12 Particle filtering

12.1 Introduction

In this chapter we discuss the filtering of non-Gaussian and nonlinear series by fixing the sample at the values previously obtained at times $\dots, t-2, t-1$ and choosing a fresh value at time t only. A new recursion over time is then required for the resulting simulation. The method is called *particle filtering*. We derive the results by classical analysis but point out that analogous methods can be obtained by linear methods and Bayesian analysis using Lemmas 2, 3 and 4 of Chapter 4. Illustrations are applied to real data in Chapter 14.

A large body of literature on particle filtering has emerged since the 1990s. An early use of the idea was given by Gordon, Salmond and Smith (1993) while the term particle appears to have first been used in this connection by Kitagawa (1996). For reviews of this field and resources of references, we refer to the book of Doucet, De Freitas and Gordon (2001) and also to the overview articles of Arulampalam, Maskell, Gordon and Clapp (2002), Maskell (2004) and Creal (2012).

We first consider in Section 12.2 filtering by the method of Chapter 11. We choose a sample from $y_1, \dots, y_t$ and use importance sampling to estimate $x_{t+1}, x_{t+2}, \dots$. This is a valid method of filtering and it is occasionally useful so we describe it here. However, it involves much more computing than particle filtering for routine use. We therefore go on to consider particle filtering and its association with importance sampling.

In Section 12.3 we discuss resampling techniques designed to reduce degeneracy in sampling. We go on in Section 12.4 to describe six methods of particle filtering, namely bootstrap filtering, auxiliary particle filtering, the extended particle filter, the unscented particle filter, the local regression filter and the mode equalisation filter.

12.2 Filtering by importance sampling

We shall consider a variety of formulations of the non-Gaussian nonlinear model, the most basic having the form

$$p[y_t|Z_t(\alpha_t), \varepsilon_t], \qquad \varepsilon_t \sim p(\varepsilon_t), \tag{12.1}$$

$$\alpha_{t+1} = T_t(\alpha_t) + R_t(\alpha_t)\eta_t, \qquad \eta_t \sim p(\eta_t), \tag{12.2}$$

for $t = 1, 2, \dots$, where Z_t, T_t and R_t are known matrix functions of α_t and where ε_t and η_t are disturbance series; as before, y_t and α_t are the observation

and state series. We adopt the same notation as in the previous chapters. For convenience we define a collection of state vectors by

$$\alpha_{1:t} = (\alpha_1', \ldots, \alpha_t')', \tag{12.3}$$

whereas we keep the notation $Y_t = (y_1', \ldots, y_t')'$.

In this section we discuss the use of importance sampling for the filtering of non-Gaussian and nonlinear time series which are generated by the model (12.1) and (12.2). Consider an arbitrary function $x_t(\alpha_{1:t})$ of $\alpha_{1:t}$ in (12.3); as for the smoothing case introduced in Section 11.2, most of the problems we shall consider, from both classical and Bayesian perspectives, amount essentially to the estimation of the conditional mean

$$\begin{aligned} \bar{x}_t &= \mathrm{E}[x_t(\alpha_{1:t})|Y_t] \\ &= \int x_t(\alpha_{1:t}) p(\alpha_{1:t}|Y_t) \mathrm{d}\alpha_{1:t}, \end{aligned} \tag{12.4}$$

for $t = \tau + 1, \tau + 2, \ldots$ where τ is fixed and can be zero. We shall develop recursions for computing estimates of $\bar{x}_t$ by simulation; the sample $y_1, \ldots, y_\tau$ can possibly be used as a 'start-up' sample for initialising these recursions. We consider the estimation of $\bar{x}_t$ in (12.4) by simulation based on importance sampling analogously to the estimation of the conditional smoother $\bar{x}_t$ in (11.3) of Chapter 11. We let $g(\alpha_{1:t}|Y_t)$ be an importance density that is as close as possible to $p(\alpha_{1:t}|Y_t)$ subject to the requirement that sampling from $g(\alpha_{1:t}|Y_t)$ is sufficiently practical and inexpensive.

