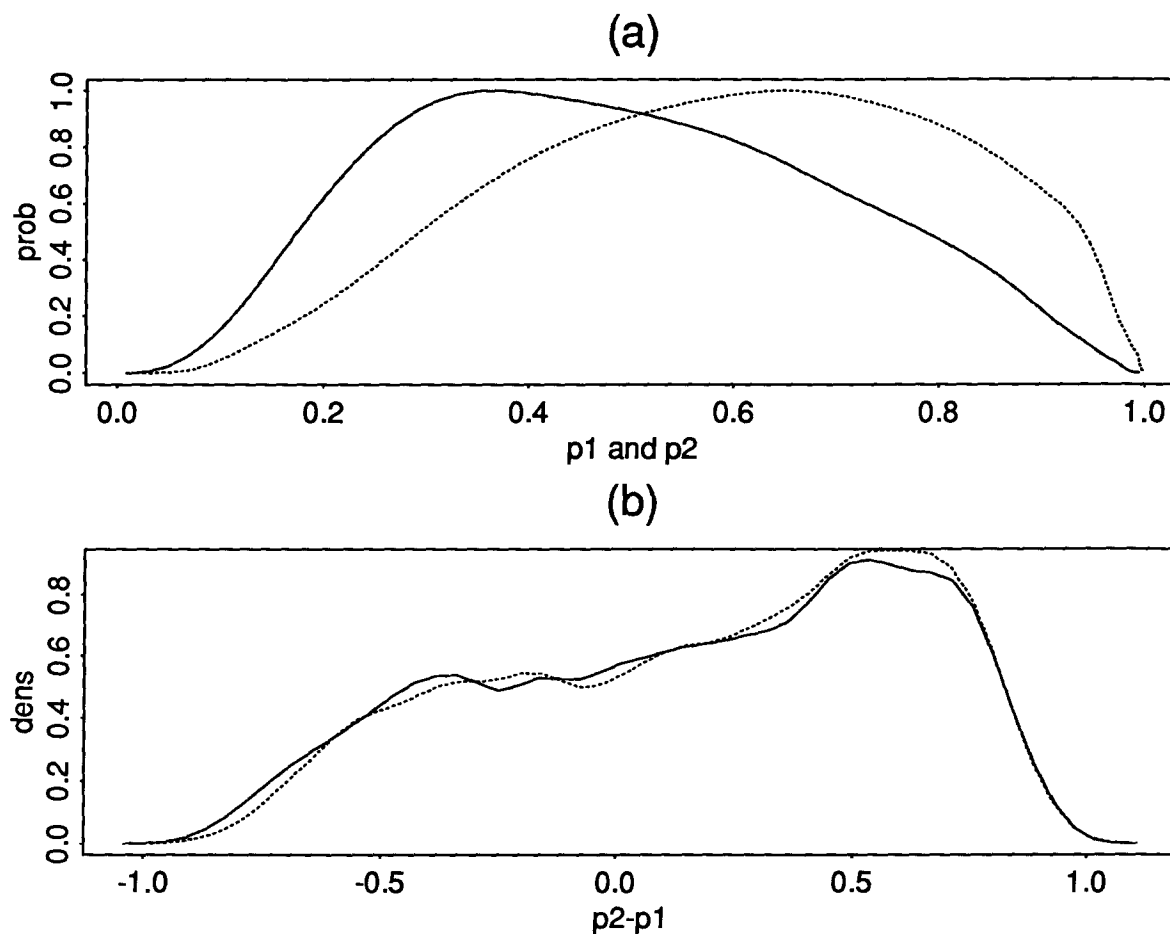
Figure 2.2: (a) Marginal posterior for p_1 (full line) and p_2 (broken line). (b) Density estimates for $(p_2 - p_1)$ for 4000 sample points (full line) and the 500 values remaining after clustering (dotted line).

values, particularly in the range 0.5 to 0.7. The dotted line on the same graph, was obtained in an analogous way, but using only the 500 clustered values $\Theta_2^{(500)}$; little information is lost and the result is essentially the same.

In my description, I have avoided results in the form of point estimates. Posterior mode, median and (Monte Carlo) mean could be used as such. But, the posterior distribution is the full expression of current uncertainties and beliefs about the parameters; why say less (with a point estimate) if I can say it all with the posterior distribution ?

2.4 Sequential Analysis

Here I shall discuss a class of models for which the parameter θ is assumed time-dependent; hence I use θ_t to associate it with time t. This topic is not needed for the developments in the next chapters, and may be omitted without loss of continuity. But it illustrates another field where the AIS can be used.

There are good biological reasons to distrust time-invariant θ for models describing the dynamics of exploited fish stocks (Walters [69], Parma [47]). Alternatively, we might take θ_t to describe the state of the system at time t. Some examples: define θ_t as the stock biomass at time t; or define $\theta_t = (X_{c_1,t}, \cdots, X_{c_k,t})'$ to be a k-dimensional vector where $X_{c_i,t}$ describes the state of classes c_i at time t. For the latter case, think of X as some abundance measurement and c_i as different locations; or think of X as number of fish (or total weight) caught at age- or size-classes c_i.

State-space models and Kalman filtering (Brockwell and Davis [4]) have been used in such situations (Walters [68], Mendelssohn [42], Sullivan [66]). The Bayesian alternatives are *dynamic Bayesian models* (DBM), the subject of this section. An extensive treatment of DBM's is given by West and Harrison [77]. I shall describe DBM in its general form and show why (and how) adaptive importance sampling (AIS) can overcome difficulties in practical implementation. I conclude with an application to a non-linear non-Gaussian example, where Kalman filtering does not apply.

In its general form, the elements of a DBM are a discrete time-series of s-dimensional observations Y_t ($t = 1, 2, \cdots$), each of which is related to a p-dimensional parameter θ_t representing the state of the system at time t. The dynamics are described as follows. At time t, the observation Y_t has a known sampling distribution.

$$(Y_t \mid \theta_t) \sim p_o(Y_t \mid \theta_t) \tag{2.6}$$

Conditional on θ_t, Y_t is assumed independent of Y_j and θ_j, for all $j \neq t$ (past or future). The

state parameter is assumed to evolve in time according to an evolution density

$$(\theta_{t+1} \mid \theta_t) \sim p_e(\theta_{t+1} \mid \theta_t) \tag{2.7}$$

Once θ_t is given, θ_{t+1} is conditionally independent of past values θ_j for $j < t$.

For each time $t = 1, 2, \cdots$ the defining probability densities are given by the *observation model* (2.6) and the *evolution model* (2.7). Let D_t be the information available at the end of period t. If, from t to $t + 1$, the only relevant information is the new data point Y_{t+1}, then the recursive relation $D_{t+1} = \{Y_{t+1}, D_t\}$ describes the dynamics of sequential information acquisition. The conditional densities of (2.6) and (2.7) typically also depend on D_t. I often exclude this term from the notation to avoid clutter.

The Bayesian element in the model is the expression of uncertainties about θ_t by probability densities. In particular, assume that

$$p(\theta_j \mid D_t), \; j = 1, 2, \cdots \tag{2.8}$$

describes the uncertainty about θ_j given all current information D_t. In the context of time-series, the probability densities (2.8) have different names depending on the j used: *filtered* for $j = t$; *smoothed* for $j < t$ and *predictive* for $j > t$. In standard Bayesian terminology (Section 1.3), filtered and smoothed densities are *posteriors* given D_t, while predictive densities are *priors* for future states. The Bayesian (subjectivistic) interpretation of these probabilities allows the modification of any such density if relevant information *outside* of D_t becomes available.

To use the Bayes Theorem, we need a sequential updating mechanism for each additional observation Y. This is done by the following 3 steps:

Step 1: Starting at some time $t-1$ with a given posterior (filtered) distribution $p(\theta_{t-1} \mid D_{t-1})$, get a prior (one-step-ahead predictive) density for θ_t as

$$p(\theta_t \mid D_{t-1}) = \int p_e(\theta_t \mid \theta_{t-1}, D_{t-1}) p(\theta_{t-1} \mid D_{t-1}) d\theta_{t-1}$$

Step 2: Get the (marginal) one-step-ahead predictive distribution for the observable Y_t as

$$p(Y_t \mid D_{t-1}) = \int p_o(Y_t \mid \theta_t, D_{t-1}) p(\theta_t \mid D_{t-1}) d\theta_t$$

Step 3: For an observed value $Y_t = y_t$, the probability $p_o(y_t \mid \theta_t, D_{t-1})$ acts as a likelihood for θ_t. Hence, using Bayes Theorem, update the prior of Step 1 to get the posterior (filtered) density for time t,

$$p(\theta_t \mid D_t) = p(y_t \mid D_{t-1})^{-1} p(\theta_t \mid D_{t-1}) p_o(y_t \mid \theta_t, D_{t-1})$$

The evaluation of the above integrals has no analytic solution in general. Important exceptions — Dynamic Linear Models and Multi-Process Models — are discussed in detail by West and Harrison [77].

In case of non-linear functions and/or non-Gaussian noise structures, two types of approximations are used: analytical and numerical. Analytical approximations are used, for instance, by Delampady et al. [11] and [10]. They apply *Laplace approximations* of Tierney and Kadane [67] to calculate posterior moments for a Bayesian version of an autoregressive time-series model AR(1), and for a model where θ_t is the intensity parameter of a Poisson process.

Numerical approximations start by considering a grid of values for θ_t. Choice of this grid is an important point for the distinction among different methods. Sorenson [64] identifies two general groups:

(*i*) *local methods* which restrict analysis to one grid point only; and

(*ii*) *global methods* which use the whole grid.

Local Methods

The best known local method is the *extended Kalman filter* (EKF) (for application to fisheries see Walters [68]). To describe it, I define

$$E[Y_t \mid \theta_t, D_{t-1}] = F(\theta_t, D_{t-1}) \tag{2.9}$$

with expectation over observation error v_t; an $(s \times 1)$ random vector; and

$$E[\theta_t \mid \theta_{t-1}, D_{t-1}] = G(\theta_{t-1}, D_{t-1}) \tag{2.10}$$

with expectation taken over system noise w_t; a $(p \times 1)$ random vector.

The functions $G(\cdot)$ and $F(\cdot)$ — which can be time-dependent — are assumed known. The EKF linearizes $G(\cdot)$ and $F(\cdot)$ with respect to θ., assumes Gaussian distributions and chooses the grid point to be some "best linear estimate" (The ordinary Kalman filter is the special case where $G(\cdot)$ and $F(\cdot)$ are linear). The updating formulas used in Kalman filtering are equivalent to the formulas used to update mean and variance for a Gaussian conjugate family of distributions. Therefore, within the Bayesian context, the use of the EKF *implies* that the posterior distributions for θ_t are Gaussian. According to Sorenson [64], the first serious step away from a Gaussian assumption involved the use of Edgeworth expansions and "seems to be most useful when the posterior density is unimodal even though not Gaussian".

Global Methods

Any global method needs a way to solve the general problem of

1. defining an initial grid,

2. giving a procedure to define the grid at the next time step, and

3. for a given grid, select a method to approximate the probability density.

Kitagawa [27] approximates integrals by a finite sum defined over a discrete equi-spaced grid for θ_t, where densities are approximated by a first-order spline function. Smith et al. [63], Pole [49], Pole and West [50], propose an adaptive grid using Gauss-Hermite quadrature. West [74] points out the difficulties with the quadrature method:

(i) the enormous computational demand in problems with more than a few parameters,

(ii) the difficulty of reconstructing smooth posterior distributions based on the evaluation of only a few points, and

(iii) the need to work with an approximately orthogonal transformation of the parameters.

In the adaptive importance sampling approach of West [75] [76] the grid points are randomly generated. Such dynamic grids will be concentrated in regions of interest in the parameter space (provided the mixture models are good approximation to the "true" underlying distribution).

How to (adaptively) find a good mixture model, was described (for time-invariant θ) in Section 2.3. I will show next how this can be modified for sequential models.

2.4.1 AIS in Sequential Models

Using the notation of Section 2.3 (with the obvious changes to include the time index t) and given the observation model (2.6), the evolution model (2.7), and the data set D_{t-1}, the posterior $p(\theta_{t-1} \mid D_{t-1})$ is summarized by $\Gamma_{t-1} = \{g_{t-1}(\cdot), n_{t-1}, \Theta_{t-1}, \Omega_{t-1}\}$, where $g_{t-1}(\cdot)$ is the final mixture model *before* clustering and n_{t-1} the corresponding sample size. The updating mechanism (Steps 1 to 3 above) can be rewritten as Monte Carlo approximations.

Step 1: *(i)* If the expectation $E_e[H(\theta_t) \mid \theta_{t-1}]$ with respect to the evolution density (2.7) can be explicitly evaluated for some function of interest $H(\cdot)$, then its predictive expectation is

$$E[H(\theta_t) \mid D_{t-1}] \approx \sum_{j=1}^{n_{t-1}} \omega_{t-1,j} E_e[H(\theta_t) \mid \theta_{t-1,j}]$$

Particularly, for mean and variance,

$$a_t = E[\theta_t \mid D_{t-1}] \approx \sum_{j=1}^{n_{t-1}} \omega_{t-1,j} E_e[\theta_t \mid \theta_{t-1,j}]$$

$$R_t = Var[\theta_t \mid D_{t-1}] \approx \sum_{j=1}^{n_{t-1}} \omega_{t-1,j}(Var_e[\theta_t \mid \theta_{t-1,j}] + (E_e[\theta_t \mid \theta_{t-1,j}] - a_t)^2)$$

(ii) The prior (predictive) distribution $(\theta_t \mid D_{t-1})$ is approximated by

$$p(\theta_t \mid D_{t-1}) \approx \sum_{j=1}^{n_{t-1}} \omega_{t-1,j} p_e(\theta_t \mid \theta_{t-1,j})$$

This approximation is used to generate a stochastic sample $\Theta_t^* = \{\theta_{t,j}^*; j = 1, \cdots, n_{t-1}\}$ from the prior distribution. It is also used to evaluate prior probabilities in step 3 below.

Step 2: The sets Ω_{t-1} and Θ_t^* are used for inferences from the (marginal) forecasting distribution $(Y_t \mid D_{t-1})$. Particularly,

$$f_t = E[Y_t \mid D_{t-1}] \approx \sum_{j=1}^{n_{t-1}} \omega_{t-1,j} E_o[Y_t \mid \theta_{t,j}^*]$$

$$Q_t = Var[Y_t \mid D_{t-1}] \approx \sum_{j=1}^{n_{t-1}} \omega_{t-1,j}(Var_o[Y_t \mid \theta^*_{t,j}] + (E_o[Y_t \mid \theta^*_{t,j}] - f_t)^2)$$

where E_o and Var_o are calculated with respect to the observation model (2.6).

The marginal probability $p(Y_t \mid D_{t-1})$ is given by

$$p(Y_t \mid D_{t-1}) \approx \sum_{j=1}^{n_{t-1}} \omega_{t-1,j} p_o(Y_t \mid \theta^*_{t,j})$$

Step 3: After $Y_t = y_t$ was observed, the goal is to obtain $\Gamma_t = \{g_t(\cdot), n_t, \Theta_t, \Omega_t\}$ as an approximation for the posterior (filtered) distribution $p(\theta_t \mid D_t)$, completing one recursion. The procedure is as follows:

1. Use Θ^*_t and Ω_{t-1} to: (*i*) calculate the Monte Carlo estimate of V as indicated in Section 2.3 and (*ii*) obtain the clustered sets $\Theta^c_t = \{\theta^c_{t,j}; j = 1, \cdots, m\}$ and $\Omega^c_t = \{\omega^c_{t,j}; j = 1, \cdots, m\}$ with $m < n_{t-1}$ elements.

2. Define an initial mixture

$$g_{t,0}(\theta) = \sum_{j=1}^{m} \omega^c_{t,j} d(\theta \mid \theta^c_{t,j}, h^2 V)$$

and produce a sample $\Theta_{t,0} = \{\theta_{t,0,j}; j = 1, \cdots, n_{t,0}\}$ from it.

3. Calculate the corresponding set of weights $\Omega_{t,0}$, from $g_{t,0}(\theta_{t,0,j})$ and

$$f_t(\theta_{t,0,j}) \propto p_o(y_t \mid \theta_{t,0,j}) p(\theta_{t,0,j} \mid D_{t-1})$$

4. The summary $\Gamma_{t,0} = \{g_{t,0}(\cdot), n_{t,0}, \Theta_{t,0}, \Omega_{t,0}\}$ is used to repeat the procedure. Repeat the procedure to obtain $\Gamma_{t,1}$, $\Gamma_{t,2}, \cdots$ until this set is considered to mimic well the true posterior. Use the final set as Γ_t.

For the initial time $t = 1$ the density $\pi(\theta_1) = p(\theta_1 \mid D_0)$ has to be specified; usually a diffuse distribution is chosen, from which a random sample can be drawn easily. This random sample is $\Theta_{1,0}$. To obtain the corresponding weights $\Omega_{1,0}$, define $g_{1,0}(\theta) = \pi(\theta)$ so that $\omega_{0,j} \propto p_o(Y_1 \mid \theta_{1,0,j})$. The above procedure is then applied to generate $\Gamma_{1,1}$, etc.

2.4.2 An Example

For illustrative purposes, I shall consider a (hypothetical) situation in which the biomass of some resource is described in time. I use a non-linear non-Gaussian model for which classical (Kalman filter) methods do not apply.

The biomass dynamic is described by

$$B_t = S_{t-1} \exp(a - bS_{t-1}^p)e^{w_t} \tag{2.11}$$

where w_t are independent random variables from a Student distribution with 9 degrees of freedom mean 0 and constant variance W. S_t denotes the biomass surviving year t. Given an annual natural survival rate ζ, and h_t the harvest rate in year t, I define S_t as

$$S_t = \zeta(1 - h_t)B_t \tag{2.12}$$

Each year some abundance index I_t is observed. The relation between I_t and the unknown biomass B_t is modelled by

$$I_t = qB_t e^{v_t} \tag{2.13}$$

where the observation errors v_t are assumed independent random variables from a Student distribution with 9 degrees of freedom, mode 0 and constant variance V.

After defining $\theta = \ln B$ and $Y = \ln I$, the following dynamic model is derived from (2.11), (2.12) and (2.13):

$$\begin{cases} \theta_t = G(\theta_{t-1}, h_{t-1}) + w_t \\ Y_t = F(\theta_{t-1}) + v_t \end{cases} \tag{2.14}$$

where

$$G(\theta, h) = d_1 + \ln(1 - h) + \theta - d_2(1 - h)^p e^{p\theta}$$

$$F(\theta) = \ln q + \theta$$

$$d_1 = \ln \zeta + a$$

$$d_2 = b\zeta^p$$

Table 2.2: List of parameters used in the simulation model for artificial data generation

Parameter	Value
ζ	0.8
a	1.65
b	0.20
p	2
B_1	4.7
q	1
V	0.13
W	0.04
d_1	1.43
d_2	0.13

Table 2.3: Output from simulation. The harvest rates h_t were fixed in advance. B_t is the Biomass at the start of year t and I_t is the corresponding observed abundance index. θ_t and Y_t are the natural logarithm of B_t and I_t, respectively.

t	h_t	B_t	I_t	θ_t	Y_t
1	0.2	4.71	3.22	1.55	1.17
2	0.4	3.39	1.67	1.22	0.51
3	0.6	3.90	2.59	1.36	0.95
4	0.6	4.62	5.81	1.53	1.76
5	0.3	5.70	5.16	1.74	1.64
6	0.3	2.12	2.51	0.75	0.92

Notice the non-linearity of $G(\cdot)$ with respect to θ. By construction the process is non-Gaussian as well. For the purpose of illustration, all parameters a, b, p , q, ζ, W and V are assumed known with values listed in Table 2.2.

I simulated the process for 6 time steps with harvest rates h_t specified for times $t = 1, \cdots, 6$. Table 2.3 summarizes the outcome.

Notice the substantial drop in biomass at time $t = 6$; possibly an effect of the heavy-tailed Student distribution modelling the random process noise w.

The initial uncertainty about θ_1 is described by a prior $\pi(\theta_1)$ which is Student with 9 degrees of freedom, mean 1 and scale 0.36 (the dotted line for $t = 1$ in Figure 2.3). After observing
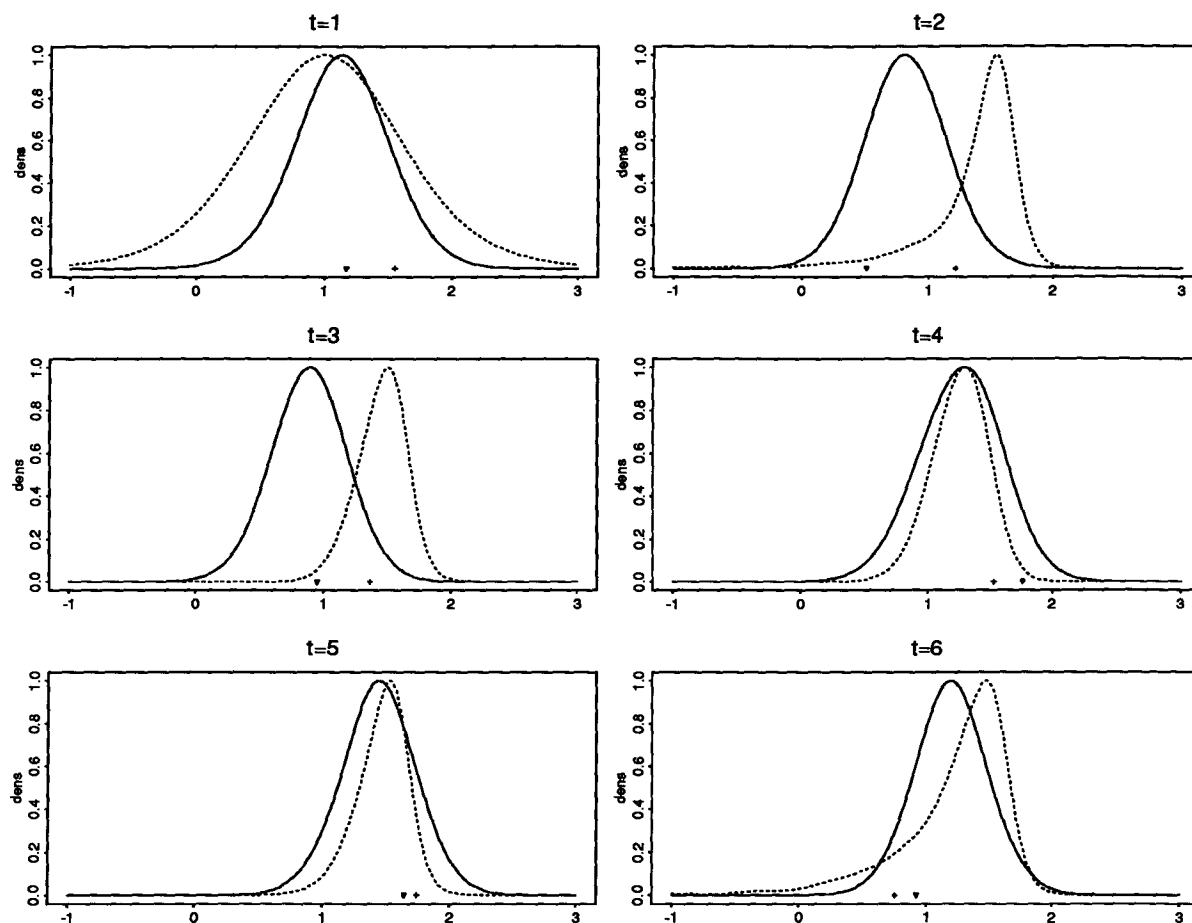
Figure 2.3: Prior (predictive) and posterior (filtered) distributions are shown respectively by dotted and full lines, for times $t = 1$ to $t = 6$. '∇' is the observation Y_t and '$+$' indicates the "true" θ_t.

Y_1, the posterior distribution $p(\theta_1 \mid Y_1)$ is obtained (the full line on the same graph). The use of $G(\theta_1, h_1)$ together with the known Student distribution for w_1 enables the calculation of the prior (predictive) distribution $p(\theta_2 \mid Y_1)$ (the dotted line for $t = 2$ in Figure 2.3). After observing Y_2, the process is repeated; then for Y_3, etc.

At each t, I repeated the AIS procedure 3 times, with sample sizes $n_0 = 250$, $n_1 = 350$ and $n = 1000$. The mixture model (2.3), used to approximate the posterior $p(\theta_t \mid Y_1, \cdots, Y_t)$, was obtained after clustering the sets Θ and Ω to $m = 100$ components. I used the conventional

bandwidth multiplied by 0.65 to allow for possible multi-modal distributions. The relative entropy (2.4) was always larger than 0.997 for the final mixture approximation, indicating a good agreement between the importance function $g(\cdot)$ and the corresponding posterior $p(\cdot)$. For times $t > 1$, densities for predictive (prior) distributions $p(\theta_t \mid Y_1, \cdots, Y_{t-1})$ where approximated by

$$\sum_{j=1}^{1000} \omega_{t-1,j} p_e(\theta_t \mid \theta_{t-1,j})$$

where $p_e(\cdot)$ is the (known) evolution model: a Student density with 9 degrees of freedom, mean $G(\theta_{t-1}, h_{t-1})$, and variance W. In Figure 2.3, prior (broken lines) and posterior for θ_t, are shown for times $t = 1$ to 6. For convenience, probabilities (y-axis) were standardized for maximum mode at unity. The information provided by $Y_1 = 1.17$ (displayed as '$\triangledown$') reduces uncertainties as is reflected in the sharper tails of the posterior. The small deviation of Y_1 with respect to the prior mode 1, caused the minor shift of the mode to the right. The true (unobserved) biomass indicated by '+' was in fact higher than prior mode or observation would indicate. The influence of $G(\cdot)$ in shaping predictions can be seen in the prior for $t = 2$. The asymmetric shape of this distribution supports the inappropriateness of a Gaussian approximation (which would be implicit in case of EKF). A second observation $Y_2 = 0.51$, surprisingly low according to predictions, causes a substantial revision in current beliefs and a smaller posterior mode. For times 3, 4 and 5 analogous considerations describe the dynamic of information acquisition. At $t = 6$, the predictive distribution again presents the asymmetric feature displayed at $t = 2$.

Two features about the predictive distributions deserve comment: (1) the consistent limitation to values below $\theta = 2$, and (2) the striking presence of sharper tails in the predictive distribution (as compared to the posterior) for times 3, 4 and 5. Both are reflection of the particular form of $G(\cdot)$. But, while (1) could be anticipated by the non-linearity in the biomass production model, (2) was a surprise.

2.5 Summary

This chapter was about the implementation of the Bayesian paradigm where numerical approximations of integrals involving some probability function $p(\cdot)$ are needed. Monte Carlo integration with adaptive importance sampling was used to calculate these approximations.

In Section 2.2, I gave a review of the concepts and properties related to *basic importance sampling*. This approach depends critically on the existence of a "good" *importance function* to replace $p(\cdot)$. A procedure to find such a function adaptively is described in Section 2.3. This method, proposed by West [75], uses a finite mixture of Student distributions as the importance function. This function is refined by a recursive mechanism. Such procedure is particularly attractive for multivariate $p(\cdot)$. An example was included to illustrate the details of the required calculations.

Section 2.4 extended the application of the adaptive method to situations where the parameter of interest changes in time and the data are acquired, one by one, in sequential order. An illustrative example was provided.

Chapter 3

Models for Biomass Dynamics

3.1 Introduction

In Chapter 1 I described the importance of Bayesian analysis for fisheries assessment and management. Some technical difficulties which have hindered a more widespread application in the past can be overcome by the numerical methods described in Chapter 2. The present chapter integrates the previous two, and proposes procedures for Bayesian analysis of biomass dynamic models.

The novelty is the Bayesian framework in which traditional fishery models are treated. I shall establish notation, describe, and justify the proposed procedures. The main result of the chapter is the posterior distribution for θ. The potential advantages and new insight gained from such an approach will be explored in Chapter 4 using case-studies.

Here is a short guide through the chapter. The rest of this introduction describes some general features of fishery models and data. The reader familiar with them can move to Section 3.2 where I introduce the first group of models. Section 3.3 presents a second group of models for which I extend the time-series fitting procedure of Pella and Tomlinson [48]. In Section 3.4 I describe delay-difference models — again an optional section. Finally, Section 3.5 is a summary of the chapter.

Fishery Data and Models

Models describing the biomass dynamics of fish stocks are of central importance in the assessment of the resource. By providing information about productivity, such models are used to

find optimal policies for exploitation. Generally speaking, these models imply that highest productivity is obtained at some intermediate biomass. If the biomass level is too low, the stock is being *overfished* in the sense of catching too many fish too early, hence wasting potential for reproduction and growth. At low biomass, the surplus due to growth typically exceeds losses caused by natural mortality. Unduly high biomasses, on the other hand, favour mortality over growth. At least two causes explain this phenomenon: (1) the large proportion of older fish, which no longer use food efficiently for body growth; and (2) the waste of energy caused by competition for limited food supply.

The above description makes it clear why estimates of biomass, and the attempt to model its dynamic, are central in assessment studies. In the past scientists worked with a static concept of "optimum biomass" obtained under equilibrium (time-steadiness) conditions. The inappropriateness of such a concept has become apparent from its failure in practical situations and from a growing number of simulation studies. Recent developments have moved away from these earlier approaches in at least two ways. First, there is the use of dynamic (as opposed to equilibrium) models and the view of optimum biomass as a "moving target" subjected to uncontrollable external factors and affected by the history of exploitation; in fact, optimality itself is being questioned. Second, there is an increasing emphasis on statistical models to describe the structure and dynamics of uncertainty.

Management strategies need to reflect these changes. The historic (and still widely used) concept of "Maximum Sustainable Yield" (MSY), defined as "the average yield which would sustain the resource at its optimum equilibrium biomass", is outdated . MSY has been replaced by concepts such as: (i) optimum fishing effort maximizing future average discounted catch (Ludwig and Walters [38]) or (ii) total allowable catch quotas (TAC's) which — subjected to some constraints — result in (estimated) smallest risk of biomass depletion (Fournier and Warburton [17]). Still more elaborate strategies using stochastic dynamic programming (Walters [68]) are possible. Such strategies explicitly acknowledge that an action today (by affecting the dynamics into the future) will affect the state of the stock tomorrow. Uncertainty reduction (learning)

as well as economic gains can be combined to determine an acceptable compromise. Finally, in order to disclose confounding effects between environmental factors and those related to fishing and stock size, management actions can be guided by general principles of experimental design (Walters and Collie [70]).

The usefulness of the Bayesian estimation procedure and its link to management decisions, was described in Chapter 1. I will not further elaborate on the topic (although I shall come back to it in the concluding chapter), but turn to other aspects relevant for stock assessment: (1) *biomass* as an aggregate measurement of the stock, (2) *data* types and acquisition, (3) *models* and estimation of parameters.

Biomass

The appeal of stock description in terms of its biomass is simplicity. But, as for all aggregate measurements, there is a price to pay: a loss of information regarding the internal structure of the stock. A large number of young and small fish can have the same biomass as a smaller number of older large fish. From a biological standpoint they can be very different in behaviour, food requirement, reproductive potential, and mortality rate. In commercial terms they might also represent distinct markets, prices, and fishing gear requirement (hence, costs). Similar comments apply to the spatial and/or temporal aggregation of local biomass measures.

Data

I shall give a short account of the type of data commonly available for fishery assessment, its sources and potential weaknesses. For a broad description of this topic including survey design, see Hilborn and Walters [25].

The data in fisheries have three sources: commercial catches, surveys and tagging. Data from the commercial fleet form the primary source of information about the exploitation history of the stock. Total catches (often better defined as "landings", because they ignore onboard discharge of economically unattractive or illegal sizes and species) together with fishing effort

and location are collected. Port samples of fish to determine age, sex, length- and weight-distribution are usually collected as well. The ratio of catch per unit of fishing effort (CPUE), is used as an index of abundance and assumed to be proportional to biomass. Such data are typically observational in the sense that there is no control over where and how they are collected at sea. The fisher will go where the activity is economically attractive, be it for high abundance of fish, convenience of location, or any other reason. Assumptions of random sampling do not apply. This fact causes enormous difficulties for reliable estimation of abundance (via CPUE) and represents a serious obstacle to stock assessment.

To avoid the problems associated with commercial data, management agencies have been using research surveys to obtain more accurate estimates of abundance and fishing effort. Surveys are comparatively expensive and usually gather less data than would be required for acceptable levels of confidence in estimates (in the context of classical statistics). But, surveys are valuable as a supplement to commercial data.

Another possibility for abundance estimation has been tagging. Because of high cost, differential fish behaviour, and difficulties with tag recovery, its performance for abundance estimation has been discouraging (although it is a helpful tool in the study of movement and migration patterns).

Models

Fishery data are usually compiled annually so that a discrete modelling of time is appropriate. Therefore the models will be presented in the form of difference (rather than differential) equations.

The levels of confidence associated with classical estimates of parameters in usually non-linear models (and non-linear functions of these parameters) have been calculated in different ways: (*i*) variations on traditional linear and non-linear regression methods and approximations; (*ii*) numerical methods (jackknife, bootstrap). Alternatively, the Bayesian approach taken here provides estimates in the form of posterior distributions; the most complete description of

uncertainties.

I shall present next a detailed description of models and procedures for the practical implementation of Bayesian analysis in fisheries.

3.2 Model 1

In this section I shall present models for stock dynamics which assume only system noise. I call them *Model 1*. The next section considers models with observation error and system noise; I shall call those *Model 2*.

Model 1 is important because it allows us to derive some useful results which build the link between adaptive importance sampling (Chapter 2) and biomass dynamic models. Model 2 takes in the time-series fitting procedure of Pella and Tomlinson [48] — which has proven quite successful for conventional parameter estimation in biomass dynamic models — to calculate the likelihood.

The biomass at time t is denoted by B_t. After harvest, the surviving biomass is S_t. The model describing biomass dynamics is formulated as

$$B_t = S_{t-1}F(\theta, S_{t-1})e^{v_t} \tag{3.15}$$

where $\theta = (\theta^{(1)}, \cdots, \theta^{(p)})' \in \Re^p$ is the unknown p-dimensional parameter of interest and $\{v_t\}$ a sequence of independent random variables with mean 0 and constant (unknown) variance V. Management policies are often given as functions $H(\theta)$ where $H : \Re^p \to \Re$ is known. Prediction into the future depend on the knowledge of θ as well.

When such models are used to describe stock dynamics they are called *surplus-production models* (e.g. Ludwig and Walters [37][38]). When used to describe the relation between the number of spawners (S_j) and the number of returning recruits k years later $(B_j = R_{j+k})$, they are denoted *stock-recruitment models*. But for the purpose of this section such a distinction is unnecessary.

I shall consider two specific forms for $F(\theta, S)$. The *Deriso-Schnute* family of models specifies

$$F(\theta, S) = a(1 - b\gamma S)^{1/\gamma} \qquad (3.16)$$

with $a > 0$ and $b > 0$. Appropriate choice of γ allows formulation of a variety of classical fishery models. For instance, traditional models due to Schaeffer, Beverton-Holt, and Ricker correspond to $\gamma = 1$, $\gamma = -1$ and $\gamma \to 0$, respectively.

An alternative formulation uses the *generalized Ricker* models of Ludwig and Walters [38]. It specifies

$$F(\theta, S) = ae^{-bS^\delta} \qquad (3.17)$$

Both groups, (3.16) and (3.17), overlap at the *standard Ricker model*, for which the limit as $\gamma \to 0$ is taken in the former and $\delta = 1$ is assumed in the latter. In the models described below, I always consider specific sub-families by fixing in advance the parameter γ in case of model (3.16) or δ in case of (3.17). According to Ludwig and Walters [38], an attempt to include these as unknown parameters would produce a worse result than assuming a plausible value instead.

The components of θ need to be defined over the real line. Therefore, I shall define θ as the appropriate transformation of parameters: for a parameter $0 < \alpha < \infty$ I use $\ln \alpha$; if $0 < \alpha < 1$ I use the logit transformation $\ln(\alpha/(1-\alpha))$, etc. Hence, in the above models I define $\theta = (\ln a, \ln b)'$.

Biomass and catch are strictly positive quantities and because of this I have modelled randomness as a multiplicative (non-negative) quantity (e^v). For mathematical tractability it is convenient to divide both sides of (3.15) by S_{t-1} and take logarithms. The following probabilistic model for $Y_t = \ln(B_t/S_{t-1})$ results:

$$p(Y_t \mid \theta, V) \sim d_p(\mu_t(\theta), V) \qquad (3.18)$$

where

$$\mu_t(\theta) = \ln F(\theta, S_{t-1})$$

and $d_p(m, C)$ is some p-dimensional distribution with mean m and covariance matrix C. The value S_{t-1} is assumed known at time of predicting Y_t; hence, $\mu_t(\theta)$ is also available.

Prior uncertainties about V and θ are assumed to be independent and given by $\pi(\theta, V) = \pi(\theta)\pi(V)$.

For a given data set $Y(\tau) = \{Y_1, \cdots, Y_\tau\}$ I will be interested in features of the posterior distribution $p(\theta \mid Y(\tau))$, and sometimes in policy variables $H(\theta)$ or utilities $U(\theta, a)$. If $l(\theta, V) = p_o(Y(\tau) \mid \theta, V)$ is the likelihood for (θ, V) then, given the assumption of prior independence of θ and V, I obtain $l(\theta) = p_o(Y(\tau) \mid \theta)$ by marginalization with respect to V. I.e:

$$l(\theta) = \int p_o(Y(\tau) \mid \theta, v)\pi(v)dv$$

$$l(\theta) = E_\theta^{\pi(V)}[l(\theta, V)]$$

I shall follow the notational convention of Berger [2], and use superscripts on expectations E to indicate the random quantity (or its density) over which the expectation is to be taken. Subscripts on E will denote fixed parameter values. For example, $E_\theta^{\pi(V)}$ means the expectation with respect to the prior for V, conditional on a fixed value θ. Superscript and/or subscript will be omitted whenever the interpretation for E is obvious.

According to Bayes Theorem (section 1.3), the posterior for θ is obtained as

$$p(\theta \mid Y(\tau)) \propto \pi(\theta)E_\theta^{\pi(V)}[l(\theta, V)] \tag{3.19}$$

To use (3.19) in the context of adaptive importance sampling (AIS), we need to evaluate its right-hand side (up to a normalizing constant) for any given θ. Technical results given as Lemmas in the next subsection will enable the practical implementation in Chapter 4, but can be omitted from a first reading without loss of continuity.

3.2.1 Two Lemmas

The results to be given in the two Lemmas of this subsection will be used throughout the applications of Chapter 4.

The first and most important of these Lemmas assumes a Gaussian model in (3.18) together with an Inverse Gamma prior $\pi(V)$. The reason for this combination is the *conjugate property*

of distributions. As was explained in Section 1.3, such property is a helpful tool to simplify calculations in general, and will be used to aid the calculation of $l(\theta)$ in (3.19).

Lemma 3.2.1 *Let $d_p(.,.)$ in (3.18) be Gaussian and $\pi(V) \sim IG(\alpha_0, \beta_0)$, an inverse gamma distribution. The following special cases of (3.19) are obtained given two different assumptions regarding the distribution of $(Y(\tau) \mid \theta, V)$:*

(a) *If $(Y_j \mid \theta, V)$, $j = 1, \cdots, \tau$ are independent, then*

$$p(\theta \mid Y(\tau)) \propto \pi(\theta)(\beta_\tau(\theta))^{\alpha_\tau})$$

where $\alpha_\tau = \alpha_0 + \tau/2$ and

$$\beta_\tau(\theta) = \frac{2\beta_0}{2 + \beta_0 Z(\theta)}$$

with $Z(\theta) = \sum_{j=1}^{\tau}(Y_j - \mu_j(\theta))^2$.

(b) *If $(Y_j \mid \theta, V) \sim N(\mu_j(\theta), V)$ for any fixed θ and $(V \mid Y(j-1), \theta) \sim IG(\alpha_{j-1}, \beta_{j-1}(\theta))$, then*

$$p(\theta \mid Y(\tau)) \propto \pi(\theta) \prod_{j=1}^{\tau} \frac{(\beta_j(\theta))^{\alpha_j}}{(\beta_{j-1}(\theta))^{\alpha_{j-1}}}$$

where $\alpha_j = \alpha_{j-1} + 1/2$ and

$$\beta_j(\theta) = \frac{2\beta_{j-1}(\theta)}{2 + \beta_{j-1}(\theta)(Y_j - \mu_j(\theta))^2}$$

Proof:

- In (a), notice that

$$l(\theta, V) = \prod_{j=1}^{\tau} p(Y_j \mid \theta, V) \propto V^{-\tau/2} e^{-Z(\theta)/2V}$$

and hence

$$E^{\pi(V)}[l(\theta, V)] \propto \int v^{-(\alpha_0 + \tau/2 + 1)} \exp\left(-(\frac{Z(\theta)}{2v} + \frac{1}{\beta_0 v})\right) dv$$

$$E^{\pi(V)}[l(\theta, V)] \propto (\beta_\tau(\theta))^{\alpha_\tau} \int p_\theta(v) dv$$

where $p_\theta(V) \sim IG(\alpha_\tau, \beta_\tau(\theta))$ so that the last integral equals 1.

- For (b) recall that $\pi(V) = \pi(V \mid \theta)$ so that

$$E^{\pi(V)}[l(\theta, V)] = \int p_o(Y(\tau) \mid v, \theta)\pi(v \mid \theta)dv$$

$$= \int p_o(Y(\tau), v \mid \theta)dv$$

$$= p_o(Y(\tau) \mid \theta) = l(\theta).$$

The pair of distributions assumed in (b) form a conjugate family, such that

$$(V \mid Y(j), \theta) \sim IG(\alpha_j, \beta_j(\theta))$$

and from the Gaussian observation model it follows that

$$(Y_j \mid Y(j-1), \theta) \sim T\left(2\alpha_{j-1}, \mu_j(\theta), (\alpha_{j-1}\beta_{j-1}(\theta))^{-1}\right)$$

where $T(k, m, C)$ denotes a Student distribution with k degrees of freedom, mean m and scale C.

The expectation in (3.19) is the marginal likelihood $l(\theta) = p_o(Y(\tau) \mid \theta)$. Because of the time-series nature in the observations, assume Y_j dependent on past data $Y(j-1)$. According to the *multiplication rule*, $l(\theta)$ can be written as the product

$$l(\theta) = \prod_{j=1}^{\tau} p_o(Y_j \mid Y(j-1), \theta)$$

From the assumptions it follows that each term in the product is given by

$$p_o(Y_j \mid Y(j-1), \theta) \propto (\beta_{j-1}(\theta))^{-1/2} \left(1 + \frac{\beta_{j-1}(\theta)(Y_j - \mu_j(\theta))^2}{2}\right)^{-\alpha_j}$$

$$\propto (\beta_{j-1}(\theta))^{-(\alpha_j + 1/2)} \left(\frac{2\beta_{j-1}(\theta)}{2 + \beta_{j-1}(\theta)(Y_j - \mu_j(\theta))^2}\right)^{\alpha_j}$$

$$\propto \frac{(\beta_j(\theta))^{\alpha_j}}{(\beta_{j-1}(\theta))^{\alpha_{j-1}}}$$

which completes the proof. ∎

What distinguishes assumptions (a) and (b), is the use of sequential learning about V in the latter. At each time t an up-to-date estimate of V, given the currently available information $Y(t-1)$, is being used. I expected improved performance under assumption (b). However, for the cases I have analyzed the difference was unimportant in practice.

The second Lemma is useful in cases where some (posterior) estimation of the nuisance parameter V is required, and the assumptions of Lemma 3.2.1 apply. It gives explicit formulas for first and second central moments of the posterior $(V \mid Y(\tau))$; a finite mixture of Inverse-Gamma distributions. These moments are expectations over $p(\theta \mid Y(\tau))$, which can be evaluated easily by Monte Carlo approximation, using the final sets Θ and Ω produced by adaptive importance sampling.

Lemma 3.2.2 *Given the assumptions of Lemma 3.2.1, the first and second central moments of $(V \mid Y(\tau))$ are:*

$$E[V \mid Y(\tau)] = \frac{1}{\alpha_\tau - 1} E^{p(\theta\mid Y(\tau))}[\beta_\tau^{-1}(\cdot)] \text{ for } \alpha_\tau > 1 \,,$$

$$E[V^2 \mid Y(\tau)] = \frac{1}{(\alpha_\tau - 1)(\alpha_\tau - 2)} E^{p(\theta\mid Y(\tau))}[\beta_\tau^{-2}(\cdot)] \text{ for } \alpha_\tau > 2 \,.$$

Proof: These expectations are a direct consequence of the following facts:

1. Under assumption of Lemma 3.2.1(a), and any fixed θ, it follows that

$$p(V \mid Y(\tau), \theta) \propto \pi(V) l(\theta, V).$$

In the proof of Lemma 3.2.1, I showed that this is proportional to an Inverse Gamma distribution with parameters α_τ and $\beta_\tau(\theta)$.

In the case of Lemma 3.2.1(b) and any fixed θ, the conjugate property of distributions implies that

$$(V \mid Y(\tau), \theta) \sim IG(\alpha_\tau, \beta_\tau(\theta)).$$

Notice that, while α_τ is the same for cases (a) and (b), the values of $\beta_\tau(\theta)$ differ.

2. Using properties of conditional expectation, the two moments of interest can be rewritten as follows:

$$E[V \mid Y(\tau)] = E_2\left[E_1[V \mid Y(\tau), \theta]\right]$$

$$E[V^2 \mid Y(\tau)] = E_2\left[Var_1[V \mid Y(\tau), \theta] + E_1^2[V \mid Y(\tau), \theta]\right]$$

where E_1 and Var_1 are expectation and variance taken with respect to $p(V \mid Y(\tau), \theta)$ and E_2 is expectation with respect to $p(\theta \mid Y(\tau))$.

3. Finally, using the fact that for $X \sim IG(a, b)$ the mean is $E[X] = 1/(b(a-1))$ if $a > 1$ and the variance $Var[X] = 1/(b^2(a-1)^2(a-2))$ if $a > 2$, we obtain the result. ∎

3.2.2 Catch and Effort Data

Model (3.18) implicitly assumes direct observation (with negligible observation errors) of the set $\{(S_t, B_{t+1}); t = 1, \cdots, \tau\}$. The stochastic component modelled by v, usually denoted "system noise" or "process error", reflects noise in the dynamic description of the system. Such model applies, for instance, in studies of salmon populations to describe relations between spawning stock S_{t-1} and the returning recruits $B_t = R_{t+k}$ after k years (*Stock-Recruitment Models*), assuming that good estimates of both S and B can be obtained. For most practical cases, however, neither S nor B are available directly.

It is common to have a data history of total catches C_t and the corresponding fishing efforts E_t. These data need to be *translated* into abundance estimates for B and S in order to use Model 1. A widely used model in fisheries defines this relation as follows:

$$C_t = B_t(1 - e^{-qE_t}) \tag{3.20}$$

$$S_t = B_t - C_t \tag{3.21}$$

The term $h_t = 1 - e^{-qE_t}$ in (3.20) denotes the harvest rate in year t as a function of fishing effort; the constant $q > 0$, the *catchability coefficient*, specifying the efficiency of each unit of effort. In (3.21) no mortality other than fishing is assumed. If the rate of survival to natural causes ζ is to be included, this expression can be replaced by $S_t = B_t \zeta h_t$.

From (3.20) and (3.21) the values Y_t and S_t appearing in (3.18) are functions of $q, C_t, C_{t+1}, E_t, E_{t+1}$; that is:

$$Y_t = y(q, C_t, C_{t+1}, E_t, E_{t+1})$$

$$S_t = s(q, C_t, E_t)$$

where these functions are as follows:

$$Y_t = \ln\left(\frac{C_{t+1}(1 - \exp(-qE_t))}{(1 - \exp(-qE_{t+1}))C_t}\right) - qE_t \qquad (3.22)$$

$$S_t = C_t\left(\frac{\exp(-qE_t)}{1 - \exp(-qE_t)}\right) \qquad (3.23)$$

Notice that these calculations require specification of q which can be incorporated into θ. The logarithm of q is the appropriate transformation and I define $\theta = (\ln a, \ln b, \ln q)'$. Given a set of catch and effort data $\{(C_j, E_j); j = 1, \cdots, \tau + 1\}$ and a fixed value for the parameter θ, the corresponding set $\{(Y_j, S_j); j = 1, \cdots, \tau\}$ can be calculated using (3.22) and (3.23). The usual calculations of the posterior $p(\theta \mid Y(\tau))$ can now be performed. But, notice that each given q will result in a different sequence of Y and S values. Also, the likelihood $l(\theta)$ is to be interpreted as the evidence for θ given catch and effort data.

Observation Errors

The model so far has not considered any error in the observation of catch and/or effort. From the discussion in Section 3.1, it should be clear that such an assumption is inadequate. To focus my discussion, I shall assume accurate observation of catches C_t but substantial observation errors in efforts E_t. Such an assumption often agrees with practice if one considers the formidable difficulties in standardizing measurements of effort over an entire fleet of differing vessels, gears, and skills. To examine the effects of these errors, assume E_t to be unknown and that E_t^o is observed instead; the difference $(E_t^o - E_t)$ being the observation error. Let Y_t^o and S_t^o be the alternatives to Y_t and S_t in (3.22) and (3.23) when efforts E are replaced by E^o. To keep the argument simple, assume that the observation errors are adequately modelled by

$$Y_t^o = Y_t + x_{1,t}$$

$$S_t^o = S_t + x_{2,t}$$

where $x_{.,t}$ are random quantities induced by errors in E^o. It is easy to see that (3.15) can be rewritten in terms of observed values as

$$Y_t^o = \ln F(\theta, S_{t-1}^o - x_{2,t-1}) + v_t + x_{1,t}$$

Particularly, model (3.17) with $\delta = 1$ (the standard Ricker model) results in

$$Y_t^o = a - bS_{t-1}^o + v_t^*$$

where $v_t^* = v_t + x_{1,t} + bx_{2,t-1}$. The parameter V in (3.18) is now the variance of the random variables v_t^* (rather than v_t), which include process *and* observation errors. To distinguish the combined variance from the variance associated to (3.15), I denote the latter by V_p: the "process" error.

Such a (convenient) linear arrangement of random terms is not likely to occur in more realistic formulations. The non-linear nature of most models make this relation quite involved and analytical analysis impractical. The impact of "errors-in-variables" on modelling and management has been recognized for quite some time (Walters and Ludwig [72]) and has caused a shift in recent developments towards the explicit modelling of statistical components (Schnute [58]). But the example demonstrates that $V > V_p$; the overall variance is *more* than just process error. How much more will depend on the magnitude of errors in observations and the particular form of the model. If an estimate of V_p is needed — maybe to determine some optimal policy — then the careless use of an estimate for V instead can result in considerable bias in the final outcome. How to overcome such difficulty ? Ludwig and Walters [38], in a similar context, side-step this problem by defining an additional parameter $\lambda = V_p/V$, the ratio between these two variances[1]. The case with no errors in observation would correspond to $\lambda = 1$, while $\lambda = 0$ assumes no process error. According to their performance evaluation, attempts to estimate this parameter together with all the others gave poorer results than to simply fix it at some intermediate value, say $\lambda = 0.5$.

---

[1]They actually use $\lambda = (V - V_p)/V$

Utility Function and Bayesian Decision

The central interest of any practical application is not θ itself. In the classical approach some function $H(\theta)$ is usually to be estimated. A pragmatic approach might seek an answer to the question: "What effort level should be recommended for future years ?" Bayesian decision theory (Section 1.3) can be used in connection with the posterior for θ and some utility function to provide an answer. I shall propose an utility function and describe how such analysis can be performed.

Given a parameter value $\theta_j \in \Theta$ and an effort $e_i \in A$, where Θ is the set of parameter values and A the set of possible management actions, I define $U(\theta_j, e_i) = U_{j,i}$ as the utility associated with such a pair. Further $p(\theta \mid \text{Data})$ denotes the posterior distribution for the parameter θ and

$$U(i) = E^{p(\theta|\text{Data})} [U(\cdot, e_i)]$$

the expected utility for a fixed effort e_i. The Bayes rule or optimal action will be the effort e_{i^*} which maximizes the expected utility; i.e.:

$$U(i^*) = \max_A U(i)$$

A Monte Carlo approximation for $U(i)$ is readily available (recall Section 2.2) and calculated as

$$U(i) \approx \sum_{j=1}^{n} \omega_j U_{j,i}$$

In defining an utility function, it seems sensible to reach two goals simultaneously:

1. To maximize (discounted) future catches.

2. The avoid dangerously low surviving spawning stocks S_t.

For a given pair (θ_j, e_i) and a starting surviving biomass S_0, — which is available from the data, and defined as "the value of S at the final year of observations" — the fitting model $F(\cdot)$ is used to predict the average annual catches $\hat{c}_t$ over a τ years period into the future, given that

a constant effort $E = e_i$ is used for the whole time. For each time $t = 1, \cdots, \tau$ the calculations go as follows:

$$B_t = F(S_{t-1})$$

$$\hat{c}_t = B_t(1 - e^{-qE})$$

$$S_t = B_t - \hat{c}_t$$

The *Cumulative Discounted Catch*(CC) is calculated as

$$CC(j,i) = \sum_{t=1}^{\tau} \delta_{\mathrm{PV}}^t \hat{c}_t$$

where δ_{PV} is the *discounting factor*.

Further I determine the smallest value of S over the same period as being

$$S_{\mathrm{min}} = \min_{1 \leq t \leq \tau} S_t$$

and fix a threshold, say SLIMIT, so that smaller spawning stocks are considered to be an unacceptable risk. Such limit might be associated with a perceived danger of stock collapse if the stock is reduced further.

Finally, I define the utility function as

$$U_{j,i} = \begin{cases} CC(j,i) & \text{if } S_{\mathrm{min}} \geq \text{ SLIMIT} \\ 0 & \text{if } S_{\mathrm{min}} < \text{ SLIMIT} \end{cases}$$

In words, the utility function is given by the total discounted predicted catch if the chosen effort is fixed over a τ year period, *if and only if* during these years the spawning stock *never* falls bellow the given threshold 'SLIMIT'. If the latter occurs, the utility is set to zero.

This utility function is simplistic in two ways: (1) it assumes a linear (i.e., risk-neutral) relation between CC and the utility derived from it; and (2) it defines the utility to be zero when S_{min} falls below the threshold, instead of using some generic penalty. For a particular fisheries, further refinements might be advisable (see the case-study for Orange Roughy). But, for the present purpose, this straightforward definition is adequate since it embraces three characteristics which, I believe, should be present in any such function. Namely,

1. A range of implementable effort levels. If too low, it might kill the fishery; if too high, it would bring overcapitalization.

2. The desire of the industry to maximize catches in order to supply the increasing market demand.

3. The intent to avoid stock collapse, manifested by the managing agency and the general public and an industry which wants to utilize the resource in the future.

3.3 Model 2

For this second formulation denoted Model 2, I shall propose an estimation procedure that builds on the concept of *observation error/time-series* (OETS) fitting, introduced by Pella and Tomlinson [48]. Their central idea (for conventional parameter estimation) is, to fix a parameter value $\theta \in \Theta$ (possibly including the biomass at time $t = 1$) and then use some model to predict the whole time-series. A best θ is selected to minimize some goodness-of-fit criteria $SS(\theta)$. Particularly, Pella and Tomlinson used

$$SS(\theta) = \sum_{t=1}^{\tau} (C_t - \hat{C}_t)^2$$

where C_t denote observed catches and $\hat{C}_t$ the corresponding prediction from the model.

Subsequent developments considered more elaborate *delay-difference* models, as well as more sophisticated goodness-of-fit criteria (Deriso [12]; Schnute [56][57]; Fournier and Doonan [16]; Ludwig et al.[39]; Fournier and Warburton [17]). In the next section I shall describe delay-difference models in more detail. Here, I focus attention on the estimation procedures used by those authors, and suggest ways of incorporating these ideas into the Bayesian framework. The function $SS(\theta)$, in its most general form, is given as

$$SS(\theta) = \lambda \sum_{t=1}^{\tau} w_t^2 + (1 - \lambda) \sum_{t=1}^{\tau} v_t^2 + \sum_{j=1}^{\nu} r_j (x_j - \hat{x}_j)^2$$

where λ denotes the "split" between process noise and observation error as described for Model 1 (previous section), with $\{w_t\}$ and $\{v_t\}$ the corresponding sets of random perturbations. The

values r_j are pre-specified penalties associated to the deviation between model predictions $\hat{x}_j$ and fixed (hopefully true) values x_j. A large penalty associated to a specific $j = j_0$ say, indicates a constraint imposed upon θ, to the subset for which $\hat{x}_{j_0} \simeq x_{j_0}$.

The method is clearly attractive given its great flexibility and conceptual simplicity, although it is at least as "subjective" as the Bayesian approach. All it requires is an efficient algorithm for minimization of $SS(\theta)$ over the set Θ. Difficulties are associated with the estimation of standard errors of such an estimate. Extensive simulation using bootstrap methods (Efron [14][15]) may be necessary. Therefore the Bayesian approach is appealing for at least three reasons:

1. By introducing a prior distribution $\pi(\theta)$, it avoids the necessity to simulate possibly hundreds of replicates to determine variability in the estimates. The effect of alternative priors (and/or likelihoods) can be examined by the more efficient procedure mentioned at the end of Section 2.2

2. It does not require the maximization[2] of $SS(\theta)$. Such maximization forces choice of one optimal parameter θ^*. This ignores the fact that often there are many choices of θ which give a good fit to the data, but would demand different management actions.