From (12.4) we have

$$\begin{aligned} \bar{x}_t &= \int x_t(\alpha_{1:t}) \frac{p(\alpha_{1:t}|Y_t)}{g(\alpha_{1:t}|Y_t)} g(\alpha_{1:t}|Y_t) \mathrm{d}\alpha_{1:t} \\ &= \mathrm{E}_g \left[x_t(\alpha_{1:t}) \frac{p(\alpha_{1:t}|Y_t)}{g(\alpha_{1:t}|Y_t)} \right], \end{aligned} \tag{12.5}$$

where E_g denotes expectation with respect to density $g(\alpha_{1:t}|Y_t)$. Since

$$p(\alpha_{1:t}, Y_t) = p(Y_t) p(\alpha_{1:t}|Y_t),$$

we obtain

$$\bar{x}_t = \frac{1}{p(Y_t)} \mathrm{E}_g \left[x_t(\alpha_{1:t}) \tilde{w}_t \right], \tag{12.6}$$

where

$$\tilde{w}_t = \frac{p(\alpha_{1:t}, Y_t)}{g(\alpha_{1:t}|Y_t)}. \tag{12.7}$$

For notational convenience we suppress the dependence of $\tilde{w}_t$ on $\alpha_{1:t}$ and Y_t in (12.6) and later. By putting $x_t(\alpha_{1:t}) = 1$, it follows from (12.6) that $p(Y_t) = \mathrm{E}_g(\tilde{w}_t)$. As a result, (12.6) can be written as

$$\bar{x}_t = \frac{\mathrm{E}_g\left[x_t(\alpha_{1:t})\tilde{w}_t\right]}{\mathrm{E}_g(\tilde{w}_t)}. \tag{12.8}$$

We propose to estimate (12.6) by means of a random sample $\alpha_{1:t}^{(1)}, \ldots, \alpha_{1:t}^{(N)}$ drawn from $g(\alpha_{1:t}|Y_t)$. We take as our estimator

$$\begin{aligned}\hat{x}_t &= \frac{N^{-1}\sum_{i=1}^{N} x_t(\alpha_{1:t}^{(i)})\tilde{w}_t^{(i)}}{N^{-1}\sum_{i=1}^{N}\tilde{w}_t^{(i)}} \\ &= \sum_{i=1}^{N} x_t(\alpha_{1:t}^{(i)})w_t^{(i)},\end{aligned} \tag{12.9}$$

where

$$\tilde{w}_t^{(i)} = \frac{p(\alpha_{1:t}^{(i)}, Y_t)}{g(\alpha_{1:t}^{(i)}|Y_t)}, \qquad w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^{N}\tilde{w}_t^{(j)}}. \tag{12.10}$$

The values $\tilde{w}_t^{(i)}$ are called *importance weights* and the values $w_t^{(i)}$ are called *normalised importance weights*. This treatment closely follows the basic ideas of importance sampling as discussed in Section 11.2. A simple method of filtering based on importance sampling is to draw fresh random samples of $\alpha_{1:t}^{(i)}$ from a suitable $g(\alpha_{1:t}|Y_t)$ at each time point t and estimate $\bar{x}_t$ by (12.9); however, for long time series this would be unduly laborious.

12.3 Sequential importance sampling

12.3.1 Introduction

To circumvent the simple method of filtering, it seems more natural in the context of filtering to retain the previous selection of $\alpha_{1:t-1}^{(i)}$ for each i and to confine the new sampling at time t to the selection of $\alpha_t^{(i)}$ only. We call this sequential process for choosing $\alpha_{1:t}^{(i)}$ and the estimation based on it *particle filtering*; the resulting sets of values $\alpha_{1:t}^{(1)}, \ldots, \alpha_{1:t}^{(N)}$ are called *particles*; thus the ith particle at time t is defined by the relation

$$\alpha_{1:t}^{(i)} = \left(\alpha_{1:t-1}^{(i)\prime}, \alpha_t^{(i)\prime}\right)',$$

where $\alpha_{1:t-1}^{(i)}$ is the ith particle at time $t-1$. The key to a filtering method is the development of recursions for selecting $\alpha_t^{(i)}$ and for computing the corresponding

importance weights $\tilde{w}_t^{(i)}$ at time t and for $i = 1, \ldots, N$. The weight $\tilde{w}_t^{(i)}$ is a function of $\alpha_{1:t-1}^{(i)}$ and Y_{t-1} together with the current observation y_t and a newly chosen $\alpha_t^{(i)}$. A 'start-up' sample for initialising the recursions is not needed.