3. For more elaborate delay-difference models, θ will be of high dimension. Furthermore, the set $\{x_j; j = 1, \cdots, \nu\}$ in $SS(\theta)$ are estimates from an independent source (outside the data set), and typically have uncertainties associated with them. Non-informative fishery data, allow for estimation of only two or three parameters in θ, with remaining parameters included as additional x_j's. The Bayesian framework allows (at least in principle) for a coherent treatment of all uncertainties about θ and x as probabilities. For convenience, the latter can be incorporated into θ.

To formalize the proposed Bayesian approach, I replace (3.15) in Model 1, by the following

---

[2]Although the maximization of an utility function will be necessary at the later stage of decision making.

pair of equations

$$B_t = G(\theta, B_{t-1})e^{w_t} \tag{3.24}$$

$$I_t = F(\theta, B_t)e^{v_t} \tag{3.25}$$

where (3.24) models the dynamics for biomass B, and (3.25) the relation between biomass and an observed abundance index I. The process noise $\{w_t; t = 1, \cdots, \tau\}$ and the observation error $\{v_t; t = 1, \cdots, \tau\}$ are assumed to be independent identically distributed random sequences with mean 0 and constant variances W and V, respectively. The random variables v_t and w_t are also assumed to be mutually independent.

Let $Y_t = \ln I_t$ and $\mu_t(\theta) = \ln F(\theta, B_t)$. I assume the observation model

$$p(Y_t \mid \theta, V) \sim d(\mu_t(\theta), V) \tag{3.26}$$

where $d(m, C)$ denotes a one-dimensional symmetric distribution with mode m and variance C. I will use Gaussian distributions in the examples of the next chapter.

In analogy to the OETS fitting procedure, I define the time-series of predicted biomasses as follows:

$$\hat{B}_t = \begin{cases} E_\theta^w[B_t \mid \hat{B}_{t-1}] & \text{if } t \geq 2 \\ \hat{B}_1 \text{ fixed} & \text{possibly } \hat{B}_1 \in \theta \end{cases} \tag{3.27}$$

and further define the likelihood for θ to be given by

$$\hat{l}(\theta) = p\left((Y_1, \cdots, Y_\tau) \mid (\hat{\mu}_1(\theta), \cdots, \hat{\mu}_\tau(\theta))\right) \tag{3.28}$$

where $\hat{\mu}_t(\theta) = \ln F(\theta, \hat{B}_t)$.

I propose the following procedure to obtain the posterior $p(\theta \mid Y(\tau))$:

1. Fix a value of θ. (recall from Chapter 2, that a sample of θ's is available from the prior $\pi(\theta)$).

2. Use the models (3.24) and (3.25) together with the definition (3.27) to obtain the sequence of predictions $\{\hat{\mu}_t(\theta); t = 1, \cdots, \tau\}$

3. Use (3.28) to calculate the likelihood $\hat{l}(\theta)$ after appropriate marginalization with respect to V.

4. Obtain the posterior $p(\theta \mid Y(\tau)) \propto \hat{l}(\theta)\pi(\theta)$

Two comments are appropriate: (*i*) Lemmas 3.2.1 and 3.2.2 can be used directly after replacing $\mu_t(\theta)$ by the corresponding $\hat{\mu}_t(\theta)$; (*ii*) the time-series $\{B_t\}$ and $\{I_t\}$ are considered of the same length in the above formulation, but the abundance data need to be available only over a subset of time points. In fact, the Orange Roughy example of Chapter 4 considers such a case. There, catch data (useful in (3.24)) are available for a certain time sequence of length τ say, while abundance estimates were obtained only for the s final years: $\tau - s + 1, \cdots, \tau$.

The side-stepping of process noise by way of definitions (3.27) and (3.28) is an *ad hoc* procedure obtained by analogy to the OETS approach. I will discuss next, "why" (or at least "when") definitions (3.27) and (3.28) are justified.

The handling of process noise

Suppose that in (3.24) I accept model $G(\cdot)$ as a fair description of the underlying dynamics on average, and that model imperfections and remaining sources of variability are adequately captured by the random noise w. It is known from basic statistics that $a^* = E_\theta^w[B_t \mid \hat{B}_{t-1}]$ minimizes

$$E_\theta^w\left[(B_t - a)^2 \mid \hat{B}_{t-1}\right].$$

But a^* is the definition of $\hat{B}_t$ given in (3.27). In Bayesian decision theory (recall Section 1.3), the optimal action (*Bayes rule*) minimizes the posterior expectation of a given loss function $L(\theta, a)$. In particular, $L(B_t, a) = (B_t - a)^2$ is the "squared-error loss"; and the definition of $\hat{B}_t$ imitates the Bayes rule (although the expectation is not taken over the posterior distribution of B_t). Based on the rationale for choosing a Bayes rule, I could, in principle, use any alternative "loss-function" and modify definition (3.27) accordingly without affecting the remaining steps of the overall procedure.

On the other hand, definition (3.28) is not so clear. For simplicity, assume a single known observation y_t so that the subscript t can be dropped. Then, according to (3.26), the likelihood for θ is $l = p(y \mid \mu(\theta))$; while in (3.28) I use $\hat{l} = p(y \mid \hat{\mu}(\theta))$ instead. I am changing the mode of the symmetric distribution from the unknown true $\mu(\theta)$ to its estimate $\hat{\mu}(\theta)$; the latter being a fixed value given θ, while the former is random. Definition (3.28) might be adequate on average if, for a given θ, the expectation of l (with respect to the random μ) is approximately given by $\hat{l}$; i.e:

$$E_\theta^\mu[l] \simeq \hat{l}$$

While unable to assess the reasonability of such an assumption in general, I shall consider a particular case which is relevant for later applications. Assuming Gaussian distributions for w and v (with respective variances W and V), and $\mu = \ln F$ satisfying $E[\mu - \hat{\mu}] = 0$ and $E[\mu - \hat{\mu}]^2 = W$, it is easy to show (using Taylor series expansion up to quadratic terms) that

$$\frac{E_\theta^\mu[l]}{\hat{l}} \simeq 1 + \frac{W}{2V}\left(1 + \left(\frac{y - \hat{\mu}}{\sqrt{V}}\right)^2\right) \tag{3.29}$$

If the left-hand side of (3.29) equals 1 there is no bias; such is trivially true for $W = 0$. For a ratio smaller than 1, $\hat{l}$ is an overestimate of $E_\theta^\mu[l]$, while a value greater than 1 denotes underestimation. Given $W > 0$, and defining the standardized deviation $\epsilon = (y - \hat{\mu})^2/\sqrt{V}$, the following holds for (3.29):

1. The smallest ratio of $1 + (W/2V) > 1$ is obtained for $\epsilon = 0$; the ratio increases with increasing values of ϵ.

2. Since $\hat{\mu}$ is a function of θ, (3.29) changes with θ for fixed y. For those values of θ causing ϵ to be large, the underestimation in $\hat{l}$ is more pronounced.

3. Larger values of V imply smaller values of ϵ (hence, less bias). For $V \gg W$ the bias is negligible.

4. For fixed y and W, different values of θ giving the same squared deviation ϵ result in identical bias.

The central message in (3.29) is that $\hat{l}$ consistently underestimates $E_\theta^\mu[l]$ and, as a function of increasing values of ϵ, the underestimation gets worse.

A bias correction can be attempted if the relation

$$\lambda = \frac{W}{V + W}$$

is assumed, and estimates of λ and V are available. Expression $W/2V$ in (3.29) is then replaced by

$$\frac{\lambda}{2(1 - \lambda)}$$

and W is eliminated from the calculations.

3.4 Delay-Difference Models

Here is an overview of delay-difference (DD) models. Such models can be used for $G(\cdot)$ in (3.24). For a more detailed analysis, see Hilborn and Walters (Chapter 9) [25].

The simple biomass dynamic models described by (3.16) and (3.17) cannot deal with time-delays and changes in population structure. The new class of DD models, introduced by Deriso [12], still describes the dynamics of the aggregate biomass but, by virtue of its construction, can deal with various time-delays. This because growth and recruitment are modelled explicitly. The aggregate biomass B_t at time t, is defined as the sum

$$B_t = \sum_{a=k}^{\infty} N_{a,t} w_a \qquad (3.30)$$

where $a \geq k$ are ages (with $k > 0$ denoting the age of recruitment to the fishery), $N_{a,t}$ the number of fish of age a at the starting of year t and w_a the average body weight for age a.

The possibility of modelling details of growth (w), recruitment $(N_{k,\cdot})$, and survival $(N_{a,t} \rightarrow N_{a+1,t+1})$ separately makes the model very flexible.

In formal terms, the three basic assumptions concerning growth, recruitment and survival to be used throughout are as follows:

- Growth: A linear relation is assumed to describe average weight dynamics for all ages $a > k$, where k is a minimum age at which an individual is first recruited to the fishery. This relation requires two growth parameters α and ρ as follows.

$$w_a = \alpha + \rho w_{a-1}$$

The weight w_k is usually also fixed as an parameter.

- Recruitment: Let S_t represent the *spawning biomass* in a given year t. Particularly, I will define S_t as " biomass which survives harvest in year t". The average recruitment dynamics will be described by a deterministic relation of the form

$$R_{t+u} = R\{S_{t-1}\}$$

where $R\{\cdot\}$ is a given known function, u is a non-negative integer that describes a possible lag between birth and recruitment, and R_{t+u} represents the average *number* of recruits joining the stock at time $t + u$.

- Survival: The total survival fraction in any year, s_t, is usually modelled as the product of a constant natural survival ζ and a time-varying survival to harvest $(1 - h_t)$, such that

$$s_t = \zeta(1 - h_t).$$

Given these basic assumptions, the following pair of dynamic equations for biomass B_t and number N_t can be derived:

$$B_{t+1} = s_t \left(\alpha N_t + \rho B_t \right) + w_k R_{t+1} \tag{3.31}$$

$$N_{t+1} = s_t N_t + R_{t+1} \tag{3.32}$$

Initial conditions need to be specified before the above recurrence equations can be used. The way in which these are established might change according to any particular application. I assume a start from equilibrium (i.e. dropping the t subscripts in (3.31) and (3.32)) and

negligible harvest prior to time $t = 1$ (i.e., $h = 0$). Such an assumption is reasonable if an initial unfished population can be assumed. Rewrite (3.32) as

$$N_e = R_e/(1 - \zeta) \tag{3.33}$$

Replacing N_e in (3.31) and solving for R_e results in an equilibrium recruitment R_e given as

$$R_e = \frac{B_e(1 - \zeta\rho)}{\left(\frac{\alpha\zeta}{1-\zeta} + w_k\right)} \tag{3.34}$$

The recruitment function I will use in the case-study of Chapter 4 is a one-parameter Ricker curve defined as

$$R_{t+u} = R_e \frac{S_{t-1}}{B_e} \exp\left\{ b\left(1 - \frac{S_{t-1}}{B_e}\right)\right\} \tag{3.35}$$

Generalization of Deriso's model was proposed by Schnute [56]. Further improvements by Fournier and Doonan [16] included the use of moments for weight and length as auxiliary information. Incidently, they also considered a Bayesian structure allowing for the inclusion of a prior distribution for θ. This information was then combined with the likelihood to determine the posterior mode as an estimate for θ (the *generalized maximum likelihood estimate* in Bayesian analysis). They did not consider, however, the possibility of obtaining the posterior distribution. Alternatively Schnute [57] proposed a model for which he used *size* rather than *age* classes. He further assumed "knife-edge" recruitment for a minimum size rather than a minimum age. Hence his largest departure from previous approaches, relates to the modelling of recruitment.

An advantage of DD models over the simpler surplus-production family relates to the underlying parameters which, in the DD case, have a direct biological meaning. This is a desirable feature for the Bayesian approach if elicitation of prior uncertainty is to be used. An disadvantage with respect to simpler models is the high dimensionality of the parameter θ. This demands more computation and, because of uninformative data, often results in poorer performance (Ludwig and Walters [37]).

3.5 Summary

In this chapter I have outlined the types of data commonly used to assess the biomass of fish stocks subjected to exploitation. They are: (1) catch and fishing effort statistics from the commercial fleet, (2) abundance estimates obtained through surveys undertaken by management agencies. While the former is not a random sample, the weakness of the latter is its high cost. The practical solution usually is a compromise between both types of data.

Posterior distributions for the parameter θ were obtained for *surplus-production* models in Section 3.2 (Model 1) assuming the time-series of biomass B_t and surviving biomass S_t are directly observed, or estimated from catch and effort data. To deal with time-series of catch and abundance index, or more elaborate *delay-difference* models, I have proposed an adaptation of the *observation error/time-series* fitting procedure of Pella and Tomlinson [48]. This approach was described in Section 3.3 (Model 2). Finally, Section 3.4 gives an account of general features of *delay-difference* models.

The implications of process noise and observation errors in measurements, are also examined throughout. Corrections are possible whenever their relative magnitudes can be assessed.

Chapter 4

Applications

4.1 Introduction

This chapter is dedicated to specific applications of the adaptive importance sampling procedure (Section 2.3) to models for biomass dynamics (Sections 3.2 and 3.3). I shall consider 5 different case-studies, one per section. Each section was written to be self-contained, so that the reader can select the case-study of his or her particular interest. But I emphasize different details of the analysis in different case-studies. This avoids excessive repetition for the reader of all five. The notation developed in the two previous chapters is used throughout. In some cases I use real data, while others rely on simulations.

Sections 4.2 and 4.3 are devoted to "Model 1" of Chapter 3. In Section 4.2 I construct a posterior distribution for the parameters in models for *stock and recruitment*, using data for Skeena river sockeye salmon. Section 4.3 refers to *surplus-production* models. There I use simulated data and compare my results to those obtained by Ludwig and Walters [38], showing that some of their main conclusions can also be verified here. Since my results are posterior distributions, I can perform a Bayesian decision analysis. I suggest a sensible utility function and show that the best decision is robust with respect to alternative models and priors.

Sections 4.4, 4.5, and 4.6 use the fitting procedure outlined in Chapter 3 as "Model 2": the Bayesian alternative to the observation error/time-series fitting method. Section 4.4 is a case study of Orange Roughy, a valuable fishery in New Zealand which has been the subject of recent controversy. I show that the determining factor in the dispute is the choice of an utility function. Different prior distributions and technical details about models are of little relevance. In Sections 4.5 and 4.6 I examine situations with multi-modal posterior distributions. In the

former I again use simulated data, while the latter uses a *delay-difference* model to analyze data for Pacific cod. Finally, Section 4.7 summarizes the results of the chapter.

4.2 Stock and Recruitment: sockeye salmon

I shall consider a Bayesian alternative to the estimation of parameters in stock and recruitment models when applied to data for Skeena river sockeye salmon (*Oncorhynchus nerka*). I use the data presented in Hilborn and Walters [25] (Fig7.1, p.243) and shall conduct my analysis to allow direct comparisons with some of the results presented in their Table 7.4 (p.279).

Pacific salmon have played an important role in the development of stock and recruitment models. This can be explained in part by their particular life history. The fish hatches in fresh water and later migrates to the ocean where it lives most of its life. At the time of sexual maturity (four years later for sockeye), the fish returns to the small watershed where it had hatched. It then reproduces and dies. The size of the spawning stock can be measured in fresh water. Ageing techniques are used to determine, for each spawning stock, the number of adults that recruit to the fishery. This gives the pair of needed observations: spawning stock and the corresponding recruits.

The data consist of a sequence $\{(R_j, S_j); j = 1, \cdots, \tau\}$ of observed data, where S_j is "the number of spawners in year-class j" and R_j "the number of recruits produced by S_j". Hence, in (3.15), I define $B_{j+1} = R_j$ and consider two alternative models to fit the data:

- **R**: Ricker model (Expression (3.17) with $\delta = 1$)

$$F(\theta, S) = ae^{-bS}$$

- **BH**: Beverton-Holt model (Expression (3.16) with $\gamma = -1$)

$$F(\theta, S) = \frac{a}{1 + bS}$$

where $\theta = (\ln a, \ln b)'$.

Table 4.4: Some biologically meaningful attributes of the parameters in the Ricker and Beverton-Holt models.

	R	BH
1. Slope at Origin	a	a
2. S_{max}	$\frac{1}{b}$	∞
3. R_{max}	$\frac{a}{b}e^{-1}$	$\frac{a}{b}$
4. S_{msy}	$\frac{\ln a}{b}(.5 - .07\ln a)$	$\frac{\sqrt{a}-1}{b}$
5. u_{msy}	$\ln a(.5 - .07\ln a)$	$1 - \frac{1}{\sqrt{a}}$

For the deterministic case (i.e.: $V = 0$ in (3.15)), and under equilibrium conditions, the parameter θ relates to quantities of biological meaning. Table 4.4 (adapted from Hilborn and Walters) gives a summary of those quantities. S_{max} is the spawning stock which maximizes the recruitment; this maximum recruitment is given by R_{max}. Notice the basic difference between the models. While **R** predicts the existence of an intermediate spawning stock able to produce maximum recruitment, in **BH** the recruitment tends towards an asymptote, as S increases. S_{msy} is the optimum (spawning) stock size for "Maximum Sustainable Yield" (MSY[1]), with u_{msy} denoting the corresponding optimal harvest rate.

I used lines 1, 2 and 3 of Table 4.4 and the data-plot of R versus S displayed in Figure 4.4 to set the range for the parameters of the prior $\pi(\theta)$. I started by determining five quantiles of an "idealized" (marginal) log-normal distribution for each parameter. If the fitting of a log-normal distribution revealed inconsistency among these quantiles, I would try a new set of values. Table 4.5 gives the quantiles which were finally chosen. For instance, the value b_R (parameter b under the Ricker model) used as benchmark the interval $380 < S_{max} < 1430$, while for b_{BH} I assumed $100 < R_{max} < 5000$. The parameters of the corresponding Gaussian distribution for θ (columns 'm' and 'C') were used as mean and variance of a heavy-tailed two-dimensional Student distribution with 9 degrees of freedom and diagonal covariance matrix.

Prior uncertainties are smaller for model **R** as compared to **BH**. This is a direct consequence of the larger difficulty in eliciting quantiles for b_{BH} as compared to b_R. The reason for such a

---

[1]See Section 3.1 for its definition.
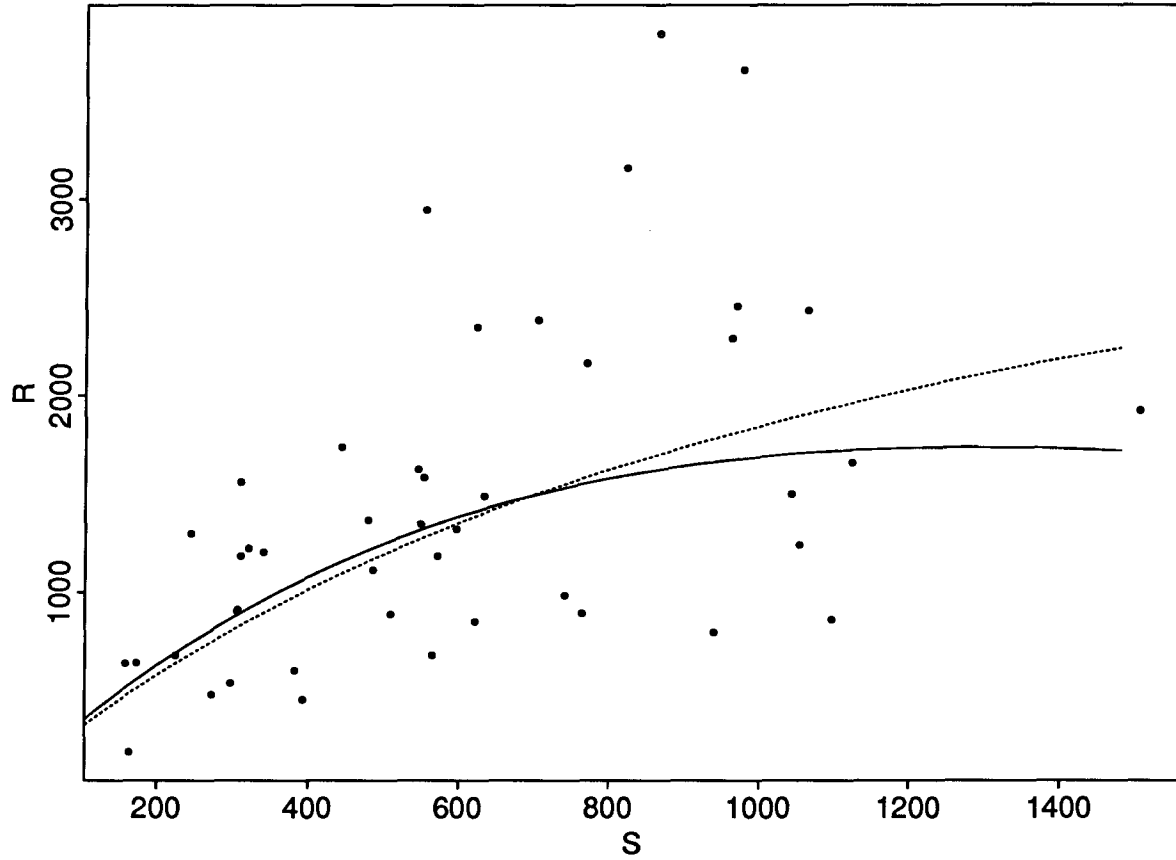
Figure 4.4: Stock and Recruitment data for Skeena river sockeye salmon (Reproduced from Hilborn and Walters [25]). The lines were fitted according to Ricker model (broken) and Beverton-Holt model (continuous) using means of posterior distribution.

Table 4.5: Five quantiles used to determine prior distribution for parameters a and b (the subscript according to the model: Ricker or Beverton-Holt). 'm' and 'C' are mean and variance used in Student prior.

Quantiles	5	25	50	75	95	m	C
a	1.0	1.5	2.0	2.7	4.0	0.70	0.18
$b_{\mathrm{R}} \times 10^{-4}$	7	11	14	18	26	-6.65	0.15
$b_{\mathrm{BH}} \times 10^{-4}$	2	7	14	31	100	-6.55	1.37

difficulty is the different biological meaning of the parameter (see Table 4.4). Prior assessment of b_{BH} needs the joint consideration of two uncertain quantities: R_{max} and the slope at origin a. For comparison, b_R relates only to S_{max}.

As a diffuse prior for V, I chose an inverse gamma distribution $IG(\alpha_0 = 2, \beta_0 = 1)$ which has mean=1, mode=0.33 and ∞ variance.

The goal is to construct the posterior $p(\theta \mid Y(\tau))$, using the AIS approximation of Section 2.3. For convenience, I define the initial importance function $g_0(\cdot)$ to be the prior $\pi(\cdot)$. This results in initial weights $\omega_{0,j}$ proportional to the marginal likelihood $p_o(Y(\tau) \mid \theta_{0,j})$. After sampling $n_0 = 1000$ points from this prior, I perform two updating recursions using $n_1 = 1500$ and $n_2 = n = 4000$, respectively. At each loop the mixture model is clustered to a smaller mixture of $m = 500$ before proceeding. The reported sample sizes were chosen empirically. Sample sizes that were too small caused instability in measurements of the relative entropies (2.4). Monte Carlo estimates for the mean and variance of θ and functions $H(\theta)$, were calculated using the full sample of $n = 4000$ points. With respect to the assumptions described in Lemma 3.2.1, I considered three different scenarios:

- **Ra** Ricker model and assumptions of Lemma 3.2.1 (a).

- **Rb** Ricker model and assumptions of Lemma 3.2.1 (b).

- **BHb** Beverton-Holt model and assumptions of Lemma 3.2.1 (b).

The values of "entropy" in Table 4.6 were calculated according to expression (2.4). This measure is an indication of the goodness-of-fit between the current importance function and the actual posterior densities at the given grid points; a value of 1 corresponds to a perfect fit. The observed entropies indicate the usefulness of the adaptive procedure in constructing a good importance function. In particular, the first recursion shows a considerable improvement, considering the relatively small sample sizes. After focusing the sampling effort on the region of interest, the refinement of a second recursion takes the value of entropy even higher. All final approximations have entropies above 0.99 (recall that 1.0 is a perfect fit). By comparison,

Table 4.6: Results for AIS applied to the Skeena river sockeye salmon data under various assumptions. The first number is the posterior mean; the posterior standard deviation is given in brackets. Columns **R*** and **BH*** are the estimates of Hilborn and Walters [25].

MODEL	Ra	Rb	BHb	R*	BH*
Entropy	.466	.540	.635		
	.955	.984	.984		
	.995	.996	.994		
$V \mid Y(\tau)$	.258(.056)	.259(.056)	.260(.056)	.22	.23
$S_{\text{msy}} \mid Y(\tau)$	718(148)	706(134)	1203(544)	851(246)	916(271)
$u_{\text{msy}} \mid Y(\tau)$	.53(.04)	.53(.04)	.45(.04)	.55(.05)	.48(.06)
$P(S_{\text{msy}} > 800)$	.21(.007)	.19(.006)	.82(.006)	–	–
$P(S_{\text{msy}} > 1000)$	.05(.003)	.03(.003)	.03(.008)	–	–

Table 4.7: Posterior mean and Covariance for different Models fitted to the Skeena river sockeye salmon data.

Model	m_0	m_1	$C_{0,0}$	$C_{0,1}$	$C_{1,1}$
Ra	1.29	-7.20	0.016	0.024	0.058
Rb	1.29	-7.18	0.015	0.021	0.052
BHb	1.20	-7.22	0.026	0.071	0.266

a one-time sample of 6500 points using the initial importance function produced an entropy of 0.60 in model **Ra** and less acceptable results for estimates. The same pattern was observed for other cases.

Posterior mean and variance for $\theta = (\ln a, \ln b)'$ are displayed in Table 4.7 and contour plots of the posteriors for **Rb** and **BHb** are shown in Figure 4.5 (a) and (b), respectively. There is positive correlation between $\ln a$ and $\ln b$. In the case of model **BHb**, one can see that for large values of $\ln a$, the conditional distribution of $\ln b$ is sharper (i.e. less variable) than for smaller values of $\ln a$.

Figure 4.5 (c) and (d), show approximations for posterior densities of S_{msy} and u_{msy}. These posterior densities were obtained by fitting an ordinary (Gaussian) kernel estimator (Silverman [59]) to the 4000 observations of the appropriate function $H(\theta_i)$ (I used the 'density()' function in Splus). The strong agreement between the posteriors obtained for models **Ra** (full line)
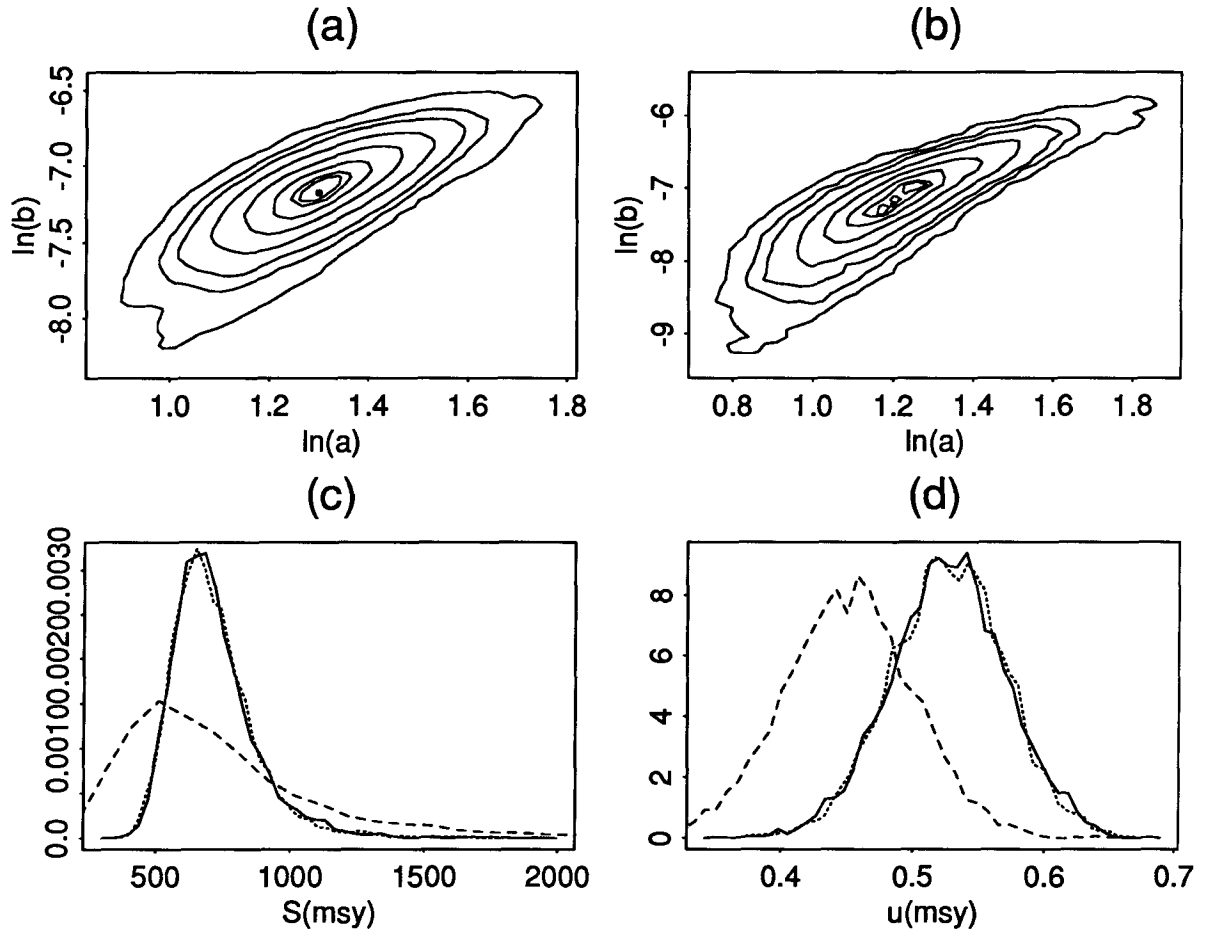
Figure 4.5: (a) and (b): Contour plots of posterior distributions for **Rb** and **BHb** (Lines at .01, .05, .10, .25, .75, .90, .95 of the largest density). (c) and (d): Approximate posterior densities for S_{msy} and u_{msy} (dashed line: **BHb**; full and dotted lines: **Ra** and **Rb**).

and **Rb** (broken line), indicate that there is no much to be gained by the more elaborate assumption (b) in Lemma 3.2.1 for this particular example. A comparison of the estimates reported in Table 4.6 for both models (**Ra** and **b**) also support this conclusion. However, comparing models **Rb** and **BHb**, show marked differences for S_{msy} and u_{msy}. Under **BHb** the posterior for S_{msy} is quite flat as compared to **Rb**. The heavy upper-tail for S_{msy} in the former is supported by the Monte Carlo estimate $P(S_{msy} > 800) = 0.82$ as compared to only 0.19 for the latter. The posterior for u_{msy} also suggests different actions under different model assumptions. If **Rb** is chosen, an optimal harvest rate in the range (0.5, 0.6) seems adequate. Alternatively, for **BHb**, smaller values within (0.4, 0.5) are more adequate. But which one is the better model ?

The two lines fitted in Figure 4.4 use Monte Carlo estimates for the posterior means of the parameters a and b (the Bayes estimators under squared-error loss) for different models: full line for **R** and broken line for **BHb**. I cannot tell from these fits which model is better. The distinction between the models becomes important when predictions for high levels of spawning stock ($S > 1000$) are made. Such model uncertainty can only be resolved by explicit experimentation in management, allowing high spawning stocks (i.e. zero harvest rate !).

The results obtained by Hilborn and Walters [25] are displayed for comparison, as columns **R*** and **BH*** of Table 4.6. There are differences in the values for estimates of both mean and standard error. The strongest disagreement is for S_{msy} under model **BHb**. But in general terms the outcome of both methods agree.

There is an advantage in describing uncertainties by posterior distributions rather than point estimates when an action has to be taken. The next two sections, where the process of decision making under uncertainty is directly addressed, will illustrate this point.

4.3 Surplus Production: Simulation Model

In this section I shall use simulations to generate artificial data for catch and effort. These data are fitted to a model of the form (3.15) according to the procedure described in Section 3.2. I

replicate some of the cases described in Ludwig and Walters [38], (denoted L&W throughout this section) and use their definition of "optimal effort" (E_{opt}), as the policy variable to be estimated from the data. In a second stage of the analysis, I shall replace this policy variable by the effort E_{B} (where B stands for 'Bayes'), chosen to maximize some specified expected utility.

L&W judge the appropriateness of a given model/estimation scheme by examining bias and variability in the estimates for E_{opt}. In order to obtain a distribution of such estimates, they replicate a large set of Monte Carlo experiments for any given fixed simulation model. While interested in the same kind of analysis, I will proceed differently to obtain the distribution of the estimates. Since a stochastic sample[2] of values for θ is available, the corresponding sample for E_{opt} — a function of θ — is available too. Therefore, I can approximate the (Bayesian) posterior distribution of E_{opt} directly, without starting the whole process from scratch.

I shall illustrate the fitting procedure for catch and effort data described in Section 3.2 and compare the outcome to results obtained by L&W. Given the artificial data generation, I have control over the amount of process noise and observation error and can determine its influence over the final outcome of the estimation. Also, the effect of different exploitation histories can be analyzed by defining different effort sequences. In the data generation process I used three different simulation models which I will now describe.

Simulation Models

Three different schemes were used to simulate the artificial data. In all cases I used the "Deriso-Schnute" model (3.16) with $a = 1.65$ and $b = .2$. The cases differ according to: (1) the model to describe the biomass dynamic (choice of γ), (2) the catchability parameter q, (3) the starting stock size S_0, and (4) the global variance $\sigma^2 = V_p + V$ defined as the sum of process noise V_p and observation error V. The values used are listed in Table 4.8. 'Case 3' differs from the other two, by having a much smaller overall variance σ^2 and a comparatively small catchability q.

The parameter $\lambda = V_p/(V_p+V)$ is fixed at $\lambda = 0.5$ in all cases; so that the global variability is

---

[2]Recall that it is a sample drawn from the posterior distribution $p(\theta \mid \text{Data})$.

Table 4.8: Specification of the three simulation models used to generate catch and effort data.

Case No.	γ	q	S_0	σ^2
1	-1	0.6	2	0.16
2	1	0.6	1.5	0.16
3	-0.5	0.31	1	0.05

equally divided between V_p and V. The observation error is associated with effort measurements only. More specifically, for $t = 1, \cdots, 26$ and a sequence of independent Gaussian random variables v_t, with mean 0 and variance V, I assume that

$$E_t^o = E_t e^{v_t}$$

where E_t is a specified effort sequence. The corresponding catch C_t^o is obtained from (3.20) when E_t^o is used as effort measurement. I further use expression (3.21) to get the surviving stock S_t, an input required for the dynamic model (3.16). The random process noise w_t, associated to the biomass dynamics, is modelled as Gaussian with mean 0 and variance V_p.

Fitting Models

After having generated the data, I use another set of models to perform estimation:

R1 The generalized Ricker model (3.17) with power $\delta = 1$.

R2 Similar to above, with power $\delta = 2$.

BH Model (3.16) with $\gamma = -1$, commonly known as the "Beverton-Holt" model.

In all but one case, the "fitting model" is different from the underlying "real" model used to generate the data. For the case in which both models are the same, the posterior distribution describes perceived uncertainties for the "true" parameter vector θ (known from the simulation model). In the remaining situations such interpretation does not apply and I focus on estimates for the policy parameters. As in the approach taken by L&W, I am concerned with the estimation of an *optimal effort* (E_{opt}) and the corresponding *optimal surviving biomass* (S_{opt}).

The exact optimal values are known from the parameters of the simulation model generating the data. I will construct an approximate posterior distribution for such a policy variable by using the AIS procedure of Section 2.3. Further, I will compare my results to those reported by L&W, which were produced by an unrelated estimation scheme.

The optimization criteria of L&W are designed to choose a constant effort to *maximize the expected present value of future harvest*. Appendix B of their paper details the calculation of such an optimal effort E_{opt} and its associated stock size S_{opt}. While the details of these calculations are not of concern here, it is important to point out that E_{opt} is a function of $\theta_0 = \ln a$, $\theta_2 = \ln q$ and the variance of the process noise V_p. Particularly, for the Beverton-Holt model an explicit solution is possible, and given by

$$E_{\mathrm{opt}}(\theta, V_p) = \frac{e^{-\theta_2}}{2}(\theta_0 + V_p - \ln \delta_{\mathrm{PV}}) \qquad (4.36)$$

where δ_{PV} is the discounting constant used in present value calculation (this value is fixed at $\delta_{\mathrm{PV}} = 0.9$ for all my calculations). From the above expression, it is clear that an overestimate of V_p causes overestimation of E_{opt}. This scenario is likely to occur whenever only process noise is assumed. To control this feature, I specify $\hat{\lambda}$, which indicates the amount of global variance to be attributed to process noise. I perform different fits using the assumptions of process noise only ($\hat{\lambda} = 1$), observation error only ($\hat{\lambda} = 0$), and an intermediate value ($\hat{\lambda} = .5$).

In order to verify the effect of different exploitation histories over the estimates, I used two alternative sequences of effort E_t displayed in Figure 4.6. Although the sequence 'Eff 1' is expected to be more informative than the slowly increasing "one-way-trip" of 'Eff 2', the latter is closer to situations encountered in practice. Table 4.9 gives a summary of the different scenarios which were analyzed. Rows 1 to 8 allow comparisons within the same simulation model, given in Table 4.8 as 'case 1'. To make comparisons among different models, one can take for instance rows 5, 10 and 14 (all are R2); or alternatively rows 7, 8, 12 and 15 (all BH). Figure 4.7 shows the simulated sequences of efforts (dotted line) and the corresponding catches (full line). In Figure 4.8 I shown the observed pairs (S_t, B_{t+1}) for the different simulations. The lines exhibit the true model — $G(\cdot)$ in (3.24) — used to generate the data.
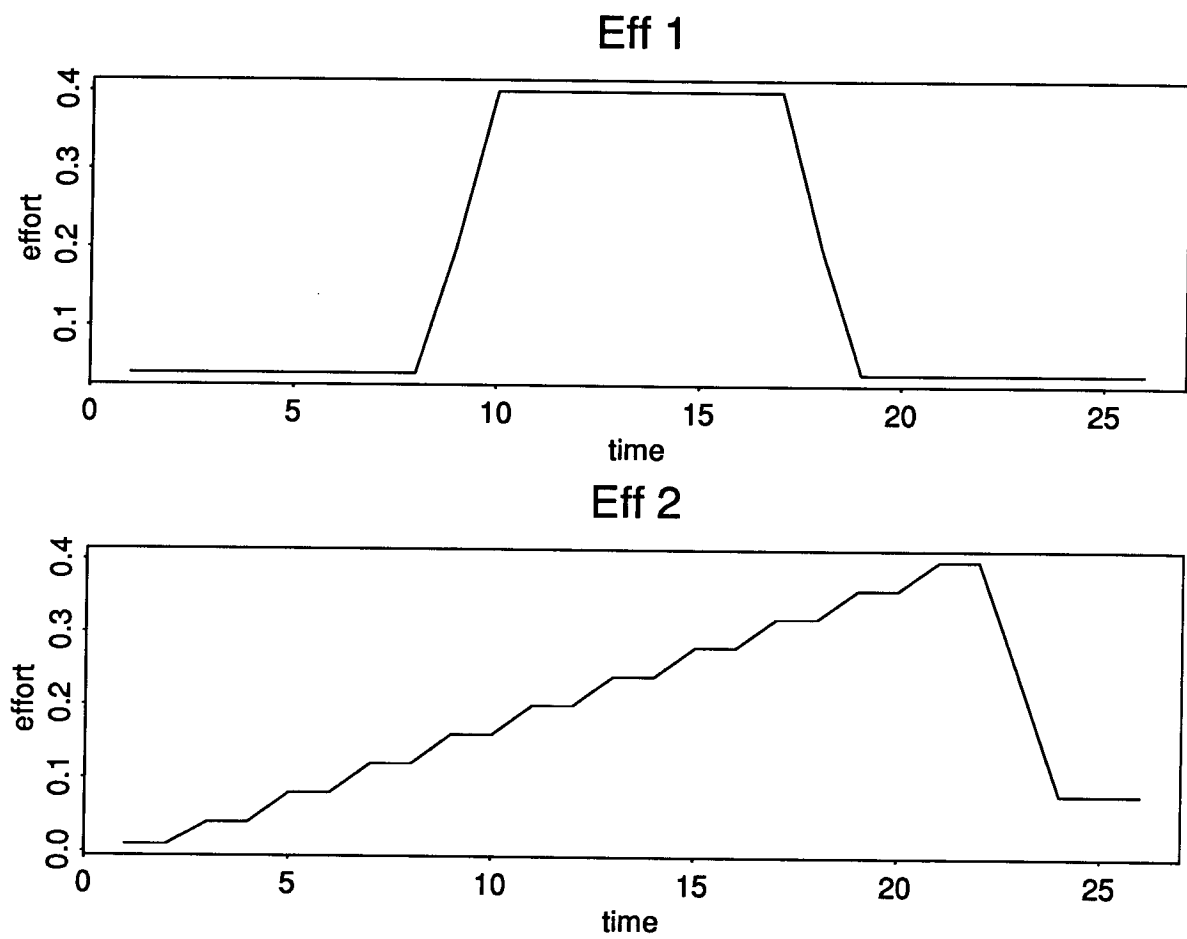
Figure 4.6: The two effort sequences used in the simulation models.
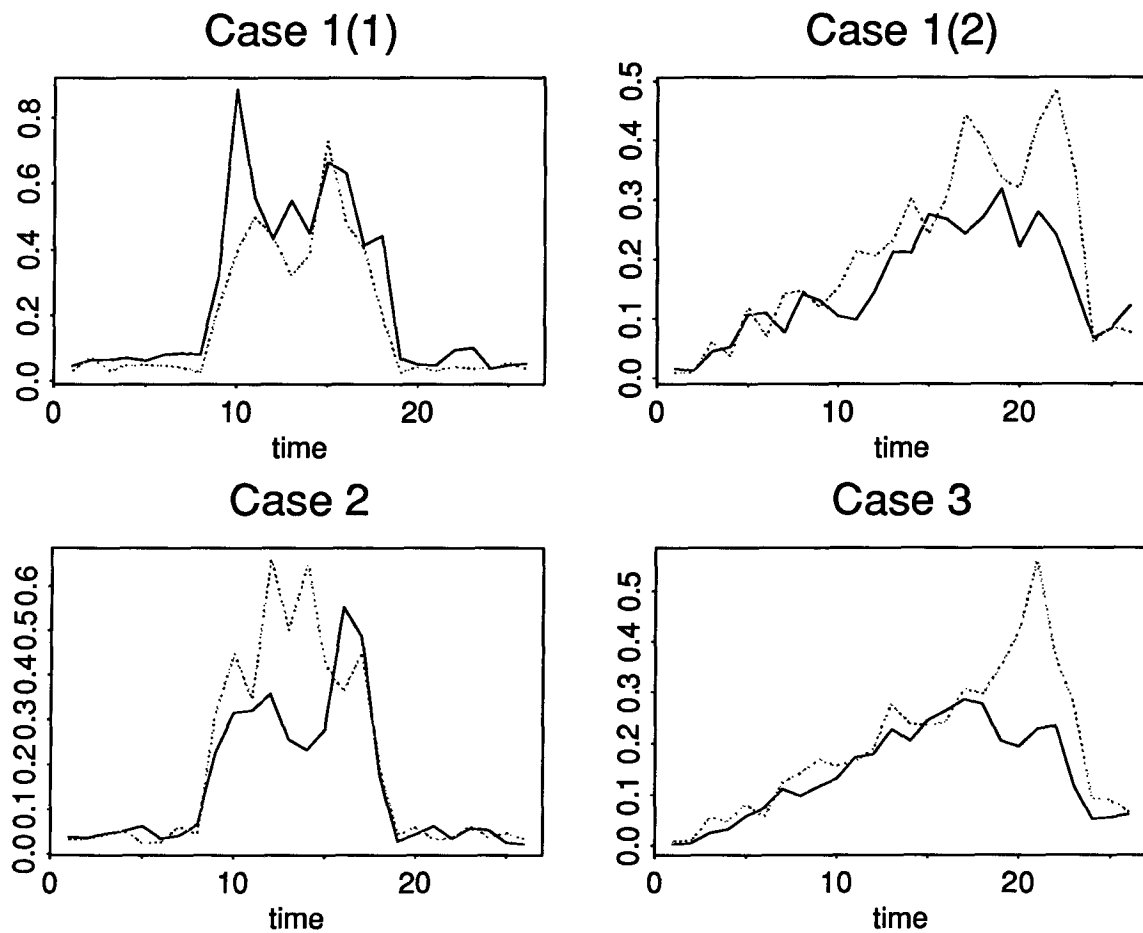
Figure 4.7: Time-series of observed efforts (dotted line) and the resulting simulated catches (full line), for different models.
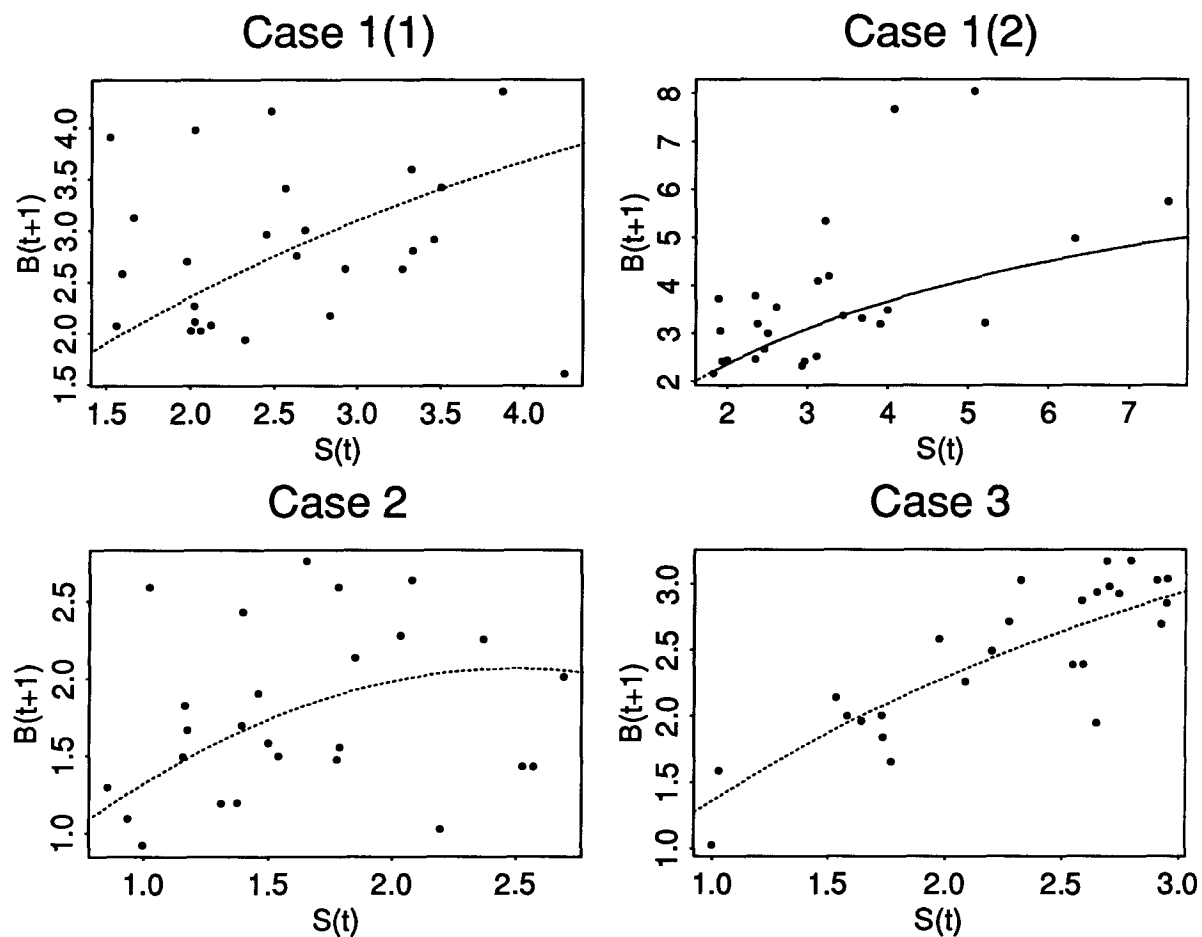
Figure 4.8: Spawning stock 'S(t)' and the corresponding biomass 'B(t+1)' for different simulation models. The line is the function $G(\cdot)$ from which the data were generated.

Table 4.9: Specification of 16 models used to fit catch and effort data. 'SIM' indicates the simulation model. 'FIT' gives the fitting model. 'Eff' specifies assumed effort sequence. $\hat{\lambda}$ is the assumed choice for λ. An omitted entrance implies the repetition from the line above.

Row No.	SIM	FIT	Eff	$\hat{\lambda}$
1	1	R1	2	1
2	-	-	2	0.5
3	-	-	1	0.5
4	-	-	2	0
5	-	R2	2	0.5
6	-	-	1	0.5
7	-	BH	2	1
8	-	-	1	0.5
9	2	R1	1	0.5
10	-	R2	-	0.5
11	-	-	-	0
12	-	BH	-	0.5
13	3	R1	2	0.5
14	-	R2	-	0.5
15	-	BH	-	0.5
16	-	-	-	0

Given the data, the approximate posterior $p(\theta \mid Y(25))$ for the parameter vector $\theta = (\ln a, \ln b, \ln q)'$ is obtained using the AIS procedure of Section 2.3 and the mixture model (2.3). I define the initial importance function $g_0(\cdot)$ to be the prior $\pi(\cdot)$. This will allow for the calculation of initial weights $\omega_{0,j}$ proportional to the marginal likelihood $p_o(Y(25) \mid \theta_{0,j})$. I draw an initial sample of size $n_0 = 1000$ points from this prior, and perform two updating recursions, using respectively $n_1 = 2000$ and $n_2 = n = 6000$. At each stage, the mixture model is clustered to a smaller mixture of $m = 800$, before proceeding with a new sample draw. The reported sample sizes were chosen empirically. Too small sample sizes, caused instability in measurements of relative entropies (2.4). Monte Carlo estimates for the mean and variance of θ and functions thereafter were calculated using the full sample of $n = 6000$ points. In all cases reported here, I used the assumption of Lemma 3.2.1(b).

As prior $\pi(\theta)$ for $\theta = (\theta_1, \theta_2, \theta_3)' = (\ln a, \ln b, \ln q)'$, I defined a three-dimensional Student distribution with 9 degrees of freedom, mean vector $m = (m_1, m2, m3)'$ and diagonal covariance matrix $C = \mathrm{diag}(C_1, C_2, C_3))$. To establish values for the moments m and C, I first determined quantiles for some biologically meaningful variables. To illustrate the procedure, I explain the derivation of m_2 and C_2 for $\theta_2 = \ln b$ in the standard Ricker model R1. According to this model, the surviving (or spawning) biomass producing maximum output, is given by $S_{\max} = 1/b$. As quantiles 5, 25, 50, 75 and 95 for $S_{\max}$, I used the values 1.1, 2.0, 3.1, 5.3, 10.0. These values were checked to approximately cohere with a log-normal distribution. Such a distribution would approximately have mode 2, median 3, mean 4 and standard deviation 3. Satisfied with these numbers, and using the transformation $\theta_2 = -\ln(S_{\max})$ at the given quantiles, I calculated the quantiles of the corresponding Gaussian distribution for θ_2. From each pair of quantiles I calculated m_2 and C_2 of the Gaussian distribution; the average of these values were my chosen parameters: $m_2 = -1.18$ and $C_2 = 0.46$. For the generalized Ricker model with power δ it holds that $b = 1/(\delta S_{\max}^\delta)$. If $b_{(2)}$ denotes the parameter b for model R2 (i.e., $\delta = 2$), then, by defining $\theta_{2(2)} = \ln b_{(2)}$, it is easy to see that

$$\theta_{2(2)} = 2\theta_2 - \ln 2$$

I used this linear relation to modify mean and variance of the parameter θ_2 accordingly. For none of the other parameters was a transformation necessary when going from R1 to R2. The remaining parameters were chosen using a similar procedure. The final values are:

R1 : $m = (0.57, -1.18, -1.18)'$ and $C = \text{diag}(0.11, 0.46, 0.46)$.

R2 : $m = (0.57, -3.05, -1.18)'$ and $C = \text{diag}(0.11, 1.84, 0.46)$.

BH : $m = (0.57, -2.14, -1.18)'$ and $C = \text{diag}(0.11, 0.30, 0.46)$.

Finally, as a prior for V, I defined an Inverse Gamma distribution with parameters $\alpha_0 = 2$ and $\beta_0 = 5$. To simplify comparisons, I used the same priors in all simulations.

Description of the Results

The entropies (2.4) of the different mixture models give a measure of the goodness-of-fit between the current importance function and the actual posterior densities at the given grid points; a value of 1 corresponding to a perfect fit. The observed entropies consistently approach 1 as the mixture model is updated. A summary of the observed entropies is given in Figure 4.9. It shows the substantial improvement at the first recursion, followed by the refinement of a later step.

The following results are summarized in Table 4.10. They are Monte Carlo estimates obtained from the sample of 6000 observations taken from the final mixture model. The effect of fixing $\hat{\lambda}$ can be verified for models R1, R2 and BH, when comparing the set of runs {1,2,4}, {10,11} and {15, 16}. Regardless of the model, there is an improvement in the estimates (i.e. reduction of bias and variance) as $\hat{\lambda}$ decreases. This result implies that, for the situations studied here, the best strategy is to assume all random disturbance as observation error. Similar findings were reported in Ludwig at al. [39].

The influence of different exploitation histories can be analyzed by comparing the outcomes for models differing only in their effort sequences. The pairs {2,3} for R1, {5,6} for R2 and {7,8} for BH, provide the output for such comparisons. In all those fits, I fixed $\hat{\lambda}$ arbitrarily at
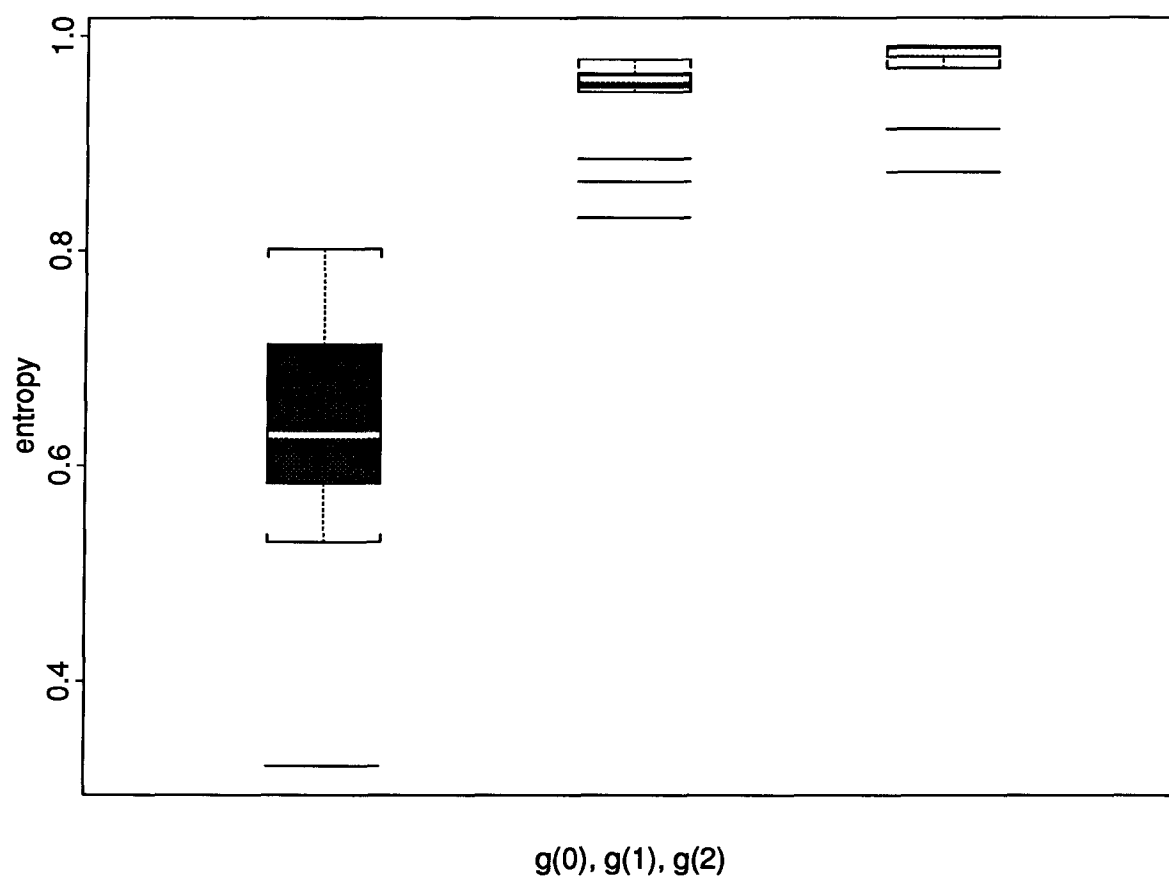
Figure 4.9: Box-plots for "relative entropy" for the sixteen fits to simulated catch and effort data.