12.3.2 Recursions for particle filtering

A new selection of $\alpha_t^{(i)}$'s need to be consistent with draws from the importance density $g(\alpha_{1:t}^{(i)}|Y_t)$. To construct a recursion for the importance density, let us temporarily drop the i index and denote a particular value of $\alpha_{1:t}^{(i)}$ by $\alpha_{1:t}$. We have

$$\begin{aligned} g(\alpha_{1:t}|Y_t) &= \frac{g(\alpha_{1:t}, Y_t)}{g(Y_t)} \\ &= \frac{g(\alpha_t|\alpha_{1:t-1}, Y_t)g(\alpha_{1:t-1}, Y_t)}{g(Y_t)} \\ &= g(\alpha_t|\alpha_{1:t-1}, Y_t)g(\alpha_{1:t-1}|Y_t). \end{aligned} \tag{12.11}$$

Now suppose that $\alpha_{1:t-1}$ is selected using knowledge only of Y_{t-1}. Moreover, given the realised values of $\alpha_{1:t-1}$ and Y_{t-1}, the value of the observational vector y_t has already been selected by a process which does not depend on the simulated sequence $\alpha_{1:t-1}$. Under these circumstances, the density $g(\alpha_{1:t-1}|Y_{t-1})$ is not affected by including y_t in its set of conditional variables Y_{t-1}. Hence, $g(\alpha_{1:t-1}|Y_t) \equiv g(\alpha_{1:t-1}|Y_{t-1})$. By adopting this equality and reversing the order in (12.11), we obtain

$$g(\alpha_{1:t}|Y_t) = g(\alpha_{1:t-1}|Y_{t-1})g(\alpha_t|\alpha_{1:t-1}, Y_t). \tag{12.12}$$

This is the fundamental recursion which underlines the practicality of particle filtering; see, for example, Doucet, De Freitas and Gordon (2001, §1.3). It is assumed that selecting $\alpha_t^{(i)}$ from the importance density $g(\alpha_t|\alpha_{1:t-1}, Y_t)$ is practical and inexpensive. In Section 12.4 we discuss different ways to sample from such an importance density.

We now develop a recursion for calculation of the weights $w_t^{(i)}$ in (12.10) for particle filtering. From (12.7) and (12.12),

$$\begin{aligned} \tilde{w}_t &= \frac{p(\alpha_{1:t}, Y_t)}{g(\alpha_{1:t}|Y_t)} \\ &= \frac{p(\alpha_{1:t-1}, Y_{t-1})p(\alpha_t, y_t|\alpha_{1:t-1}, Y_{t-1})}{g(\alpha_{1:t-1}|Y_{t-1})g(\alpha_t|\alpha_{1:t-1}, Y_t)}. \end{aligned}$$

Due to the Markovian nature of the model (12.1) and (12.2), we have

$$p(\alpha_t, y_t|\alpha_{1:t-1}, Y_{t-1}) = p(\alpha_t|\alpha_{t-1})p(y_t|\alpha_t),$$

so that

$$\tilde{w}_t = \tilde{w}_{t-1} \frac{p(\alpha_t|\alpha_{t-1})p(y_t|\alpha_t)}{g(\alpha_t|\alpha_{1:t-1}, Y_t)}.$$

For $\tilde{w}_t^{(i)}$ in (12.10) we therefore have the recursion

$$\tilde{w}_t^{(i)} = \tilde{w}_{t-1}^{(i)} \frac{p(\alpha_t^{(i)}|\alpha_{t-1}^{(i)})p(y_t|\alpha_t^{(i)})}{g(\alpha_t^{(i)}|\alpha_{1:t-1}^{(i)}, Y_t)}, \tag{12.13}$$

from which we obtain the normalised weight $w_t^{(i)}$ by

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^N \tilde{w}_t^{(j)}}, \tag{12.14}$$

for $i = 1, \ldots, N$ and $t = \tau + 1, \tau + 2, \ldots$. The recursion (12.13) is initialised by $\tilde{w}_\tau^{(i)} = 1$ for $i = 1, \ldots, N$. We then estimate x_t by $\hat{x}_t$ from (12.9) for which a value for $\alpha_t^{(i)}$ is selected from importance density $g(\alpha_t|\alpha_{1:t-1}, Y_t)$, for $i = 1, \ldots, N$. The value $\tilde{w}_{t-1}^{(i)}$ in (12.13) can be replaced by $w_{t-1}^{(i)}$; this gives the same end result due to the normalisation. Importance sampling based on this approach is called *sequential importance sampling* (SIS). It was originally developed by Hammersley and Morton (1954) and applied to state space models by Handschin and Mayne (1969) and Handschin (1970).

12.3.3 Degeneracy and resampling

The following problem can occur in the practice of simulation when recursion (12.13) is adopted without modification. As t increases, the distribution of the weights $w_t^{(i)}$ becomes highly skewed. It is possible for all but one particle to have negligible weights, for t large. We then say that the sample has become *degenerate*. The problem of degeneracy typically occurs when the likelihood function $p(y_t|\alpha_t)$ is highly peaked relatively to the density $p(\alpha_t|\alpha_{t-1})$, with the effect that few of the values $\alpha_t^{(1)}, \ldots, \alpha_t^{(N)}$ lead to non-neglible values of $w_t^{(i)}$. It is obviously wasteful to retain particles in the recursion that are contributing negligible weights to the estimate $\hat{x}_t$ in (12.9).

A way to combat this degeneracy is to proceed as follows. We first notice that $\hat{x}_t$ in (12.9) has the form of a weighted mean of a function $x_t(\alpha_{1:t}^{(i)})$ of a random vector $\alpha_{1:t}$ which takes values $\alpha_{1:t}^{(i)}$ with probabilities $w_t^{(i)}$ for $i = 1, \ldots, N$ with $\sum_{i=1}^N w_t^{(i)} = 1$. Now take the values of $\tilde{\alpha}_{1:t}^{(i)} = \alpha_{1:t}^{(i)}$ that have a been selected as *old values*, and select *new values* $\alpha_{1:t}^{(i)}$ from the old values $\tilde{\alpha}_{1:t}^{(i)}$ with probabilities $w_t^{(1)}, \ldots, w_t^{(N)}$ with replacement. This procedure is called *resampling*; it eliminates particles that have a negligible effect on the estimate.

The sample mean of the new $x_t(\alpha_{1:t}^{(i)})$'s is the old $\hat{x}_t$ defined by (12.9). In terms of the new $\alpha_{1:t}^{(i)}$'s we therefore may consider a new estimate of x_t,

$$\hat{x}_t = \frac{1}{N}\sum_{i=1}^{N} x_t(\alpha_{1:t}^{(i)}), \tag{12.15}$$

which has the same form as (12.9) with the normalised weights reset at $w_t^{(i)} = N^{-1}$. Although resampling is a technique that combats degeneracy in particle filtering, it clearly introduces additional Monte Carlo variation into the estimate $\hat{x}_t$ as computed by (12.15). It has been shown by Chopin (2004) that the estimate $\hat{x}_t$ computed before resampling as in (12.9) is more efficient and therefore is the preferred estimate. Resampling can take place after the computation of $\hat{x}_t$.

Where α_t is one-dimensional a gain in efficiency can be achieved by employing *stratified sampling* instead of random sampling for selection of the $\alpha_t^{(i)}$'s; this was pointed out by Kitagawa (1996). Here, we merely indicate the underlying idea. Let $W(\alpha)$ be the distribution function of the discrete distribution with probabilities $w_t^{(i)}$ at old values $\alpha_t^{(i)}$, that is $W(\alpha) = \sum_i w_t^{(i)}$ for all $\alpha_t^{(i)} \leq \alpha$. Choose a value c_1 from the uniform distribution between 0 and $1/N$. Then take as the new values $\alpha_t^{(i)}$ those values of α corresponding to $W(\alpha) = c_1 + (i-1)/N$ for $i = 1, \ldots, N$. Alternative resampling schemes are the systematic resampling method of Carpenter, Clifford and Fearnhead (1999) and the residual resampling method of Liu and Chen (1998). A detailed discussion on resampling with more relevant references to the statistical literature is given by Creal (2012).

Resampling can take place at each time t but it is not necessary. When sufficient particles remain into the next period, we can proceed without the resampling step. To assess whether a sufficient number of particles has remained, Liu and Chen (1998) introduce the *effective sample size* which is given by

$$ESS = \left(\sum_{i=1}^{N} w_t^{(i)\,2} \right)^{-1},$$

where $w_t^{(i)}$ is the normalised importance weight. By construction, ESS is a value between 1 and N and it measures weight stability. When only a few weights are relatively large, the ESS is small. When the weights are uniformly distributed, the value of ESS is close to N. In practice, it is often advocated that resampling should take place when $ESS < k \cdot N$ for some fraction $k = 0.75$ or $k = 0.5$.

While resampling increases the number of effective particles, it does not eliminate the degeneracy problem altogether. Consider, for example, the simple case where $x(\alpha_{1:t}) = \alpha_t$ and a particular value of (the old) $w_t^{(i)}$ is relatively high. Then in the resampling the corresponding value of $\alpha_t^{(i)}$ will be selected a large number of times so the variance of the estimate (12.15) is accordingly high. Nevertheless,

it appears that on balance resampling is effective on a routine basis and we shall use it in all the examples considered in this chapter.