Table 4.10: Posterior expectation (standard error) for the biases in S_{opt}, the bias in E_{opt}, and the global variance $\sigma^2 = V_p + V$.

Row No.	$S_{\text{opt}} - S^*_{\text{opt}}$	$E_{\text{opt}} - E^*_{\text{opt}}$	σ^2
1	0.43(1.01)	0.12(0.46)	0.23(0.07)
2	0.43(1.01)	0.12(0.46)	0.24(0.07)
3	0.36(0.77)	0.66(0.68)	0.15(0.04)
4	0.08(0.85)	0.01(0.40)	0.24(0.07)
5	0.87(1.60)	-0.19(0.34)	0.24(0.07)
6	1.10(1.15)	0.46(0.57)	0.15(0.04)
7	1.87(1.52)	1.05(1.10)	0.23(0.07)
8	1.48(1.40)	1.59(1.67)	0.16(0.05)
9	0.29(0.59)	0.61(0.70)	0.19(0.06)
10	0.84(0.84)	0.27(0.49)	0.22(0.06)
11	0.44(0.94)	0.07(0.48)	0.22(0.06)
12	1.24(1.15)	1.72(1.93)	0.21(0.06)
13	-0.21(0.44)	-0.15(0.57)	0.07(0.02)
14	0.16(0.69)	-0.31(0.47)	0.07(0.02)
15	0.30(0.77)	0.36(1.10)	0.07(0.02)
16	0.15(0.72)	0.30(1.03)	0.07(0.02)

0.5. The results for estimates of S_{opt} and E_{opt} oppose each other. In all cases, going from 'Eff 2' to 'Eff 1' reduced the variance for S_{opt} and increased bias and variance for E_{opt}. For S_{opt} the bias increased only in the case of model R2. The sequence 'Eff 1' definitively improved the estimates for the global variance σ^2. Under the simulation model described as 'case 1', only the data generated from 'Eff 1' (runs 3, 6 and 8) have expectations close to the real value 0.16, the general tendency being towards overestimation. Since 'Eff 1' would theoretically produce a more informative data sequence than 'Eff 2', the inconclusive results came as a surprise.

Another question to be answered: "How do different fitting models (R1, R2 or BH) perform under different circumstances (i.e. different simulation models) ?" This addresses the robustness of a given estimation model to deviations from the correct underlying biomass dynamics. Such relative comparisons are possible for the sets {2,5,7}, {3,6,8}, {9,10,12}, and {13,14,15}. Considering both, bias and variance, model R1 performs best for S_{opt}, while R2 is the better model for E_{opt}. It is important to notice that model BH failed, even in 'case 1' where BH was the "correct" model from which the data were generated. Similar findings in Ludwig et al. [39] and L&W, indicate that the generalized Ricker model is more effective in producing good estimates for policy parameters E_{opt} and S_{opt} than some of its competitors. This includes cases for which the data were effectively generated from such competing models.

As was already mentioned, the posterior estimates of variance tend to overestimate its true value. In the different situations investigated here, such overestimation is roughly one standard error. Exceptions are runs 3, 6 and 8 (already discussed), and runs 13 to 16 corresponding to simulations in 'case 3'.

In runs 7 and 8, the simulation and fitting models are the same. I have used 'run 8' to construct the posterior distribution for the parameter vector $(\ln a, \ln b, \ln q)'$. The two-dimensional marginals are shown in Figure 4.10 (a) to (c); the correct parameter combination used in the data generation (i.e. simulation model) is shown as a dot in each contour plot. In (d), I display a histogram of the 6000 estimates for E_{opt}; the true value is 0.54. The strong asymmetry of the histogram is a consequence of a few extremely large values (the maximum being $E_{opt} = 16.3$).
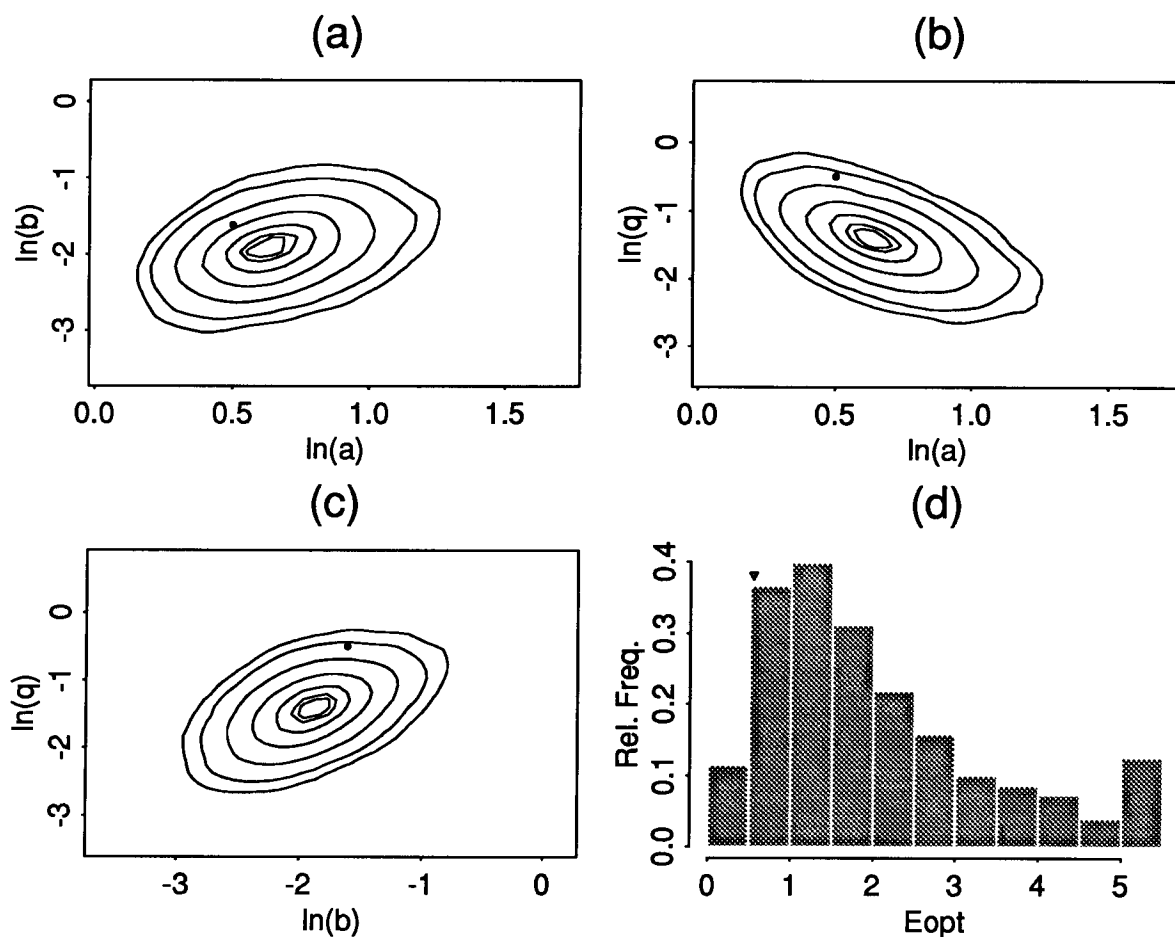
Figure 4.10: Contour plots of two-dimensional marginal posterior distributions for 'run 8'(lines at .05, .10, .25, .50, .75, .90, .95 of largest density). The dots indicate the true parameter used in the simulation model. In (d) a histogram of the 6000 Monte Carlo estimates for E_{opt} (the column to the far right includes all $E_{opt} > 5$).

Utilities and Bayesian Decision

So far I have concentrated my analysis on policy criteria of L&W in order to compare the Bayesian procedure and the "total least-squares" method used in their paper. But, in order to adequately explore the power of Bayesian decision making, I need a different approach. According to the description of Section 1.3, an optimal policy corresponding to a"Bayes rule" has to maximize the expected utility. To describe this approach, I use the utility function introduced in Section 3.2. It remains to define a set of decision rules A and the threshold SLIMIT.

For A, I specify a finite set of efforts, over a range — from 0.05 to 2.50 — which I assume implementable for the imaginary fishery under consideration. This set (50 points) is fixed as

$$A = \{0.05, 0.10, 0.15, \cdots, 2.45, 2.50\}$$

More precision could be achieved with a finer grid. But when it comes to enforcement of such policy, it might be of little use to distinguish between 0.68 and 0.71 — from a practical viewpoint 0.7 serves both cases.

The expected utilities for six simulations are displayed in Figure 4.11. Each row corresponds to a different simulation model. The first column (Runs 1, 3 and 5) relates to the fitting model R2; BH is the fitting model in the second column (Runs 2, 4 and 6). Table 4.11 also accounts for the models — notice that, in contrast to calculations of E_{opt}, no value of $\hat{\lambda}$ is needed — and gives the effort E_B which maximizes the expected utilities. In square brackets is the corresponding "true effort". It gives the best choice (in the set A) if complete information were available, the dynamics deterministic, and the above defined utility function were used as the maximizing criteria. All runs use SLIMIT = 0.3 — the threshold for surviving stock in the utility function.

The evaluation of R2 and BH is done by comparing E_B with the corresponding "true" policy. For simulation models 'case 1' and 'case 2' R2 performed best while in 'case 3' the correct estimate was given by BH. For model BH two features are observed: (1) the value
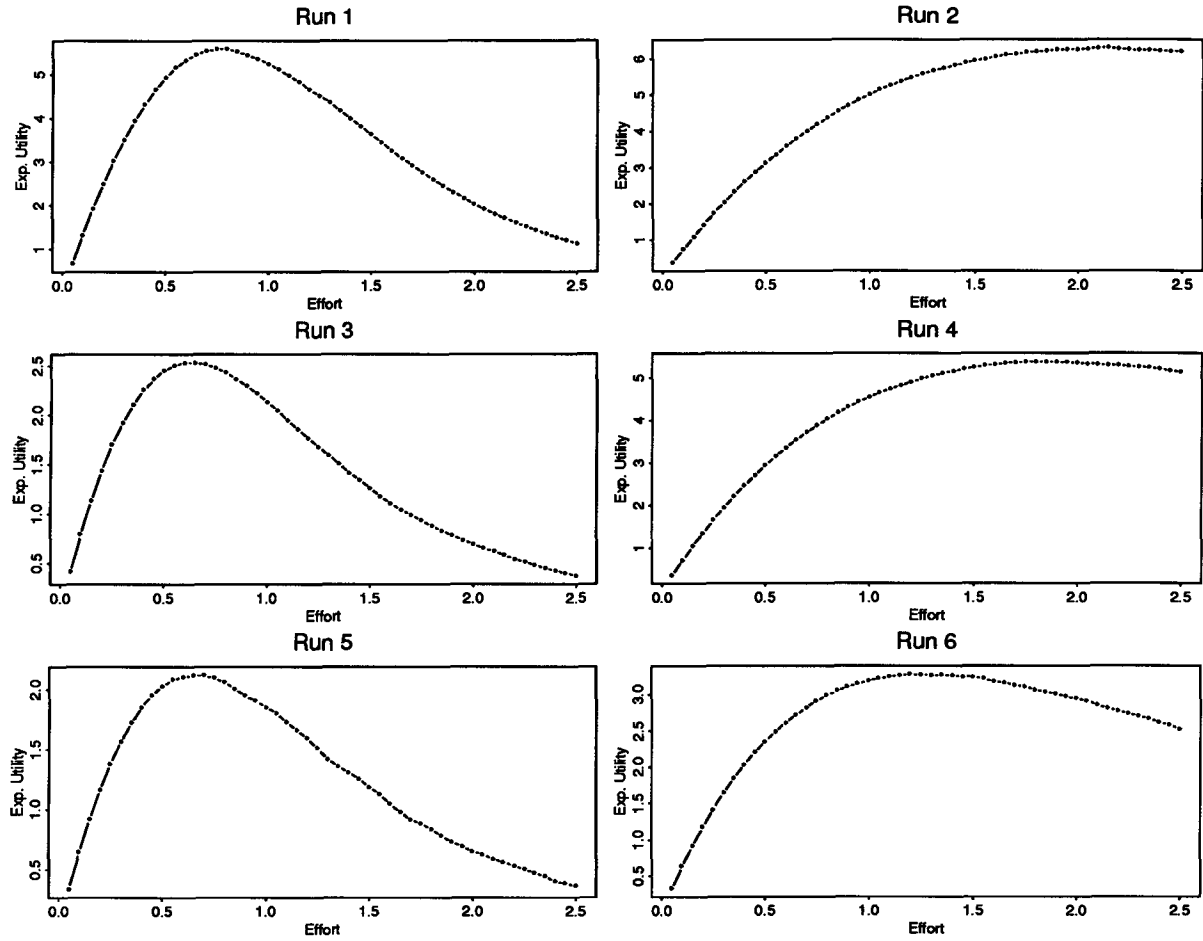
Figure 4.11: Expected posterior utilities over the set A of possible choices for effort. First column uses fitting model R2 and second column uses BH. Data from simulation models 'case 1', 'case 2' and 'case 3' were used in rows 1, 2 and 3.

Table 4.11: The Bayes rule E_B for different cases is displayed together with its "true" alternative [value in square brackets]. Columns 2 to 4 specify model assumptions. All 6 runs use SLIMIT=0.3.

No.	SIM	Eff	FIT	E_B
1	1	1	R2	0.75 [0.60]
2	1	2	BH	2.15 [0.60]
3	2	1	R2	0.60 [0.60]
4	2	1	BH	1.80 [0.60]
5	3	2	R2	0.70 [1.20]
6	3	2	BH	1.20 [1.20]

E_B is always larger than the corresponding value for R2; (2) the expected utilities (near the maximum) are very close over a wide range of effort selection — in run 2, for instance any value from effort 1.5 to 2.5 would do. Further notice that, similarly to the findings in the first part of this section, the "correct" fitting model of 'run 2' is outperformed by the "incorrect" R2 of 'run 1'.

I also experimented with different values for SLIMIT —not shown here. A sufficiently large increase of SLIMIT will tend to produce smaller values for E_B, reflecting a more conservative management. Finally, there is no effect of different exploitation histories (Eff 1 versus Eff 2) over the final outcomes in all cases I have analyzed.

Conclusion

Here is a summary of the results for this section. Two distinct analysis are considered. The first estimates the *optimal effort* E_{opt} as defined by L&W and the results are discussed in the light of earlier findings reported by them. The second analysis is new; it uses an utility function for which the best effort is the *Bayes rule* E_B, chosen to maximize the expected posterior utility.

First Analysis: Given the goal of estimating the *optimal effort*, the three central findings agree with L&W. I summarize them as follows:

1. When deciding on how to "split" the overall variance into process noise and observation error, it is preferable to favour the observation error. In particular, observation error only ($\hat{\lambda} = 0$), was shown to outperform the assumption of process noise only ($\hat{\lambda} = 1$).

2. An "incorrect" fitting model can lead to superior estimates of policy variables. For catch and effort data with large error and/or low variation in stock size, one might give up fidelity of the fitting model to the underlying dynamics in order to obtain improved estimates for the policy parameter of interest. In particular, the Ricker models (R1 and R2) have outperformed the Beverton-Holt (BH) alternative.

3. Different exploitation histories (here represented by distinct effort sequences) can affect policy parameters in different ways. Overall, there is a trade-off between bias and variance. For the cases analyzed here, there is no single model which is superior for both parameters E_{opt} and S_{opt} under all circumstances.

Second Analysis: Here I use Bayesian decision-making. The utility function incorporates the central goals that a fishery manager would like to achieve: maximization of catches and reduction in the risk of stock collapse. In order to determine the best solution, there is no need for equilibrium assumptions, neither is it necessary to specify a parameter $\hat{\lambda}$ dealing with process noise and observation error. The results show that:

1. The "incorrect" fitting model performed better than the correct model. This feature was also observed in the previous analysis.

2. For all three simulation models, at least one of the fits identified the "true" (perfect information) effort. It seems that the generalized Ricker model (particularly R2) does better than model BH while always staying on the conservative side (smaller efforts). Only in 'case 3', for which the effort has less effect over stock depletion[3], did BH give the correct effort.

4.4 Surplus Production: Orange Roughy

This section is a case-study of the Orange Roughy (*Hoplostethus atlanticus*) fishery on the Chatham Rise, New Zealand. The analysis uses data previously published by Francis [18]. The risk analysis performed by Francis in that paper has caused some recent controversy (Hilborn et al. [24]) regarding the state of the fishery and consequently the management actions to be taken. Bayesian analysis which is made possible by the procedures developed in Chapters 2 and 3, can help to clarify some of the controversy.

Using an age-structured model with many parameters (which were estimated elsewhere and

---

[3]The catchability $q = .31$ is half the value used in cases 1 and 2.

are assumed known), Francis built a cumulative distribution function for the virgin (unfished) biomass B_1. This construction was done by replicating simulations 100 times and using a frequentist approach to make empirical estimates of cumulative probabilities. He proceeded by fitting a Johnson's distribution (Johnson and Kotz [26] p.23) to the simulated cumulative distribution. Taking the first derivative of this fit, he obtained a probability density for B_1. The mode of this distribution was used as a maximum likelihood estimate of B_1. His goal was the assessment of the risk of collapse within the next five years, given different rates in the reduction of annual *total allowable catches* (TAC's) within that period. I shall give further details about these different decision options later.

Hilborn et al. [24] performed a series of Monte Carlo experiments for scenarios in which Francis' method of constructing probability distributions for B_1 fails: cases with low harvest rate and uninformative data. Based on these results, they question the estimate for B_1 produced by Francis.

A key element in the controversy is the uncertainty about the virgin biomass B_1 and its use in decision making. Francis describes uncertainty by a density for B_1, from which he then draws random values of B_1 to produce replicates for a frequentist construction of his risk function. Further uncertainties are ignored in the process. I shall use the Bayesian alternative described as 'Model 2' (Section 3.3) to construct a posterior distribution for the parameter vector θ which includes $\ln B_1$ as one of its elements. This multi-dimensional posterior distribution is further used in the context of Bayesian Decision Theory to determine an optimal action.

The data of Francis are reproduced in Table 4.12. They include a time series of catches $\{C_t; t = 1, \cdots, 11\}$ and estimates of abundance from trawl surveys $\{I_t; t = 6, \cdots, 11\}$. According to Francis' assumptions, the errors in C_t are negligible, so that all observation error is related to I_t.

My general model, is given by (3.24) and (3.25). Particularly, I define

$$\theta = (\ln a, -\ln B_1, \ln q)'$$

$$b = 1/B_1$$

Table 4.12: Estimated catch history, and trawl survey biomass for Orange Roughy (*Hoplostethus atlanticus*) for the period 78/79 to 88/90. Data from Francis [18].

Time	Catch (t)	Trawl Index
1	15,340	–
2	40,430	–
3	36,660	–
4	32,354	–
5	20,064	–
6	32,263	164,835
7	38,142	149,425
8	39,098	102,975
9	39,896	80,397
10	31,478	97,108
11	42,621	66,291

$$S_t = B_t - C_t$$

$$G(\theta, B_{t-1}) = S_{t-1} + aB_{t-1}(1 - bB_{t-1}) \quad \text{for } t = 2, \cdots, 11$$

$$F(\theta, B_t) = qB_t \quad \text{for } t = 6, \cdots, 11$$

Model $G(\cdot)$ describing biomass dynamics, is the discrete form (i.e. difference equation) of the Schaeffer model. The observation model $F(\cdot)$ assumes direct proportionality between index I and the corresponding true biomass B, with constant of proportionality q.

Determining Priors

I used two different priors for θ. Both are three-dimensional Student distributions, with 9 degrees of freedom. They represent my interpretation of the information provided and general assumptions used by Francis. The first prior, denoted "informative", assumes a relatively high confidence in prior identification of the mean values of the parameters. It is based on a guess for mean as well as a range to represent possible smallest and largest values. For the second prior, denoted "diffuse", I made a more careful (and pessimistic) elicitation. Starting with guesses for possible values of five pre-specified quantiles — 5, 25, 50, 75 and 95 — I found the parameters

of an idealized log-normal distribution which would best agree to those values (If these numbers were not coherent with such a distribution, I tried to improve with a new set of values). The parameters corresponding to the underlying Gaussian distribution were used as parameters in the Student prior. I also assumed a diagonal prior covariance.

For the parameter q I used as quantiles the values 0.20, 0.30, 0.41, 0.56, and 0.90 — these values more than cover the range of possibilities considered by Francis. Similarly, for B_1 I used the values (in thousands) 200, 400, 650, 1000, and 2000 and for a the values 0.04, 0.10, 0.19, 0.49, and 0.95.

The prior parameters for the Inverse Gamma distribution of the observation variance V were the same in all runs. They were based on Francis' assumption of an coefficient of variation of 0.20 for observation errors in the abundance index. By fixing $\alpha_0 = 2$ and setting the mode for V at 0.40, the value $\beta_0 = 8.0$ was calculated. The values for the prior were as follows:

- $V \sim IG(\alpha_0 = 2.0, \beta_0 = 8.0)$

- Informative: $\pi_0(\theta) \sim T_3(9, m_0, C_0)$ with $m_0 = (-1, 61, -13.36, -0.66)'$ and $C_0 = \text{diag}(0.30, 0.20, 0.05)$

- Diffuse: $\pi_0(\theta) \sim T_3(9, m_0, C_0)$ with $m_0 = (-1, 62, -13.37, -0.88)'$ and $C_0 = \text{diag}(0.97, 0.48, 0.21)$

Results

Table 4.13 summarizes model specifications (columns 2 and 3) and the results for different runs. Various bandwidths between $h^2 = 1$ and $h^2 = .6$ were used. I wanted to make sure not to smooth out potential "ill behaved" features, such as multimodality, by using too large a bandwidth. Variations in estimates were small within the given range for h^2. For the results presented here, I used $h^2 = 0.64$ and $h^2 = 0.90$.

In the AIS procedure (Section 2.3) I used three updating steps with sample sizes $n_0 = 1000$, $n_1 = 2000$, $n_2 = 2500$, and $n_3 = n = 6000$. At each stage the mixture model was clustered

Table 4.13: Model specifications (columns 2 and 3), final entropy (Ent.) and posterior estimates for virgin biomass B_1 for Orange Roughy (values in 1000 t) and the probability that B_1 exceeds 500 thousand t. Values in brackets are estimates for posterior standard error. Run '1a' uses assumption (a) of Lemma 3.1.; the remaining runs use assumption (b). 'D' denotes *diffuse* prior and 'I' the *informative* alternative.

Run	h^2	Prior	Ent.	B_1	$P(B_1 > 500)$
1a	0.64	D	0.94	450 (108)	0.25 (.006)
2	0.90	D	0.98	436 (92)	0.19 (.005)
3	0.64	D	0.98	431 (90)	0.18 (.005)
4	0.64	I	0.99	385 (48)	0.02 (.002)
5	0.64	I	0.99	384 (49)	0.02 (.002)

to $m = 800$ components of three-variate Student distributions with 9 degrees of freedom before proceeding. The relative entropy (2.4) follows the usual pattern described in previous sections. A typical outcome (Run 5 of Table 4.13) gives the values 0.597, 0.947, 0.987, and 0.991 for the importance functions $g_i(\cdot)$ for $i = 0, \cdots, 3$, respectively.

The prior clearly influences the final estimates. If diffuse, the Monte Carlo estimates of posterior mean and standard error are higher (approx. 433,000 t and 90,000 t), than estimates obtained under the informative alternative (approx. 384,000 t and 48,000 t). Similarly, the estimate of the probability that B_1 exceeds 500,000 t, is about 0.2 for the diffuse prior, while in the informative case the estimate is 10 times smaller. The marked differences due to different priors are also shown in Figure 4.12. Three different marginal posteriors are displayed: for 'run 1a' (line), 'run 2' (dots) and 'run 4' (broken line). If the diffuse prior is accepted, values of B_1 up to 800,000 t are possible. This is quite different from the informative case, where values of B_1 exceeding 600,000 t are unlikely. A comparison with the estimated density reported by Francis (his Figure 1(B)), suggests the density which was obtained for 'run 4' (broken line). In other words: the acceptance of Francis' distribution corresponds to the acceptance of the informative prior in my model.

The marginal distribution for 'run 1a' (line) shows an interesting feature suggesting bimodality. This was an atypical feature. It is included as an illustration of a particularly interesting
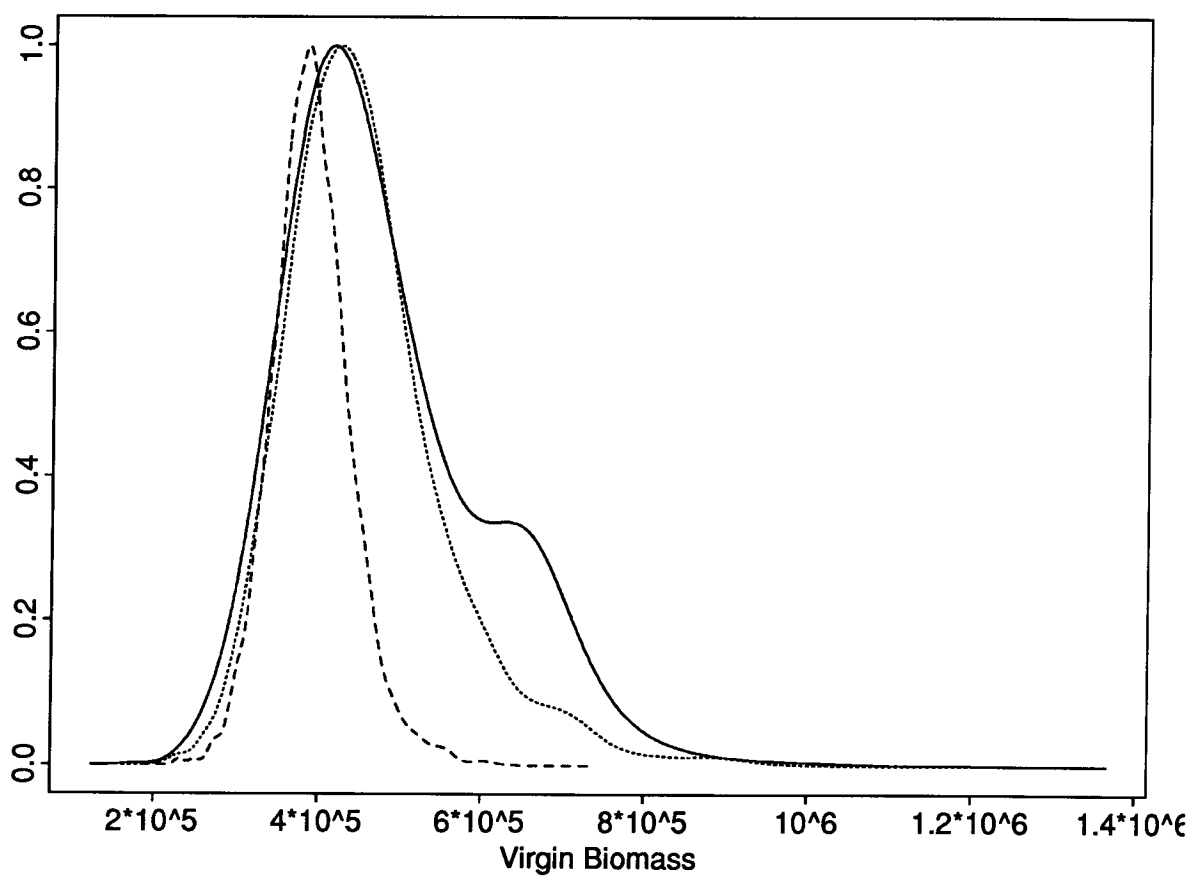
Figure 4.12: Marginal posterior distributions for virgin biomass B_1 for different runs: '1a'(line), '2'(dots) and '4' (broken line).
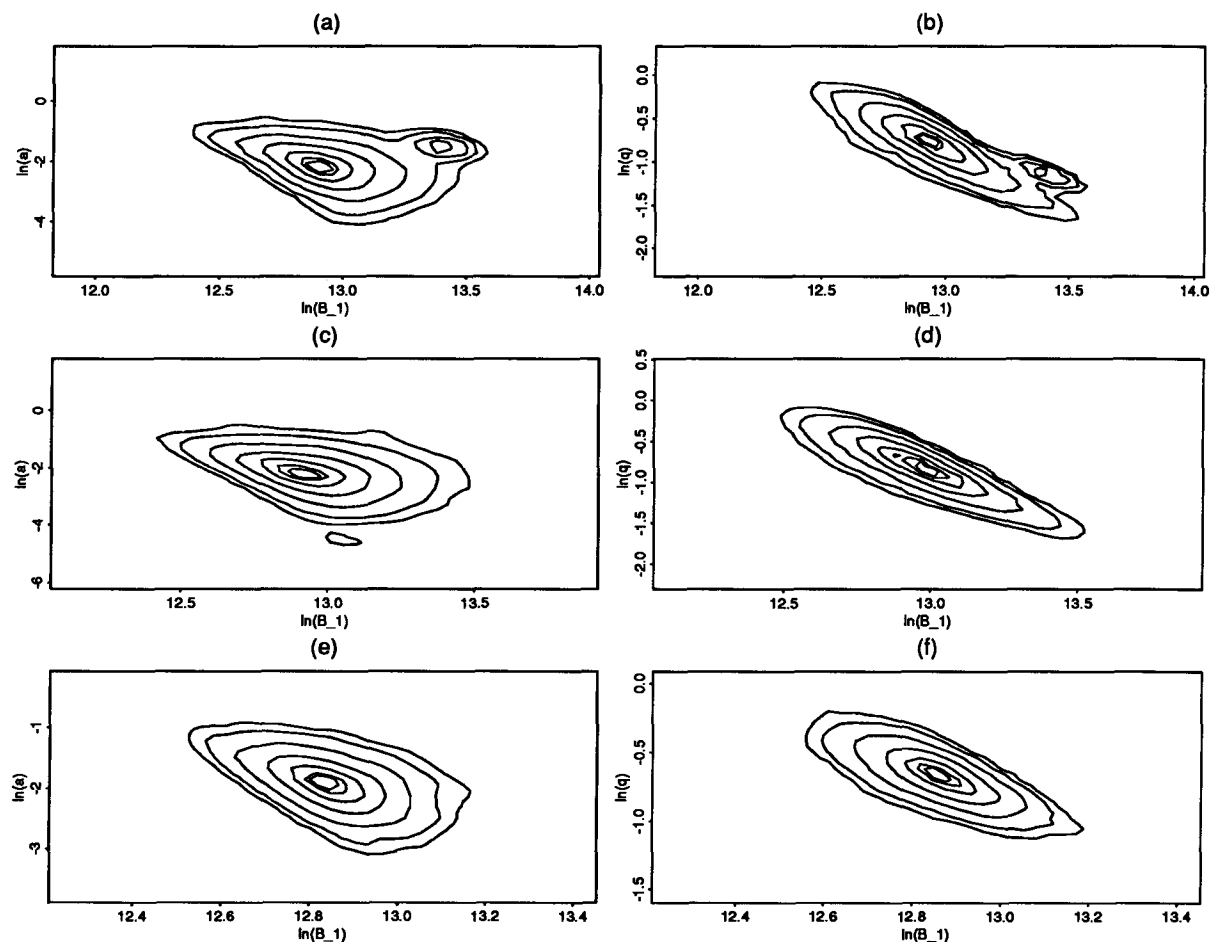
Figure 4.13: Contour plots for two-dimensional marginal posterior distributions: 'run 1a':(a) and (b); 'run 2':(c) and (d); 'run 4':(e) and (f). Contour lines are drawn at .05, .10, .25, .50, .75, .90 and .95 of the highest density.

possibility. To better see this feature, I included contour plots of the two-dimensional marginal posterior distributions for $(\ln B_1, \ln a)$ and $(\ln B_1, \ln q)$ in Figure 4.13 for the same three cases. For 'run 1a' (displayed in (a) and (b)), the two modes are clearly shown. All remaining cases are unimodal. The negative correlation between B_1 and q is consistently present (contours (b), (d), and (e)).

Decision-making

The analysis so far was able to describe some general features of marginal uni- and bi-dimensional posterior distributions and its relation to prior specification. But the final goal of Francis' analysis was the use of his risk-assessment to make management decisions. A key element in Hilborn et. al.'s criticism was the absence of a coherent framework for decision-making under uncertainty. I shall use the three-dimensional posterior distribution for θ together with an appropriate utility function to perform a Bayesian decision analysis and choose the *Bayes rule* policy from among the five different options considered by Francis. But first I shall describe these management options.

The fishery for Orange Roughy is managed by annually setting the TAC. For instance, the fishing season of 89/90 (the year immediately following the data sequence of Table 4.12), had its TAC fixed at 28,637 t. According to Francis, there seems to be an agreement to reduce TAC. But two questions remain: (1)"to what *level* should the TAC be reduced" and (2)"at what *rate* should the reduction take place".

The target level was fixed at a *maximum constant yield* of 7,500 t — accounting for question (1). Francis defined this value to be $\frac{2}{3}$MSY, where MSY denotes the deterministic *maximum sustainable yield*. This value is questionable on two grounds: (i) it is an arbitrary choice and (ii) its calculation uses the estimate $B_1 = 411,000 \, t$, ignoring the uncertainties of this estimate.

What I will try to answer here is question (2): the rate at which such reduction should be made. Table 5 of Francis displays five alternative options. I will use the same options to define my set of actions $A = \{a_1, a_2, a_3, a_4, a_5\}$. Given action a_1, the rate of TAC reduction is 3 thousand tons per year. Similarly for a_2 up to a_5, the rates are respectively 5, 7, 9 and 12 thousand tons.

Francis considers "the risk to the fisheries" which each of these options represents by estimating (using a frequentist empirical estimate) "the risk that the fishery would collapse within a five year period". The fishery is said to have collapsed if, within this period, in some year the TAC is greater than 2/3 of the corresponding biomass.

To construct my utility function I assume that, for a fixed a_i and in any given year, the catch will equal the prescribed TAC; I denote this total quota by $T^{(i)}$. Hence, starting from the current $T_0 = 28,637$, future quotas for $s = 1, \cdots, 5$ are given by

$$T_s^{(i)} = \max\{T_0 - sa_i; \ 7500\}$$

Defining the *discount rate* of future catches by δ_{PV}, the *cumulative discounted catch* over the five year period is denoted $CC^{(i)}$ and given by

$$CC^{(i)} = \sum_{s=1}^{5} \delta^s T_s^{(i)}$$

I further need to identify the "occurrence of collapse given decision a_i". This is done by the indicator function

$$A^{(i)} = \begin{cases} 1 & \text{if } T_s^{(i)} > \frac{2}{3} B_{12+s} \text{ for some } s = 0, \cdots, 5 \\ 0 & \text{otherwise} \end{cases} \tag{4.37}$$

Given a pair $(\theta_j, a_i) \in \Theta \times A$ (recall that Θ is the random sample from the posterior distribution), the utility function can be stated in terms of the notation just introduced:

$$U(\theta_j, a_i) = (1 - A^{(i)})\phi(CC^{(i)}) \tag{4.38}$$

where ϕ is some known non-decreasing function.

The utility (4.38) is zero whenever collapse is predicted ($A^{(i)} = 1$). Otherwise it is $\phi(CC^{(i)})$. The dependence on the posterior $p(\theta \mid \text{Data})$ becomes clear when considering the algorithm used to calculate (4.38):

1. Fix $\theta_j \in \Theta$ and use $G(\cdot)$ together with the observed catch data $\{C_t; t = 1, \cdots, 11\}$ to predict the average time series for $\{B_t; t = 2, \cdots, 12\}$.

2. Fix a decision rule $a_i \in A$ and define $\{T_0, T_1^{(i)}, \cdots, T_5^{(i)}\}$ to be the catches into the future.

3. Evaluate $A^{(i)}$ according to (4.37).

4. Calculate $U(\theta_j, a_i)$ using (4.38).

Table 4.14: Risk-averse utilities ϕ associate to the five possible *cumulative discounted catch* $CC^{(i)}$.

i	a_i	$CC^{(i)} \times 10^{-3}$	$\phi(CC^{(i)})$
1	3	83	.981
2	5	63	.950
3	7	52	.916
4	9	46	.888
5	12	40	.850

5. Repeat the whole procedure for the next pair (θ_j, a_i).

I consider 3 alternatives for ϕ:

1. $\phi(x) = x$.

2. $\phi(x) = 1$.

3. $\phi(x)$ is a *risk-averse* function defined in Table 4.14.

The first assumes $\phi(x) = x$; so that the utility is linear in CC whenever the possibility for collapse is excluded. A more conservative approach, which is concerned *only* with the risk of collapse defines $\phi(x) = 1$. This second alternative assumes the same utility for different catches regardless of its size; an unrealistic assumption. A compromise between both extremes can be obtained by a concave function for ϕ and is given as the third option.

Some comments on Table 4.14. Since the values $CC^{(i)}$ are fixed for each i, I only need to define ϕ for those particular points. Notice that the utilities become smaller with smaller values of CC, but the reduction is non-linear. This values were determined assuming $\phi(160) = 1$ and using Table A.1 for constant risk-aversion in Lindley [31] (p.196). For comparison, the linear (or risk-neutral) values would be .52, .39, .33, .29, .25 for $i = 1$ to 5, respectively.

For the manager who needs a practical answer, the results in Table 4.15 are the most important of the section — with complementing details on Table 4.16. The following features are highlighted there:

- The decisions are robust over different bandwidths [1a,2] and prior assumptions ([1a,5] and [3,4]).

- What defines the optimal rate is the utility function. If the *linear* relation $\phi(x) = x$ is chosen, the optimal decision is a_1 which recommends a reduction in TAC of 3000 t per year over the next 5 years. On the opposite end, the conservative '0—1' option ($\phi(x) = 1$) recommends a_5, a drastic rate of annual reduction by 12000 t. The intermediate *risk-averse* utility recommends rates in between: it chooses a_3, a reduction rate of 7000 t per year, as optimal.

- The expected utilities in the '0—1' case, have a special interpretation: the complement of the probability of collapse. For any decision a_i, the probability of collapse p_i, is estimated as

$$p_i = 1 - E[U(\theta, a_i)]$$

If, for instance, I take 'run 5' in Table 4.16, these estimated are .27, .13, .08, .06, .05 for actions a_1 to a_5, respectively (these probabilities change little for diffuse prior:'run 2'). Such values are in disagreement with the results presented in Francis' Figure 3: his estimates of the risk of collapse exceed .6 for a_1, while probabilities smaller than .2 are observed only for a_4 and a_5. Since both methods are derived under different paradigms, a direct comparison is not possible. However, Francis' model is complex (many parameters) and the estimates of risk are sensitive to recruitment assumptions (as indicated in his Figure 3(d)). My model is comparatively simple and does not need special assumptions for recruitment. The findings in Ludwig and Walters [37] [38] indicate that simple models are likely to be more robust for management purposes.

The results suggest that, for practical decision-making, the focus of discussions should be the assessment of the most adequate utility function to which all parts can agree (industry, agency, fisher, environmentalist, etc.). Given the available data, details regarding technicalities in parameter estimation and modelling seem to be of minor consequence.

Table 4.15: Recommended decisions a_i (rate of TAC reduction in thousand t/year), which maximize the expected posterior utilities, using three alternative utility functions

Run	Linear	Risk-Averse	0—1
1a	-	7	12
2	-	-	12
3	3	7	-
4	3	7	-
5	-	7	12

Table 4.16: Expected utilities obtained for all 5 decision rules, according to the form of the utility function (expressed by ϕ).

ϕ	Run	a_1	a_2	a_3	a_4	a_5
Linear	3	.622	.538	.462	.420	.373
	4	.609	.540	.470	.427	.378
Risk-Averse	1a	.740	.814	.823	.814	.793
	3	.733	.811	.824	.817	.796
	4	.717	.825	.838	.830	.808
	5	.716	.820	.837	.828	.810
0—1	1a	.767	.871	.913	.931	.945
	2	.753	.862	.906	.922	.938
	5	.730	.870	.918	.934	.951

Summary

There are 4 conclusions from this case-study for Orange Roughy:

1. The simple production of probabilities of collapse, as done by Francis, is not enough for decision making. The uncertainties (given by probabilities), need to be linked to measures of the consequences (utilities) resulting from different actions. The Bayesian paradigm further says that, for rational and coherent decision making, the maximization of expected utilities is the correct criterion to use. Such analysis was carried out with three distinct utility functions. By taking this additional step, my analysis completes the work of Francis.

2. The shape of the marginal posterior distributions for B_1 is affected by the choice of a prior. I used two alternative priors: informative and diffuse. The distribution used by Francis corresponds to the informative prior. The diffuse prior resulted in a heavier upper tail, increasing the possibility of values $B_1 > 500,000t$ from 5% to about 20%.

3. Decision making is robust with respect to choices for prior. Given any of the utility functions, the optimal decision was not affected by the particular choice of prior (informative or diffuse). This indicates that for the current model/data situation, and assuming consensus about the set of actions A, the tail behaviour in the distribution for B_1 was not strong enough to change the overall optimal decision.

4. The construction of a sensible utility function is an important tool in the negotiating process among all interest groups. My results suggests that this aspect of the analysis overshadows technicalities concerning the choice of a model for biomass dynamics.

4.5 Multimodality: Simulation Model

The advantage of replacing classical parameter estimation by the Bayesian alternative is most apparent when the posterior distribution is multimodal. It indicates that more than one model

describes data reasonably well. Such information is an important tool in the design of adaptive management policies (Walters [68]), where one of the requirements is the identification of alternative hypothesis. A procedure that forces the choice upon a single best model, as the classical approach does (least-squares, maximum likelihood), is unable to capture the uncertainty present in the problem. Such ideas were the motivation to examine practical situations for which multimodality might emerge.

In this section and the next, I give examples for which the posterior distribution is multimodal. Here I use simulations to produce artificial data. The next section is a case-study of Pacific cod. I employ two different approaches:

1. In the simulated example I use an observation mixture-model (MM) given as:

$$(Y_t \mid \theta, V) \sim r \cdot N(\mu_t^{(1)}(\theta), V) + (1 - r) \cdot N(\mu_t^{(2)}(\theta), V) \text{ with } r \in (0, 1) \qquad (4.39)$$

where $\mu_t^{(1)} \neq \mu_t^{(2)}$ for all t. Such model might be adequate in situations where available knowledge suggests a mixture of two possible hypothesis (or stocks): a very productive stock with low virgin biomass, against a less productive stock with large virgin biomass.

2. The approach of the next section maintains the original observation model (without mixture) but modifies the description of biomass dynamics by replacing the *surplus-production* model by a more elaborate *delay-difference* (DD) model (see Section 3.4).

The observation mixture model (MM)

For data generation I used a standard logistic (Schaeffer) model for biomass dynamics. The simulation model is as follows:

$$B_t = S_{t-1} + aB_{t-1}(1 - bB_{t-1})$$

$$S_t = B_t - C_t$$

$$C_t = B_t(1 - e^{-Qe_t})$$

$$E_t^o = E_t e^{v_t}, \quad \text{where } v_t \text{ are i.i.d. } N(0, \sigma_e^2 = 0.05)$$

$$I_t = C_t / E_t^o$$

where B_t is biomass, C_t the catch, E_t the effective effort, E_t^o the observed effort and I_t the observed abundance index. Parameters in the model are $a = 0.2$, $B_1 = 1/b = 4000$ and catchability Q. The catchability Q is modelled as a random variable from a mixture of uniform densities, namely

$$Q \sim \epsilon U(.20, .40) + (1 - \epsilon) U(.55, .65)$$

with $\epsilon = 0.4$. The real (but unknown) effort sequence is defined for $t = 1, \cdots, 26$ to be

$$E_t = \begin{cases} 0.40 & \text{if } 9 < t < 18 \\ 0.20 & \text{if } t = 9 \text{ and } t = 18 \\ 0.04 & \text{otherwise} \end{cases}$$

Figure 4.14 shows the time series of C_t, E_t^o and I_t for the 26 years of data. The last plot displays the abundance index against the corresponding observed effort for the last 13 years. For fitting purposes I assume that only these 13 abundance observations are known along with all 26 observations for catch.

The observation model for $Y_t = \log I_t$ uses MM defined in (4.39) with $r = 0.5$. I further define $\mu_t^{(i)}(\theta)$ as

$$\mu_t^{(i)}(\theta) = \ln(q^{(i)} B_t^{(i)}) \text{ for } i = 1, 2$$

where for $i = 1$ I use $a^{(1)}$, $b^{(1)}$, $q^{(1)}$ according to the relation

$$(\ln a^{(1)}, \ln b^{(1)}, \ln q^{(1)}) = (\theta_0, \theta_1, \theta_2)$$

while for $i = 2$, I take

$$(a^{(2)}, b^{(2)}, q^{(2)}) = \left(\frac{a^{(1)}}{2}, \frac{b^{(1)}}{2}, 2 \cdot q^{(1)}\right)$$

In the above construction I am mixing the hypothesis ($i = 1$) of high productivity, low virgin biomass, and a given catchability, with an alternative hypothesis ($i = 2$) of low productivity, large virgin biomass, and high catchability.
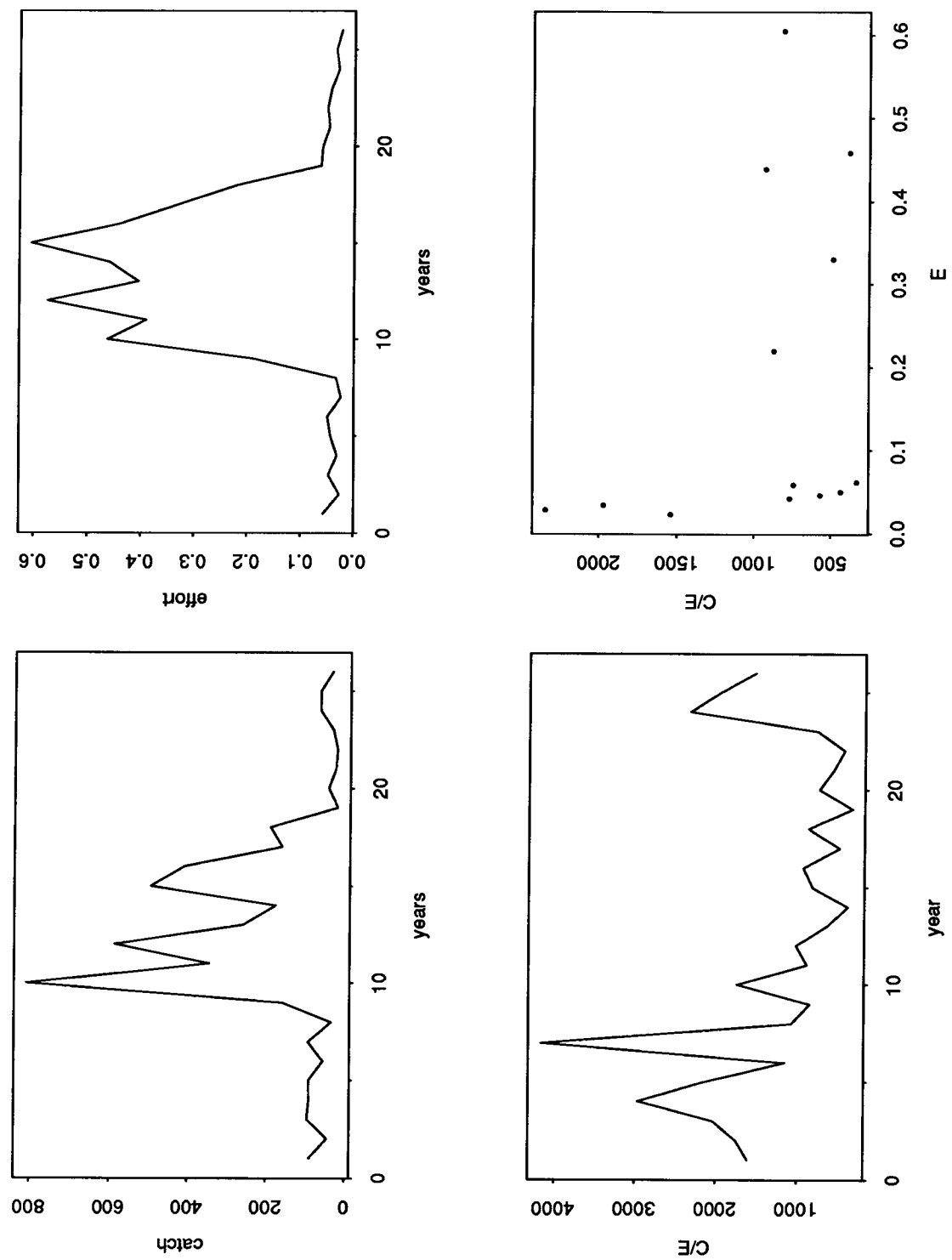
Figure 4.14: Time series for Catch (C_t), observed effort (E_t) and CPUE (I_t). Also a plot of 'CPUE vs. E' for the final 13 years

The posterior distribution for the parameter vector θ is obtained by using the algorithm of Section 3.3. What changes is the use of (4.39) as the observation model.

Some two- and three-dimensional[4] marginal posteriors are displayed in Figure 4.15. The perturbation caused by the stochastic behaviour of the catchability Q is translated into uncertainties regarding the virgin biomass. A secondary mode is estimated for a biomass near 1600 while the primary mode is estimated near a biomass of 3200. Also notice the negative correlation between $\ln a$ and $\ln B_1$ which reflects the confounding between a small productive stock versus a larger less productive one.

If both modes are interpreted as summarizing conflicting hypothesis, the highest one indicates the hypothesis with the strongest support. In this example, the large stock (virgin biomass near 3200) with a smaller intrinsic growth rate (parameter a) is the most likely of both alternatives. This hypothesis is closest to the true virgin biomass of 4000.

4.6 Multimodality: Pacific Cod

One might argue that the mixture model (4.39) used in Section 4.5 is artificial. Therefore, I investigate the emergence of a multimodal posterior in a more natural way, without using a specific mixture model for observations. Possibly *surplus-production models* are too simplistic[5].

A next step towards more flexible models, capable of incorporating characteristics of growth, recruitment, and time-delays, is the class of *delay-difference* (DD) models. By establishing assumptions for growth, survival and recruitment, these models can better depict the interactions of these elements and the resulting time-delays. For more details on the DD model and the notation used throughout, see Section 3.4.

Here I shall present a case-study using the data for Pacific cod (*Gadus macrocephalus*) presented in Hilborn and Walters [25] (p.345). For the purpose of the analysis, parameters for growth and natural survival are assumed known (Walters pers. comm.) and given in Table 4.17.

The observed data are two time-series: catches $\{C_1, \cdots, C_\tau\}$ and abundance index

---

[4]For three-dimensional plots I used 'persp()' function in S-Plus.

[5]But recall that in the case-study for Orange Roughy one of the fits also suggested bimodality.
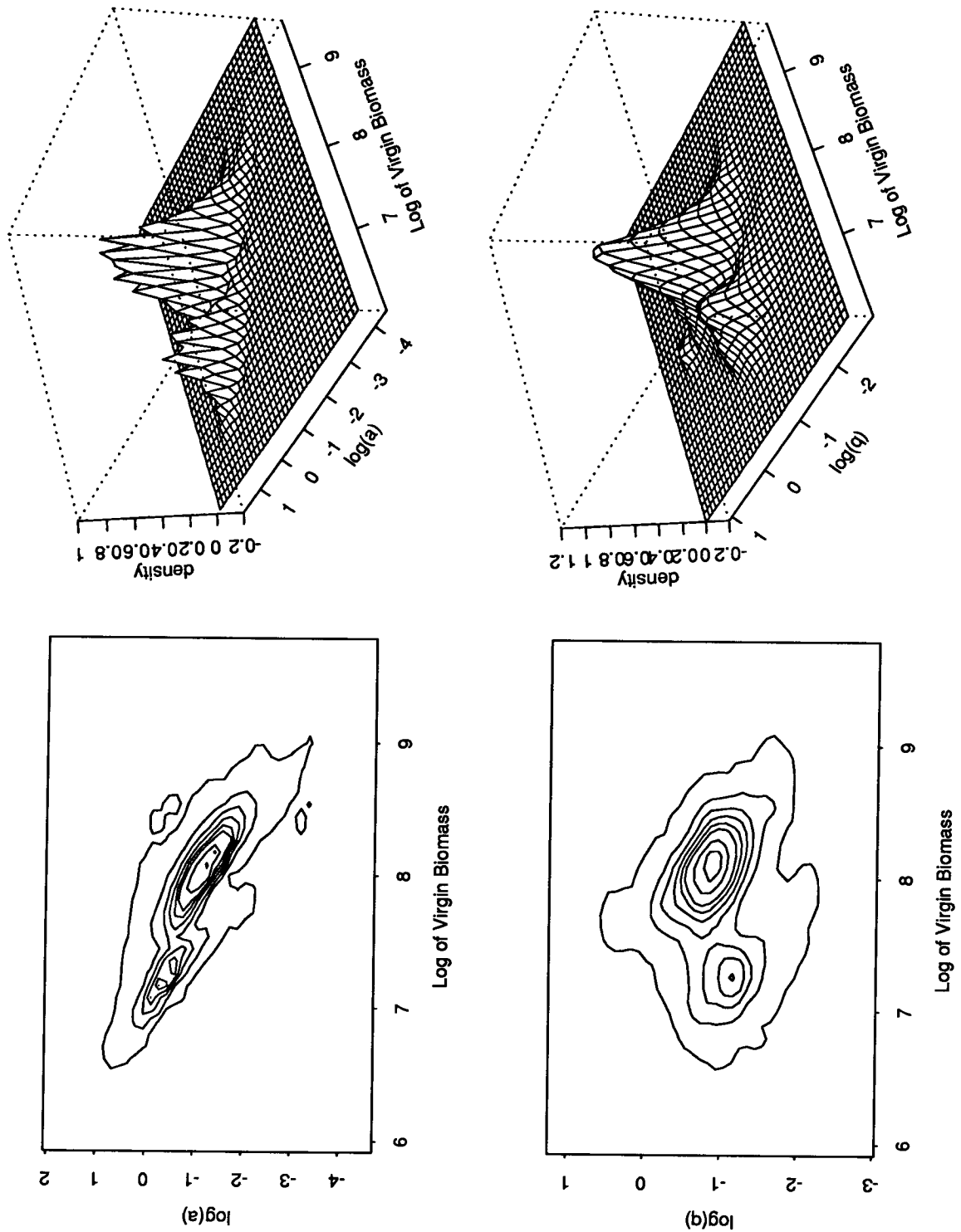
Figure 4.15: Some marginal posteriors for data simulated using the logistic biomass model and fitted by the mixture model described in text. The contours are at .01,.10,.20,.30,.40,.50,.70,.90 of the largest observed posterior density.

Table 4.17: Parameters which are fixed in the delay-difference model for Pacific cod (*Gadus macrocephalus*).

Parameter	Value
α	1.0
ρ	0.8
w_k	0.5
ζ	0.6
u	3

$\{I_1, \cdots, I_\tau\}$ for a total of $\tau = 22$ years. The abundance index is *catch per unit effort*, CPUE. I shall define harvest rates in terms of catches by

$$h_t = C_t/B_t \text{ for } t = 1, \cdots, \tau$$

The abundance index I_t is assumed proportional to the corresponding true biomass B_t but subjected to i.i.d. observation errors v_t. Hence, I use the standard model

$$I_t = qB_t e^{v_t} \text{ for } t = 1, \cdots, \tau$$

where q is the constant of proportionality.

The unknown parameters are: b in the recruitment equation (3.35); $B_1 = B_e$ the (initial) equilibrium biomass needed in (3.34) and (3.35); and the constant of proportionality q. The parameter vector $\theta = (\theta_1, \theta_2, \theta_3)'$ is given by

$$\theta = (\ln b, \ln q, \ln B_1)'.$$

The observation model, written as a function of θ, is defined by:

$$(Y_t \mid \theta, V) \sim N(\mu_t(\theta), V) \tag{4.40}$$

For any fixed value θ and time-series of catches $\{C_t\}$, the sequence $\{\mu_t(\theta) : t = 1, \cdots, \tau\}$ is obtained recursively, using the following steps:

1. Start with $B_1 = e^{\theta_3}$ calculate R_1 using (3.34).

Table 4.18: Quantiles specified for the parameters. The mean m and variance C are parameters for the corresponding Gaussian distribution for the logarithm of the parameter.

Quant.	5	25	50	75	95	m	C
b	1.0	1.4	2.5	3.5	5.0	0.82	0.28
$q \times 10^{-3}$	1	6	25	100	500	-3.74	3.73
B_1	1	3	6	9	18	1.60	0.75

2. Substitute R_1 into (3.33) to obtain N_1.

3. For $t < 5$, define $R_t = R_1$; otherwise use (3.35) and the corresponding "lagged" spawning stock S_{t-u-1}.

4. Use (3.31) and (3.32) recursively to obtain the sequences $\{\hat{B}_t : t = 1, \cdots, \tau\}$ and $\{\hat{N}_t : t = 1, \cdots, \tau\}$.

5. Calculate $\{\hat{\mu}_t(\theta) : t = 1, \cdots, \tau\}$, where $\hat{\mu}_t(\theta) = \ln(q\hat{B}_t)$ and $q = e^{\theta_2}$.

The posterior distribution for θ is obtained using the adaptive importance sampling (AIS) of Section 2.3. For data fitting I use the algorithm proposed in Section 3.3.

To describe background information about the parameter vector θ, I constructed the prior distribution based on educated guesses for specified quantiles of the original parameters to which biological meaning can be attached. Table 4.18 gives these values for five different quantiles of an "imaginary" marginal log-normal distribution for each parameter. The listed quantiles were the outcome after a series of minor adjustments until reaching consistency among all five values. The parameters which would correspond to the underlying Gaussian distribution were used as the mean m and variance C in a three-dimensional Student distribution with 4 degrees of freedom and diagonal covariance matrix.

In order to evaluate the range of values for b, notice that

$$\frac{1}{b} = \frac{S_{\text{opt}}}{B_e}$$

where S_{opt} is the spawning stock which produces the maximum number of recruits R_{opt}. It is easy to see from (3.35) that

$$\frac{R_{\mathrm{opt}}}{R_e} = \frac{1}{b} e^{b-1}$$

so that for $b = 5$ the recruitment R_{opt} would be eleven times larger then its value at unfished equilibrium R_e. It is also reasonable to assume $0 < S_{\mathrm{opt}} < B_e$, which implies $b > 1$. For q and B_1 the quantiles can be interpreted directly.

As prior for the variance V I assume an Inverse-Gamma distribution with parameters $\alpha_V = 2$ and $\beta_V = 8$ which is diffuse (infinite variance) and corresponds to a coefficient of variability of approximately 20% in the observed abundance.

The approximation for the posterior distributions were obtained after 4 iterations using $n_0 = 1000$, $n_1 = 2000$, $n_2 = 2500$, $n_3 = 3500$, and $n_4 = n = 6000$. At each stage, a reduction to a smaller mixture model of $m = 800$ components was performed before proceeding. The heavier tails (4 degrees of freedom) were chosen empirically after trying different possibilities while aiming for stable measures of relative entropy (2.4). For the results discussed here, I set the bandwidth $h^2 = 0.9$. The entropy of the final approximations was 0.91 (recall that a value of 1 corresponds to a perfect fit at the posterior sample points). In Figure 4.16 (a, b and c), contour-plots for the two-dimensional marginal posterior distributions are displayed. The pair $(\ln b, \ln B_1)$ (Fig. 4.16(a)) has a primary mode near $(1.4, 1.3)$, while a secondary (less likely) mode is near $(0.5, 2.1)$. Bi-modality is also shown in (c), the marginal posterior for $(\ln q, \ln b)$. In (b) a strong (negative) correlation between $\ln B_1$ and $\ln q$ features the confounding effect typically found for those parameters. Figure 4.16 (d) shows the one-dimensional marginal for $\ln B_1$. The comparison of this one-dimensional marginal distribution with the contour plots (a) and (b) gives a good indication of loss in information incurred by the reduction in dimensions. Still, such a one-dimensional marginal posterior distribution has much more to say than a point estimate as, for instance, the Monte Carlo mean 1.89 with standard deviation 0.36.

The confounding effect between B_1 and q suggests that a good prior knowledge about q is likely to reduce posterior uncertainty about B_1. In order to see this, I have defined a more
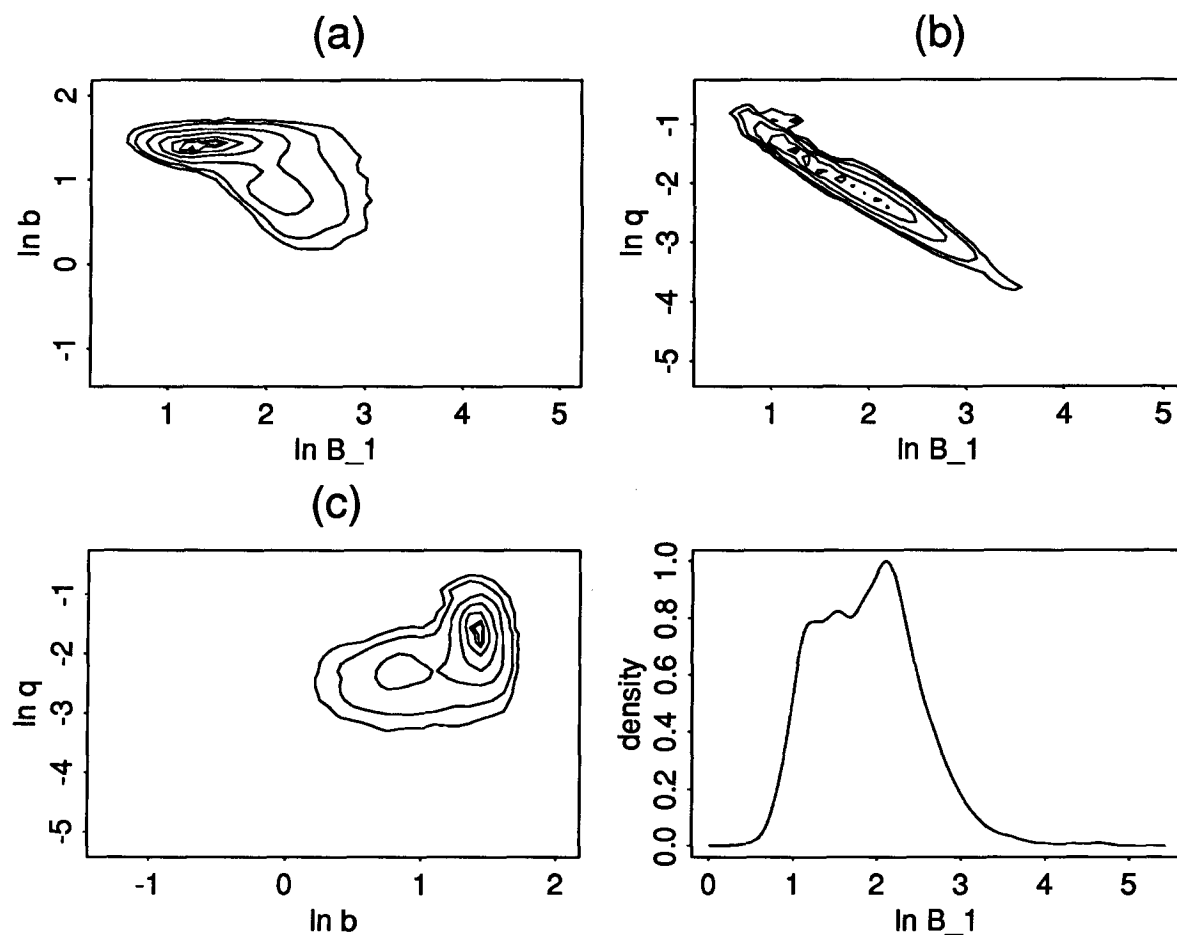
Figure 4.16: Estimated posterior marginal densities for Pacific Cod (*Gadus macrocephalus*). The contours represent .05, .10, .25,.50, .75, .90 and .95 of the largest observed density.

Table 4.19: Monte Carlo estimates of posterior mean (standard deviation) for diffuse and informative priors. The entropies are for the final importance function.

Estimates	Diffuse	Informative
V	0.28 (0.10)	0.30 (0.10)
B_1	8.1 (7.5)	8.7 (4.0)
$P(B_1 > 9.0)$	0.29 (.006)	0.40 (.006)
$P(B_1 > 10.0)$	0.23 (.005)	0.28 (.006)
Entropy	0.91	0.95

informative prior for q, setting the quantiles 5, 25, 50, 75, 90 and 95 at values .03, .05, .08, .12, .22 respectively. Such resulted in the prior moments $m = -2.53$ and $C = 0.38$ for $\ln q$. This prior favours the less likely of both modes. Figure 4.17 shows the contours (and perspective plot) for the marginal posterior between $\ln b$ and $\ln B_1$, together with the one-dimensional marginals for the individual parameters. Notice the increased credibility associated to this secondary mode, confirming my predictions. At the same time the mode corresponding to the smaller biomass dominates — a consequence of the strong favouritism towards the recruitment parameter b associated to that mode. The change in the shape of the marginal posteriors for $\ln B_1$ and $\ln b$ are most noticeable. Figure 4.18 serves this purpose: the marginal posteriors for the "informative" prior (broken line) increase possibilities for the smaller mode for $\ln b$ and substantially reduce the chances for values of $\ln B_1$ below 1.5.

Table 4.19 and Table 4.20 give additional summaries for estimates obtained for both fits. The case with diffuse prior suggests a slightly smaller expected initial biomass of 8.1 as compared to 8.7 obtained for the informative case. The increased background information assumed in the latter case is also reflected in the overall reduction in variances; the exception is $\ln b$ for which the variance increased slightly — possibly due to the more pronounced bimodality.

4.7   Final Remarks

Throughout this Chapter I have examined a series of case studies for Bayesian inference and decision making. These applications were intended to provide a range of illustrations for the
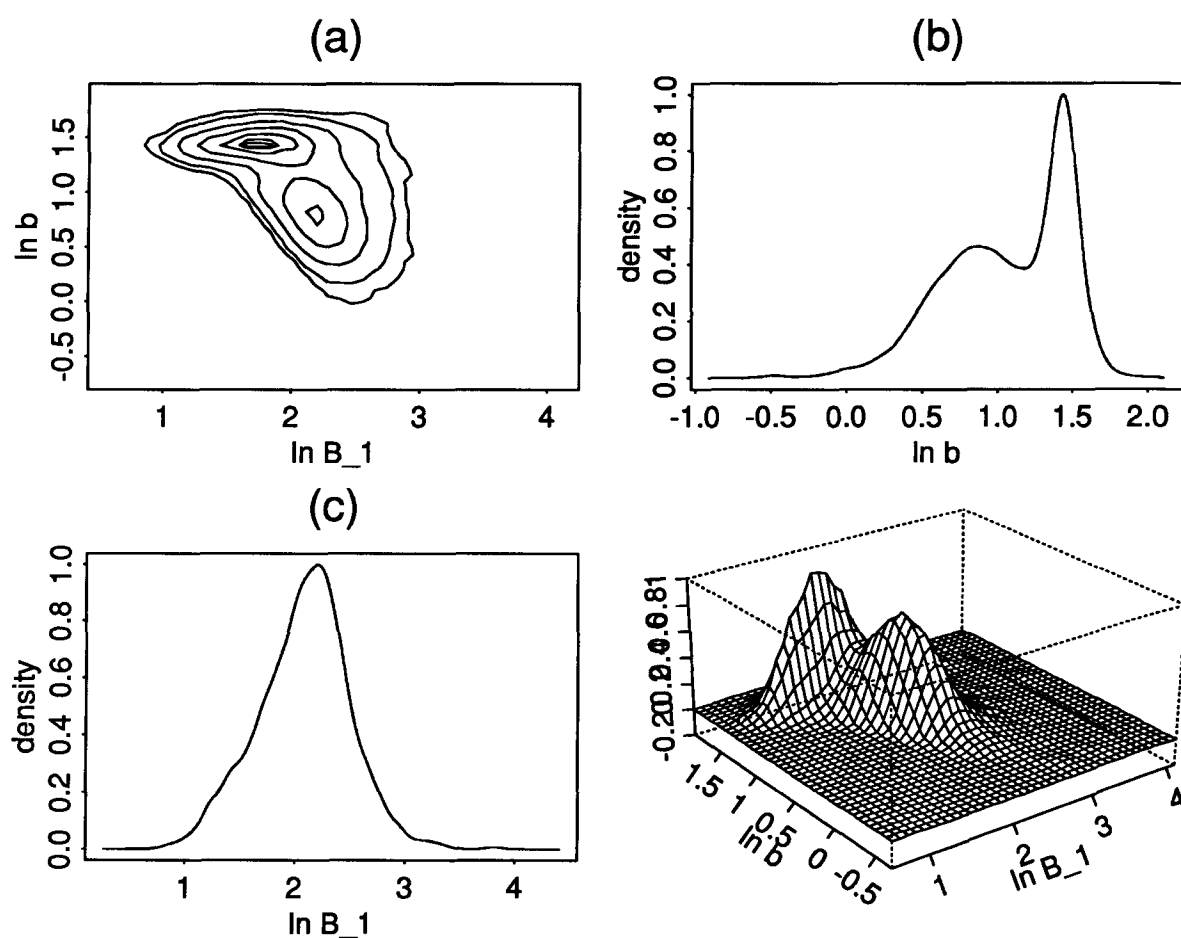
Figure 4.17: Estimated posterior marginal densities for Pacific cod, using an informative prior for the catchability parameter q. The contours represent .05, .10, .25, .50, .75, .90 and .95 of the largest observed density.
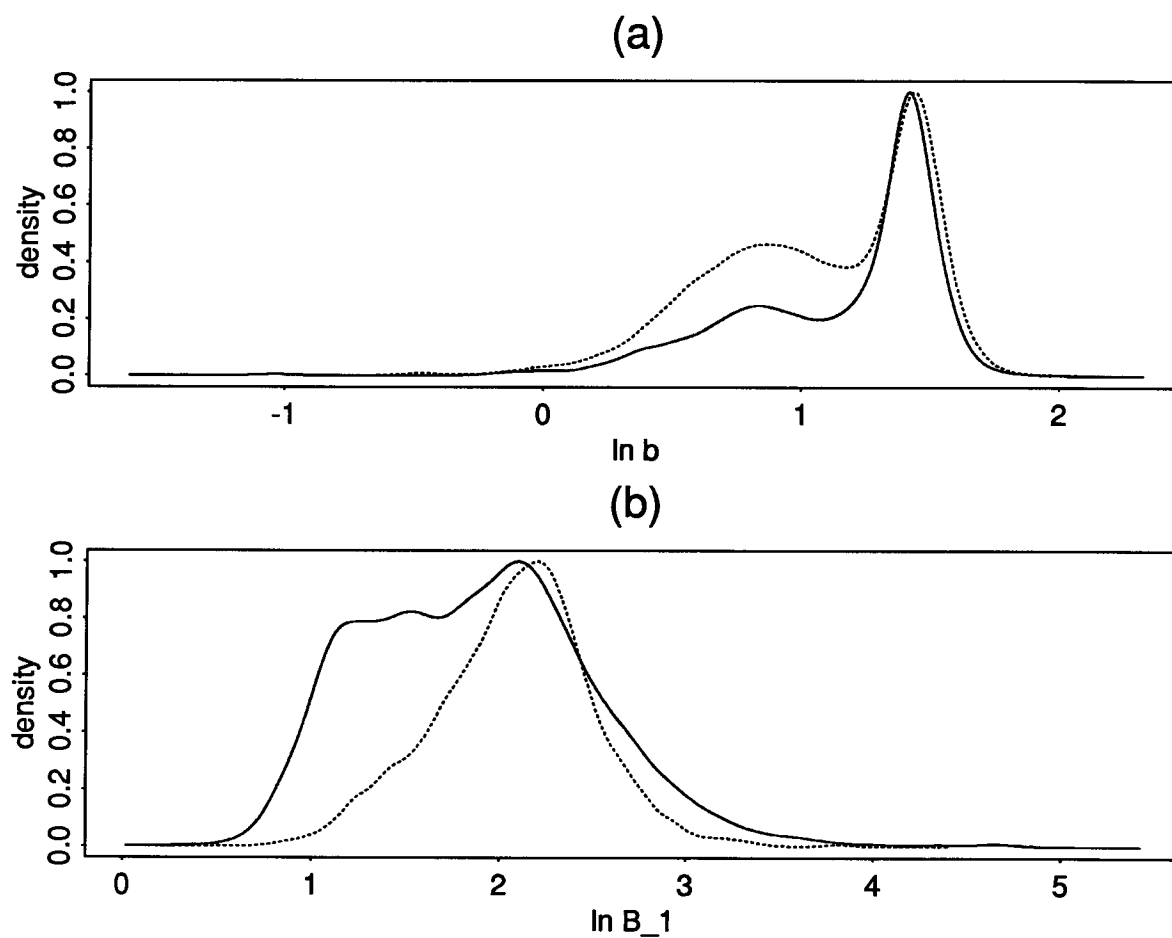
Figure 4.18: Marginal posterior densities for: (a) $\ln b$, (b) $\ln B_1$. The full and dotted lines correspond to the "diffuse" and "informative" priors respectively.

Table 4.20: Monte Carlo estimates of moments of the posterior distribution $p(\theta \mid Y(\tau))$ for Pacific cod, given diffuse and informative priors. The values m_0, $m1$ and m_2 refer to $\ln b$, $\ln q$, and $\ln B_1$ respectively. The values $C_{i,j}$ are the corresponding covariances.

Parameter	Diffuse	Informative
m_0	1.16	1.07
m_1	-2.09	-2.28
m_2	1.89	2.08
$C_{0,0}$	0.14	0.16
$C_{0,1}$	0.11	0.05
$C_{0,2}$	-0.13	-0.09
$C_{1,1}$	0.36	0.15
$C_{1,2}$	-0.35	-0.15
$C_{2,2}$	0.36	0.17

procedures described in Chapters 2 and 3, especially the use of adaptive importance sampling (AIS) to approximate posterior distributions of parameters in biomass dynamic models. The full benefits of such a rich description of uncertainties, are realized in decision making. In two case-studies — the simulation of Section 4.3 and the Orange Roughy fishery of Section 4.4 — I take this extra step after defining a suitable utility function and a set of possible management actions.

The first case-study (Section 4.2) considered stock and recruitment data for Skeena River sockeye salmon (*Oncorhynchus nerka*). I constructed a posterior distribution for the two-dimensional parameter and used standard density estimation techniques to approximate posterior distributions for two policy parameters. The estimates were discussed in the light of a previous analysis performed by Hilborn and Walters [25] for the same data. I further used this example to illustrate the advantage of adaptive importance sampling over one-time importance sampling.

A second case-study (Section 4.3) was based on simulated data. The effect of process noise and observation error and the robustness of different fitting models were examined. For the study I replicated some of the simulation models used by Ludwig and Walters [38], but used the Bayesian models to estimate the "optimal effort" defined in their paper. I further included

a new criterion to choose some "best effort", based on utility theory and Bayesian decision analysis. This new approach is attractive given its simplicity of concept and the avoidance of unrealistic equilibrium assumptions. Since the approach requires a posterior distribution, it is an exclusive feature of the Bayesian method. If a fitting model is judged by its capability of correctly estimating the policy parameters of interest, one is compelled to give up the fidelity of the model to the "true" underlying dynamics in order to improve the performance of estimation. Such findings were reported previously by Ludwig and Walters [38] and are confirmed within the Bayesian setting. The same feature was also identified for the new approach by comparing the estimated best decision to its *perfect information* alternative.

The third case-study (Section 4.4) refers to the fishery of Orange Roughy (*Hoplostethus atlanticus*) off the New Zealand coast. I used the data of Francis [18] and a surplus-production model to examine the risk of different management options. Francis produced estimates of risk of collapse from a sophisticated age-structured model. Such risks are too high if compared to my results based on a much simpler and possibly more robust model. Furthermore, the Bayesian framework of my analysis can better account for uncertainties in underlying parameters and its consequences for decision making. My results suggest that the choice between various management options is primarily driven by the utility function which compromises between risk of collapse and future catches. The more one favours the former (large risk-aversion), the more conservative (i.e. larger rate of annual reduction in total allowable catches (TAC)) the solution.

The richness of information available through a posterior distribution becomes very evident in cases where this distribution is multimodal. Such situations highlight the weakness of any approach which forces estimation to one point only. Decision making that is based on point estimates ignores the uncertainty among a multitude of reasonable answers. The last two case-studies examine multimodality. In Section 4.5 I once again resort to a simulation model. I consider an imaginary fishery for which the effectiveness of fishing effort (measured by the *coefficient of catchability*) changes randomly from year to year. The results show a clear bimodal

posterior, indicating the duality which the fitting model is incapable to resolve. Section 4.6 returns to real data. This time I consider Pacific cod[6] (*Gadus macrocephalus*) and use a more elaborate *delay-difference* model (see Section 3.4) to describe biomass dynamics. The posterior distribution indicates two modes. The analysis of one- and two-dimensional marginal posterior densities also depicts the loss in information incurred by dimensionality reduction.

---

[6]The potential for bimodality in this particular data set was pointed out to me by C. J. Walters.

Chapter 5

Discussion and Conclusion

"Fisheries management is a decision problem, not an assessment problem", say Hilborn and Walters [25] after a detailed description and analysis of where and why it has failed in the past. The confusion between management and assessment, they claim, is at the root of the problem. Why ?

People used their expertise to build more complex, realistic models. It was thought that inclusion of any new insight about biological or ecological phenomena, would enhance the predictive capabilities of such models. In seeking to maintain the credibility of fishery sciences, there was a tendency towards suppression or elimination of all elements of doubt. Confession of uncertainty was considered to be a sign of weakness. But failures in the real world revealed that this procedure was not effective.

In the 1970's and 1980's, Hilborn, Ludwig, Walters, and others ([71], [34], [37], [33], [39], [38]) indicated where to look for possible flaws. They used simulation studies to show that typical fishery data are not informative enough to allow for reliable estimation in realistic but highly parameterized models. It was recognized that a compromise between performance of estimates and complexities of models had to be made. Traditional management practices, based on equilibrium assumptions and goals for stable catches, were in collision course with the desire for learning about the systems potential. Less realistic but simpler models were shown to have a better performance in management than more sophisticated alternatives. Uncertainty was exposed; its description presented a new challenge. Such a major change in thinking about models simultaneously changed the concept of management.

Traditional notions and management practices were questioned. "It is far more important to

explicitly consider uncertainty than to apply sophisticated analytical techniques", say Hilborn and Walters [25]. The pretence of certainty was abandoned and fishery management seen as decision making under uncertainty. It became important to identify competing models and compare possible actions. The distinctive roles of assessment and management became clear: *assessment* should produce the alternative models (hypothesis) and their relative credibility, while *management* had to establish a set of actions and the measurement of their consequences. All these insights led naturally to use of Bayesian statistics.

Lindley [31] states that, for coherent decision making in the presence of uncertainty, there is only one way to go:

1. All uncertainties *must* be quantified by *probabilities.*

2. The consequences of various actions *must* be described by *utilities.*

3. The action, expected to give the largest utility, *must* be taken.

There is considerable theoretical evidence — see Berger [2] and de Finetti [8] for lengthy treatments and further references — to support the conclusion that Bayesian decision theory gives the framework to accomplish this goal. It has been shown that any decision making which does not *correspond* to a Bayesian analysis, is liable to lead to incoherent procedures, and hence repeated use may lead to the collapse of the fishery.

From a Bayesian standpoint, probabilities are the expression of beliefs about uncertain events. The individualistic perception which goes into such interpretation, gave it the name: *subjective probability.* Classical non-Bayesian views, reject such probabilities on two grounds:

1. The frequentist interpretation of probability (e.g., the probability $\frac{1}{2}$ of heads in tossing a fair coin) as the only acceptable one; and

2. the impression of lack in scientific objectivity, caused by the inclusion of personal opinions, perceived as arbitrary.

In answer to argument 1, it can be said that the flexibility gained from subjective probabilities is in fact an advantage of Bayesian statistics. In a frequentist approach there is no room for statements like "high probability that the Orange Roughy fishery will collapse within the next 5 years" while it offers no difficulty to the subjectivist. As to the criticism in 2, *any* statistical analysis leading to decision making, relies on subjective (personal) inputs. The Bayesian approach makes such inputs explicit and open to evaluation by other experts; this is certainly a scientific attitude. The reader can find more details about the controversy between 'objectivists' and 'subjectivists' and the philosophical foundations underlying the Bayesian approach, in Berger [2] (sections 1.6, 3.1 and 4.1) or, in a less technical language, in Lindley [31].

If we agree with subjective probabilities as a coherent way to express uncertain beliefs, Bayes' Theorem becomes an important conceptual link between (prior) beliefs and the (statistical) information provided by data. The resulting *posterior* probability gives the complete description of current beliefs. The theorem tells how beliefs have to change in the light of a new piece of evidence (data); it is a mathematical description for coherent inductive reasoning.

But, in public decision making — as is the case in fisheries — it can be difficult to find a prior or utility to which all participants would agree. The Bayesian statistician may use a "diffuse" (or "vague") prior or may report the outcome of the analysis over a set of reasonable alternatives. Often — as was observed for Orange Roughy — the management decision is robust to such alternatives. If, however, different reasonable priors give conflicting answers, then the scientific uncertainty is to be treated as a feature of the problem and should never be eliminated by artificial or hidden assumptions. A set of sensible utilities, on the other hand, is an useful negotiating tool among various interest groups.

Management: description or prescription

Our modern thinking about assessment and management includes concepts of adaptive management and replication. Such tools are needed to develop a better understanding of stock potential and enable more reliable predictions. The answers sought are therefore *descriptive*;

they explain how nature behaves.

Management decisions are needed, and will always be based on beliefs, no matter how good the understanding of the system or how reliable the predictions. In the worst scenario, uncertainties are simply ignored. To ensure better decision making, *prescriptive* answers are needed.

According to de Finetti [8], the subjectivistic theory of probability is prescriptive rather then descriptive. A posterior distribution is not the description of a real phenomena but an expression of the perceived uncertainties and beliefs about it, given all the available evidence. Such distribution becomes of practical relevance when used in combination with measures of consequences (utilities), to *prescribe* a best action.

Adaptive management uses experimentation to enhance understanding (description) of the mechanisms acting on renewable resources under exploitation. There are, however, recognized difficulties with the practical implementation of experimental management programs. The difficulties go from technical (insufficient replicates) to institutional (coordination, enforcement) and political (who sets priorities and accepts the responsibilities for the decision). But how can we predict the effect of an action if there are no replicates and controls to compare with ? How to respond to unexpected (or contradictory) predictions from different working models ?

Regardless of the ability to overcome these problems, the prescriptive role of management will still be necessary. Although often preferred, inaction is merely one of many possible choices. I believe that Bayesian statistics has much to offer in helping to improve the decision making process. If people are trained in better elicitation of uncertainties, if utilities give a fair description of preferences and attitudes towards risk, then a coherent framework is available. How can we accomplish the complex task (decision making) if we can not do the elementary ones (describe uncertainties and utilities) ?

The contribution of this work

Technical difficulties have hindered the implementation of the Bayesian approach. For fishery sciences in particular, the non-linearity of models and plethora of parameters have impeded easy implementation.

This technical challenge had to be tackled so that the Bayesian paradigm could become of practical significance. By using the adaptive importance sampling procedure introduced by West [75] [76], I was able to construct multivariate posterior distributions for the parameters of biomass dynamic models. Since the procedure gives a random sample of values from the posterior distribution, this output can be coupled easily with an utility function to calculate approximate Monte Carlo expectations — a requirement for Bayesian decision making.

The application of these results to a variety of case-studies and its combination with reasonable utility functions served to illustrate the new possibilities. For instance, it shed new light on the controversy about the New Zealand Orange Roughy fishery. The (frequentist) results of Francis [18] were shown to agree with the Bayesian analysis only when a particular (informative) prior was assumed. But the sensitivity of the model to different reasonable priors turned out to be unimportant when deciding upon the best policy. For this case in particular, my analysis suggests that the decisive issue is the choice of an utility function; neither prior uncertainties nor technical details about models are influential. Because this choice is political and not technical, the scientist doing assessment should not impose his or her judgement of "adequate" risk taking. He or she are entitled, however, to assist in the construction of an adequate utility function and in the elicitation of prior beliefs as well as the calculation of expected utilities. In this context the role of assessment is more than data analysis; it includes the analysis of beliefs and attitudes towards risk.

Future Work

There are many aspects of the Bayesian method where practical difficulties remain and where more work is needed. Some are general: elicitation of priors and utilities or the problem of public decision (i.e., more than one decision maker). Others are specific: the extension to other classes of fishery models and/or importance functions. I shall comment on each separately:

Elicitation: In his description of the importance of probabilities as the adequate measurement of uncertainty, Lindley [31] uses an analogy in geometry. He states that in the same way as the laws of geometry have to be satisfied in the construction of a building, the laws of probability are to be satisfied if a coherent description of uncertainties is to be achieved. Pursuing the analogy further, he compares the tools to measure lengths and angles to methods for elicitation of priors. Such methods might be simple or sophisticated, inaccurate or precise. The decision maker, likewise the architect, will need tools to make the required measurements. Much needs to be done to develop better tools for measurement of uncertainty. Nevertheless, concludes Lindley, while technological applications are still deficient, the scientific principles are clear.

Public Decision: In the process of decision making it is usually a committee that has to arrive at a consensus, each committee member having his or her own utility function and prior uncertainty. All analyses considered throughout previous chapters made the implicit assumption of one decision maker only. What can we do when beliefs and interests of different individuals have to be used ? The theory for dealing with group decision (Weerahandi and Zidek [73], Zidek [81], de Waal et al. [9]) was developed to address this issue and can be tried out here.

Fishery Models: Surplus-production models were central to this work. I commented that there was no conceptual difficulty in extending the procedure to more elaborate alternatives. In fact, a standard delay-difference model was used in the case-study for Pacific cod. We have also seen throughout that, given the common limitations in quantity and

quality of fishery data, simpler (i.e., less parameterized) surplus-production models tend to be more robust to correctly identify the best policy choice. But if we can improve prior elicitation of parameters in the more demanding delay-difference models, the potential gain is considerable. To train fishery biologists for a better prior elicitation might include the following steps:

1. Acquaint the elicitors with probabilities and its role in describing uncertainties.

2. Generate artificial data using detailed and realistic models.

3. Get elicitation of parameters for specified models and based on the artificial data from previous step.

4. Perform analysis to identify failures of model and elicitation. Try to answer questions like:

   - Which parameters are more critical ?

   - Which models are more robust ?

   - Can elicitors learn from past mistakes ?

The extension to nonstationary models (i.e. time-dependent θ_t) is another promising area for further research, as is the search for alternative mixture models which enable to work with strictly positive variables directly (i.e., no need to use log-transformations).

Conclusion

Active adaptive strategies are the essence of modern fishery management. Such experimental approach can increase our understanding (description) of the system and improve our capabilities for prediction. But, it needs the features of Bayesian analysis to determine how uncertainties propagate in time as new evidence becomes available through new data.

Policy choices need to be made, independent of the progress in adaptive management practices. Bayesian analysis provides coherent prescriptions of how such choices are to be made."In all cases where details are available", says Wilimovsky [78] in (non-Bayesian) discussion about

fishery management, "the output of formal predictions is modified by judgment. The methodology of judgment is not formalized so that it is unlikely that two or more individuals would arrive at the same management recommendation or even have the ability to trace the other's logic".

Appendix A

Pseudo-code for Adaptive Importance Sampling

This appendix presents a pseudo-code which describes the implementation of the adaptive importance sampling (AIS) of Chapter 2. In particular, it refers to the procedures described in Section 2.3. For concreteness, the code considers a fixed number of updating steps (modifications for an interactive procedure are easy to implement in principle) and assumes two time-series of data (not necessarily of the same length): (i) total catches and (ii) some abundance index. I use the fitting procedure described in Section 3.3 as "Model 2" and take the functions $G(\cdot)$ and $F(\cdot)$ to be known.

I shall specify some notation to be used in the pseudo-code:

- A vector $(a_1, \cdots, a_P)$ of order P, will be denoted `a[1...P]`. A single element a_i from this vector is identified as `a[i]`.

- A matrix $(N \times P)$ with N lines and P columns with its elements given by $\{a_{j,i} : j = 1, \cdots, N; i = 1, \cdots, P\}$, will be denoted `a[1...N,1...P]`. A single element $a_{j,i}$ is identified as `a[j,i]`.

- The j-th line of the above matrix — the vector of order P given by $\{a_{j,1}, \cdots, a_{j,P}\}$ — is denoted `a[j,1...P]` or simply `a[j,*]`. Any operation performed on such a line j is understood to be applied to each of its P elements individually.

- Let S be a symmetric positive-definite matrix of order P. I define a vector T of order $K = P*(P+1)/2$, containing the elements of the lower triangle of S, taken by row. That is, for each element $(S_{j,i} : j \geq i)$ define $r = (j(j-1)/2) + i$ to get

$$T_r = S_{j,i}$$

127

Covariance matrices will be stored in the form T rather than S throughout the code.

- The different steps in the pseudo-code are numbered. Global numbers (1, 2, etc.) if read in isolation, give a summary description of the sequential structure of the algorithm. Lower levels of numbering (2.1, 2.1.1, etc.) give details about how the task of '2' is accomplished.

- The conditional operator IF (condition) GOTO (x.x) is to be interpreted as follows:

  1. Verify condition.

  2. If condition is TRUE, than go to line number x.x.

  3. If condition is FALSE, than go to next line.

The description of the code is sub-divided into three parts: (I) - a "list of variables" for which the values need to be specified in advance; (II) - The pseudo-code for the main program; (III) - Pseudo-code describing subroutines used in (II).

I - LIST OF VARIABLES:

```
P: dimension of the parameter vector 'Theta'.
K = ( P*(P+1)/2 ): the dimension of a vector giving the
    lower triangle of PxP Covariance matrix.
G: number of iterations in the AIS procedure.
NO: size of sample to be drawn from prior.
N[1...G]: sample size to be drawn in each iteration.
M: number of Student distributions to be used in the clustered
    mixture model.
DFO: degrees of freedom in prior Student distribution.
DFk: degrees of freedom of the Student kernel in the mixture
    model.
H: the bandwidth of kernel variance.
```

DATA_C: number of observed catch data.

DATA_A: number of observed abundance index data.('DATA_A' is
smaller or equal to 'DATA_C)

THETA_MEAN[1...P]: prior mean for 'Theta'.

THETA_VAR[1...K]: lower triangle of the prior covariance matrix
for 'Theta'.

ALPHAO and BETAO: parameters for Inverse Gamma prior.

CATCH[1...DATA_C]: time-series of observed catches.

ABUN[1...DATA_A]: time-series of observed abundance index data.

II - PSEUDO-CODE:

1: Draw a sample of size NO from the prior distribution and store
the results in THETA[1...NO,1...P].

1.1: Perform a "Choleski Decomposition" (see final remarks) of
THETA_VAR[1...K] and put the result in CHOLESKI[1...K].

1.2: For j=1 to j=NO

1.2.1: Draw a random vector from a "Multivariate Student
Distribution" (see final remarks) using as input
THETA_MEAN[1...P], CHOLESKI[1...K] and DFO.

1.2.2: Store the output in THETA[j,*].

Next j

2: Calculate the likelihood for ABUN[1...DATA_A] for each
given value of THETA[1...NO,*] and store the normalized

```
    results in OMEGA[1...NO].

    2.1: For i=1 to DATA_A

         2.1.1. Get Y[i] = ln( ABUN[i] )

         Next i

    2.2: Initialize SUM = 0

    2.3: For j=1 to NO

         2.3.1: Get the time series of deterministic average

                biomass dynamics by using THETA[j,*], ALPHAO,

                BETAO and CATCH[1...DATA_C] in the assumed

         biomass model (G(.) and F(.)). Store the

         relevant values in MU[1...DATA_A].

         2.3.2: Store the likelihood value in OMEGA[j].

         2.3.3: SUM = SUM + OMEGA[j]

         Next j

    2.4: For j=1 to NO

         2.4.1: OMEGA[j] = OMEGA[j]/SUM

         Next j


3: Calculate the "relative entropy" of the prior and store it as

   EO or send it to the output device.

   3.1: Initialize ENTROPY = 0

   3.2: For j=1 to NO

        3.2.1: ENTROPY = ENTROPY + ( OMEGA[j]*ln(OMEGA[j]) )

        Next j

   3.3: EO = -ENTROPY/ln(NO)

   3.4: Send 'EO' to output device.
```

4: Perform the G recursions of the adaptive importance sampling
   procedure. In the process keep track of the entropies E[1...G].
   Store the final values THETA[1...N[G]] and OMEGA[1...N[G]] for
   further use.

   4.1: Get the "bandwidth" for kernel variance and store the
        outcome in WINDOW.

        4.1.1: X1 = (4/(1+2P))^(1/(1+4P))

        4.1.2: X2 = M^(1/(1+4P))

        4.1.3: WINDOW = H*(X1*X1)/(X2*X2)

   4.2: For k=1 to G

        4.2.1: IF (k>1) GOTO (4.2.5)

        4.2.2: Use current THETA[1...NO,*] and OMEGA[1...NO]
        to get Monte Carlo mean and covariance. Store the

               mean as KER_MEAN[1...P] and the lower triangle of

               the covariance matrix as KER_VAR[1...K].

   4.2.3: Use the subroutine "CLUSTER()" (described below)

          and the input of THETA[1...NO,*] and

          OMEGA[1...NO] to get the smaller sets

          T_CL[1...M,*] and 'O_CL[1...M].

          4.2.4: Go to (4.2.7)

          4.2.5: Use current T_OLD[1...N[k],*] and

                 O_KER[1...N[k]] to get Monte Carlo mean and

                 covariance. Store the mean as KER_MEAN[1...P] and

                 the lower triangle of the covariance matrix as

                 KER_VAR[1...K].

          4.2.6: Use the subroutine "CLUSTER()" and the input of

                 T_OLD[1...N[k],*] and O_KER[1...N[k]] to get

```
                    T_CL[1...M,*] and O_CL[1...M].

        4.2.7: For i=0 to K

               *.1: KER_VAR[i]=WINDOW*KER_VAR[i]

               Next i

        4.2.8: Perform a "Choleski Decomposition" of

               KER_VAR[1...K] and put the result in

KER_CHOL[1...K].

        4.2.9: For j=1 to N[k]

               *.1: Draw a random integer from the set {1,...,M}

                    according to probabilities given by

                    O_CL[1...M]. Store it in A.

               *.2: Draw a random vector from a "Multivariate

                    Student distribution" using as input

                    T_CL[A,*], KER_CHOL[1...,K] and DFK.

               *.3: Store the output in THETA_KER[j,1...P].

               Next j

        4.2.10: Calculate the weights of the random sample

                THETA_KER[1...,N[k],*] which was drawn from the

current "mixture model". Store this value in

                O_KER[1...N[k]].

               *.1: Initialize SUM=0

               *.2: For j=1 to N[k]

                    *.2.1: Calculate the "density" of

                           THETA_KER[j,*] for a P-dimensional

                           Student distribution with parameters

                           THETA_MEAN[], THETA_VAR[] and DFO.

                           Store the result in PRIOR.
```

```
*.2.2: Initialize MIXTURE = 0

*.2.3: For i=1 to M

          *.2.3.1: Calculate the "density" of

                   THETA_KER[j,*] for a

                   P-dimensional Student

                   distrib. with parameters

                   T_CL[i,*], KER_VAR[] and DFK.

                   Store the result in KERNEL.

          *.2.3.2: KERNEL = O_CL[i]*KERNEL

          *.2.3.3: MIXTURE = MIXTURE + KERNEL

          Next i

     *.2.4: Get the time series of average

            biomass dynamics using THETA_KER[j,*],

ALPHAO, BETAO, and CATCH[1...DATA_C]

in equation G(.).

Store the mean values in

MU[1...DATA_A] and get the likelihood

                   LIK for data ABUN[1...DATA_A].

     *.2.5: F = LIK*PRIOR

     *.2.6: O_KER[j] = F/MIXTURE

     *.2.7: SUM = SUM + O_KER[j]

     Next j

*.3: For j=1 to N[k]

     *.3.1: O_KER[j] = O_KER[j]/SUM

     Next j

*.4: Calculate the "relative entropy" of the

     current "mixture model" and store it as
```

```
                    E[k] or send it to output.

                    *.4.1: Initialize ENTROPY = 0

                    *.4.2: For j=1 to N[k]

                                *.4.2.1: ENTROPY = ENTROPY +

                                         O_kER[j]*ln(O_KER[j]).

                           Next j

                    *.4.3: E[k] = -ENTROPY/ln(N[k])

                    *.4.4: Send E[k] to output.

              *.5: IF (k=G) GOTO (*.8)

              *.6: For j=1 to N[k]

                    *.6.1: T_OLD[j,*] = THETA_KER[j,*]

                    Next j

              *.7: Go to (4.2.1)

              *.8: For j=1 to N[G]

                    *.8.1: THETA[j,*] = THETA_KER[j,*]

                    *.8.2: OMEGA[j] = O_KER[j]

                    Next j

           Next k


    5: Use THETA[1...N[G],*] and OMEGA[1...N[G]] to get Monte Carlo

       mean and covariance. Store the mean in POST_MEAN[1...P] and

       the lower triangle of the covariance matrix as

       POST_VAR[1...K]. Send this values to the output device for

       further use.


    6: (Optional): Get Monte Carlo estimates for mean and variance of

       functions "H(Theta)".
```

7: Use THETA[1...N[G],*] and OMEGA[1...N[G]] in the subroutine

   "CLUSTER()" to get the final "mixture model". Store these values

   as T_CL[1...M,*] and O_CL[1...M] and keep them in a File for

   further use.

III - SUB-ROUTINES:

CLUSTER(): This subroutine performs the clustering algorithm of

      section 2.4 (Chapter 2).

Input: - The dimension P of the parameter vector.

        - A positive integer N denoting the size of the data set.

        - A positive integer M (M < N) defining the desired size of

          the data set after clustering.

        - A data matrix THETA[1...N,1...P]

        _ A vector of weights OMEGA[1...N]

Output: - A matrix T_CL[1...M,1...P]

         _ A vector O_CL[1...M]

1: Initialize D=N

2: Make a copy of OMEGA[]. Also create

   a vector of identification indices.

   2.1: For j=1 to N

        2.1.1: O_COPY[j] = OMEGA[j]

```
          2.1.2: ID[j] = j

          Next j


3: Sort (see final remarks about "sorting") the vector

   O_COPY[1...N] into ascending numerical order and

   make the corresponding rearrangement of

   ID[1...N].


4: Sort the values in T_COPY[1...N,*] to agree with the

   ordered vector O_COPY[1...N].

   4.1: For j=1 to N

        4.1.1: S = ID[j]

        4.1.2: T_COPY[j,*] = THETA[S,*]

        Next j


5: IF ( D=M ) GOTO (7)


6: Reduce the data set by one unit, by merging the first case

   with its "nearest neighbour" and reallocating the new value

   within the sorted set.

   6.1: For j=2 to D

        6.1.1: Calculate the "distance" between T_COPY[1,*] and

               T_COPY[j,*] and store its value in DIST.

        6.1.2: IF ( j>2 ) GOTO (6.1.5)

        6.1.3: Initialize: DMIN = DIST

                           WEIGHT = O_COPY[2]

                           S = 2
```

```
          6.1.5: IF ( DIST>=DMIN ) GO TO (6.1.9)

          6.1.6: DMIN = DIST

          6.1.7: WEIGHT = O_COPY[j]

          6.1.8: S = j

          Next j

    6.2: O_TEST = O_COPY[1] + O_COPY[S]

    6.3: For i=1 to P

          6.3.1: T_TEST[i] = (O_COPY[1]*T_COPY[1,i] +

                 O_COPY[S]*T_COPY[S,i])/O_TEST

          Next i

    6.4: For j=2 to D-1

          6.4.1: IF ( j>=S ) GOTO (6.4.5)

          6.4.2: O_COPY[j-1]=O_COPY[j]

          6.4.3: T_COPY[j-1,*]=T_COPY[j,*]

          6.4.4: Go to (6.4.13)

          6.4.5: IF ( O_TEST<O_COPY[j+1] ) GOTO (6.4.9)

          6.4.6: O_COPY[j-1] = O_COPY[j+1]

          6.4.7: T_COPY[j-1,*] = T_COPY[j+1,*]

          6.4.8: Go to (6.4.13)

          6.4.9: O_COPY[j-1] = O_TEST

          6.4.10: O_COPY[j] = O_COPY[j+1]

          6.4.11: T_COPY[j-1,1...P] = T_TEST[1...P]

          6.4.12: T_COPY[j,*] = T_COPY[j+1,*]

          Next j

    6.5: D = D-1

    6.6: Return to (5)
```

```
7: Store T_COPY[1...M,*] and O_COPY[1...M] in the arrays defined
   for output.
   7.1: For j=1 to M
        7.1.1: O_CL[j] = O_COPY[j]
        7.2.2: T_CL[j,*] = T_COPY[j,*]
```

Final Remarks

Choleski Decomposition: The Choleski decomposition of a covariance matrix Σ results in a lower-triangular matrix L satisfying

$$\Sigma = LL^T$$

Such decomposition is helpful here in two ways: (i) to perform random draws from a multivariate distribution with covariance matrix Σ and (ii) to calculate the inverse matrix Σ^{-1}. For further details and algorithms I suggest Ripley [53] and Nash [44].

Student Distribution: In order to generate random draws from uni- and multidimensional Student densities, algorithms can be found in Ripley [53].

Sorting: To perform the "sorting" in the CLUSTER() subroutine, I used the "Heapsort" algorithm in Press et al. [52].

Bibliography

[1] M. Aoki. *State space modelling of time series.* Springer Verlag, 1987.

[2] J. O. Berger. *Statistical decision theory and Bayesian analysis.* Springer Verlag, 1985.

[3] G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control.* Holden-Day, revised edition, 1976.

[4] P. J. Brockwell and R. A. Davis. *Time series: Theory and Methods.* Springer Verlag, 1987.

[5] A. Charles. In-season fishery management: a Bayesian model. *Nat. Res. Mod.*, 2:457–498, 1988.

[6] C. A. Clark. *Bioeconomic Modelling and Fisheries Management.* John Wiley & Sons Inc., 1985.

[7] C. A. Clark. *Mathematical Bioeconomics: the optimal management of renewable resources.* John Wiley & Sons Inc., 2nd edition, 1990.

[8] B. de Finetti. *Probability, induction and statistics.* John Wiley & Sons, 1972.

[9] D. de Waal, P. Groenewald, D. van Zyl, and J. V. Zidek. Multi-Bayesian estimation theory. *Statistics and Decisions*, 03, 1985.

[10] M. Delampady, I. Yee, and J. V. Zidek. Hierarchical Bayesian analysis of a discrete time series of Poisson counts. Technical Report 111, Dept. of Statistics – University of British Columbia, 1991.

[11] M. Delampady, I. Yee, and J. V. Zidek. A smoothness-prior model for discrete time series. Technical Report 110, Dept. of Statistics – University of British Columbia, 1991.

[12] R. B. Deriso. Harvesting strategies and parameter estimation for an age-structured model. *Can. J. Fish. Aquat. Sci.*, 37:268–282, 1980.

[13] R. B. Deriso and A. M. Parma. Dynamics of age and size for a stochastic population model. *Can. J. Fish. Aquat. Sci.*, 45:1054–1068, 1988.

[14] B. Efron. Non parametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika*, 68:589–599, 1981.

[15] B. Efron. Bootstrap confidence intervals for a class of parametric problems. *Biometrika*, 72:45–58, 1985.

[16] D. A. Fournier and I. J. Doonan. A length-based stock assessment method utilizing a generalized delay-difference model. *Can. J. Fish. Aquat. Sci.*, 44:422–437, 1987.

139

[17] D. A. Fournier and A. Warburton. Evaluating fisheries management models by simulated adaptive control – Introducing the composite model. *Can. J. Fish. Aquat. Sci.*, 46:1002–1012, 1989.

[18] R. I. C. C. Francis. Use of risk analysis to assess fishery management startegies: a case study using Orange Roughy (Hoplostehus atlanticus) on the Chatham Rise, New Zealand. *Can. J. Fish. Aquat. Sci.*, 49:922–930, 1992.

[19] S. M. Fried and R. Hilborn. Inseason forecasting of Bristol Bay, Alaska sockeye salmon (Oncorhynchus nerka) abundance using Bayesian probability theory. *Can. J. Fish. Aquat. Sci.*, 45:850–855, 1988.

[20] H. J. Geiger and J. P. Koenings. Escapement goals for sockeye salmon with informative prior probabilities based on habitat considerations. *Fish. Res.*, 11:239–256, 1991.

[21] A. E. Gelfand and A. F. M. Smith. Sampling based approaches to calculating marginal densities. *J. Am. Stat. Assoc.*, 85:398–409, 1990.

[22] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. Pattn. Anal. Mach. Intell.*, 6:721–741, 1984.

[23] J. Geweke. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 57:1317–1339, 1989.

[24] R. Hilborn, E. K. Pikitch, M. M. McAllister, and A. Punt. A comment on "Use of risk analysis to assess fishery management startegies; a case study using Orange Roughy, (*hoplosthethus atlanticus*) on the Chatham Rise, New Zealand" by R.I.C.C. Francis. *Can. J. Fish. Aquat. Sci.*, 1993. To appear.

[25] R. Hilborn and C. J. Walters. *Quantitative fisheries stock assessment: choice, dynamics & uncertainty.* Chapman and Hall, 1992.

[26] N. L. Johnson and S. Kotz. *Continuous univariate distributions.* Houghton Mifflin, 1970.

[27] G. Kitagawa. Non-gaussian state-space modeling of nonstationary time series. *J. Amer. Stat. Assoc.*, 82:1032–1050, 1987.

[28] T. Kloek and H. K. van Dijk. Bayesian estimates of equation system parameters: an application of integration by Monte Carlo. *Econometrica*, 46:1–19, 1978.

[29] P. M. Lee. *Bayesian statistics: an introduction.* Oxford University Press, 1989.

[30] D. V. Lindley. Approximate Bayesian methods. In J. M. Bernardo et al., editor, *Bayesian Statistics 2*, pages 223–245. Valencia University Press, 1980.

[31] D. V. Lindley. *Making Decisions.* John Wiley & Sons, 2nd edition, 1985.

[32] H. Linhart and W. Zucchini. *Model Selection.* John Wiley & Sons, Inc., 1986.

[33] D. Ludwig. Small models are beautiful: efficient estimators are even more beautiful. In Springer Verlag, editor, *Unknown*, 1988. Conference held at Cornell University.

[34] D. Ludwig and R. Hilborn. Adaptive probing strategies for age-structured fish stocks. *Can. J. Fish. Aquat. Sci.*, 40:559–569, 1983.

[35] D. Ludwig, R. Hilborn, and C. J. Walters. Uncertainty, resource exploitation, and conservation: Lessons from History. *Science*, 260:17, 1993.

[36] D. Ludwig and C. J. Walters. Optimal harvesting with imprecise parameter estimates. *Ecol. Model.*, 14:273–292, 1982.

[37] D. Ludwig and C. J. Walters. Are age structured models appropriate for catch-effort data? *Can. J. Fish. Aquat. Sci.*, 42:1066–1072, 1985.

[38] D. Ludwig and C. J. Walters. A robust method for parameter estimation from catch and effort data. *Can. J. Fish. Aquat. Sci.*, 46:137–144, 1989.

[39] D. Ludwig, C. J. Walters, and J. Cooke. Comparison of two models and two estimation methods for catch and effort data. *Nat. Res. Model.*, 2:457–498, 1988.

[40] M. Mangel and C. W. Clark. Uncertainty, search and information in fisheries. *J. Cons. Ciem.*, 41:93–103, 1983.

[41] P. McCullagh and J. A. Nelder. *Generalized linear models*. Chapman and Hall, 2nd edition, 1989.

[42] R. Mendelssohn. Some problems in estimating population sizes from catch-at-age data. *Fishery Bulletin*, 86:617–630, 1988.

[43] T. S. Nakamura, S. Ohnishi, and Y. Matsumiya. A Bayesian cohort model for catch-at-age data obtained from research takes of whales. *Rep. Int. Whal. Comm.*, 39:375–382, 1989.

[44] J. C. Nash. *Compact Numerical Methods for Computers*. Adam Hilger Ltd., 1979.

[45] J. C. Naylor and A. F. M. Smith. Econometric illustrations of novel numerical integration strategies for Bayesian inference. *J. of Econometrics*, 38:103–125, 1988.

[46] M. S. Oh and J. O. Berger. Adaptive importance sampling in Monte Carlo integration. *J. Statist. Computn. Simuln.*, 1991. (to appear).

[47] A. M. Parma. Optimal harvesting of fish populations with non-stationary stock-recruitment relationships. *Nat. Resource Mod.*, 4:39–76, 1990.

[48] J. J. Pella and P. K. Tomlinson. A generalized stock production model. *Bull. Inter-Am. Trop. Tuna Comm.*, 13:419–496, 1969.

[49] A. Pole. Efficient Bayesian learning in non-linear dynamic models. Technical report, Dep. of Statistics – University of Warwick, 1988.

[50] A. Pole and M. West. Efficient Bayesian learning in non-linear dynamic models. *J. Forecasting*, 9:119–136, 1990.

[51] S. J. Press. *Bayesian statistics: principles, models and applications*. John Wiley & Sons, 1989.

[52] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical recipies in C*. Cambridge University Press, 1988.

[53] B. Ripley. *Stochastic Simulation*. John Wiley & Sons, 1987.

[54] D. B. Rubin. Comment on "The Calculation of posterior distributions by data augmentation", by M.A.Tanner and W.H.Wong. *J. Am. Statist. Assoc.*, 82:543–546, 1987.

[55] L. J. Savage. *The foundations of statistics*. John Wiley & Sons, 1954.

[56] J. Schnute. A general theory for analysis of catch and effort data. *Can. J. Fish. Aquat. Sci.*, 42:414–429, 1985.

[57] J. Schnute. A general fishery model for size-structured fish population. *Can. J. Fish. Aquat. Sci.*, 44:924–940, 1987.

[58] J. Schnute. The importance of noise in fish population models. *Fisheries Research*, 11:197–223, 1991.

[59] B. W. Silverman. *Density estimation for statistics and data annalysis*. Chapman and Hall, 1986.

[60] A. F. M. Smith. Bayesian computational methods. *Phil. Trans. R. Soc. Lond. A*, 337:369–386, 1991.

[61] A. F. M. Smith and A. E. Gelfand. Bayesian statistics without tears: a sampling-resampling perspective. *The Am. Statistician*, 46:84–88, 1992.

[62] A. F. M. Smith and G. O. Roberts. Bayesian computation via the Gibbs sampler and related Markov ChainMomte Carlo methods. *J. R. Statist. Soc. B*, 55:3–23, 1993.

[63] A. F. M. Smith, A. M. Skene, J. E. H. Shaw, and J. C. Naylor. Progress with numerical and graphical methods for practical Bayesian statistics. *The Statistician*, 36:75–82, 1987.

[64] H. W. Sorenson. Recursive estimation for nonlinear dynamic systems. In J. C. Spall, editor, *Bayesian Analysis of time-series and dynamic models*, pages 127–165. Marcel Dekker, 1988.

[65] A. M. Starfield and A. L. Bleloch. *Building models for conservation and wildlife*. McMillan Pub. Co., 1986.

[66] P. J. Sullivan. A Kalman-Filter approach to catch-at-length analysis. *Biometrics*, 48:237–257, 1992.

[67] L. Tierney and J. B. Kadane. Accurate approximations for posterior moments. *J. Amer. Statist. Assoc.*, 81:82–86, 1986.

[68] C. J. Walters. *Adaptive management of renewable resources*. McMillan Pub. Co., 1986.

[69] C. J. Walters. Nonstationarity of production relationships in exploited populations. *Can. J. Fish. Aquat. Sci.*, 44:156–165, 1987. Suppl. 2.

[70] C. J. Walters and J. S. Collie. Design of an experimental program for groundfish management in the face of large uncertainty about stock size and production. In R. J. Beamish and G. A. MacFarlane, editors, *Effects of Ocean Variability on Recruitment and an Evaluation of Parameters used in Stock Assessment Models.*, pages 13–25. Canadian Special Publications in Fisheries and Aquatic Sciences, 1989. No. 108.

[71] C. J. Walters and R. Hilborn. Ecological optimization and adaptive management. *Ann. Rev. Ecol. Syst.*, 9:157–188, 1978.

[72] C. J. Walters and D. Ludwig. Effects of measurement errors on the assessment of stock-recruitment relationships. *Can. J. Fish. Aquat. Sci.*, 38:704–710, 1981.

[73] S. Weerahandi and J. V. Zidek. Elements of multi-Bayesian decision theory. *Ann. Statist.*, 11:1032–1046, 1983.

[74] M. West. Mixture models, Monte Carlo, Bayesian updating and dynamic models. Discussion paper, ISDS – Duke University, 1992.

[75] M. West. Modelling with mixtures. In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 4*, pages 503–524. Oxford University Press, 1992.

[76] M. West. Approximating posterior distributions by mixtures. *J. R. Statist. Soc. B*, 55:409–422, 1993.

[77] M. West and J. Harrison. *Bayesian forecasting and dynamic models*. Springer Verlag, 1989.

[78] N. J. Wilimowsky. The need for formalization of decision algorithms and risk levels in fishery research and management. *Can. J. Fish. Aquat. Sci.*, 42:258–262, 1985. Suppl. 1.

[79] R. L. Wolpert. Monte Carlo integration in Bayesian statistical analysis. *Contemp. Math.*, 115(101–116), 1991.

[80] J. E. Zeh, A. E. Raftery, and Y. Qing. Assessment of tracking algorithm performance and its effect on population estimates using bowhead whales, Balaena mysticetus, identified visually and acoustically in 1986 off Point Barrow, Alaska. *Rep. Int. Whaling Comm.*, 40:411–421, 1990.

[81] J. V. Zidek. Multi-Bayesianity. Technical Report 05, Dept. of Statistics – University of British columbia, 1984.