12.3.4 Algorithm for sequential importance sampling

For the implementation of sequential importance sampling, resampling does not take place with respect to the entire path $\alpha_{1:t}$ but only to the most recent value of α_t. In most cases of practical interest, the function $x_t(\alpha_{1:t})$ can be defined such that

$$x_t(\alpha_{1:t}) \equiv x_t(\alpha_t), \qquad t = 1, \ldots, n.$$

Furthermore, at time t, we are interested in estimating $x_t(\alpha_t)$ rather than $x_1(\alpha_1), \ldots, x_t(\alpha_t)$. Similarly, in the case of the Kalman filter, we also concentrate on the filtering distribution $p(\alpha_t|Y_t)$ rather than $p(\alpha_{1:t}|Y_t)$. It is shown by Chopin (2004) that the particle filter provides a consistent and asymptotically normal estimate of all moment characteristics of $p(\alpha_j|Y_t)$ only for $j = t$, not for $j < t$. In our particle filter implementations below, we therefore only resample α_t.

A formal description of the *sequential importance sampling resampling* (SISR) procedure is given by the following steps, for all $i = 1, \ldots, N$ and at fixed time t.

(i) Sample α_t : draw N values $\tilde{\alpha}_t^{(i)}$ from $g(\alpha_t|\alpha_{t-1}^{(i)}, Y_t)$ and store $\tilde{\alpha}_{t-1:t}^{(i)} = \left\{\alpha_{t-1}^{(i)}, \tilde{\alpha}_t^{(i)}\right\}$.

(ii) Weights : compute the corresponding weights $\tilde{w}_t^{(i)}$

$$\tilde{w}_t^{(i)} = \tilde{w}_{t-1}^{(i)} \frac{p(\tilde{\alpha}_t^{(i)}|\alpha_{t-1}^{(i)}) p(y_t|\tilde{\alpha}_t^{(i)})}{g(\tilde{\alpha}_t^{(i)}|\alpha_{t-1}^{(i)}, Y_t)}, \qquad i = 1, \ldots, N,$$

and normalise the weights to obtain $w_t^{(i)}$ as in (12.14).

(iii) Compute the variable of interest x_t: given the set of particles $\left\{\tilde{\alpha}_t^{(1)}, \ldots, \tilde{\alpha}_t^{(N)}\right\}$, compute

$$\hat{x}_t = \sum_{i=1}^{N} w_t^{(i)} x_t(\tilde{\alpha}_t^{(i)}).$$

(iv) Resample : draw N new independent particles $\alpha_t^{(i)}$ from $\left\{\tilde{\alpha}_t^{(1)}, \ldots, \tilde{\alpha}_t^{(N)}\right\}$ with replacement and with corresponding probabilities $\left\{w_t^{(1)}, \ldots, w_t^{(N)}\right\}$.

The procedure is recursively repeated for $t = \tau + 1, \tau + 2, \ldots, n$. The simulation from the importance density $g(\alpha_t|\alpha_{t-1}, Y_t)$ in step (i) is discussed in Section 12.4.

12.4 The bootstrap particle filter

12.4.1 Introduction

The selection of $\alpha_t^{(i)}$ given $\alpha_{1:t-1}^{(i)}$ and Y_{t-1} is crucial in particle filtering as it affects the computation of $\hat{x}_t$ and the recursive computation of the weights $\tilde{w}_t^{(i)}$, for $i = 1, \ldots, N$. In this section we present the main methodology of the particle filter based on a basic way of selecting the $\alpha_t^{(i)}$'s. The method is referrd to as the bootstrap filter. We shall focus attention on the development of the recursion (12.13), the estimation formula (12.15) and the computer algorithm for their implementations.

12.4.2 The bootstrap filter

The first particle filter to be developed was the *bootstrap filter* of Gordon, Salmond and Smith (1993). It is sometimes called the *sampling importance resampling* (SIR) filter; see, for example, Arulampalam, Maskell, Gordon and Clapp (2002). The key to the construction of a particle filter that relies on the weight recursion (12.13) is the choice of the importance density $g(\alpha_t|\alpha_{1:t-1}, Y_t)$. Since the joint density of α_t and y_t given $\alpha_{1:t-1}$ and Y_{t-1} depends only on α_{t-1}, we can restrict ourselves to importance densities of the form $g(\alpha_t|\alpha_{t-1}, y_t)$. Ideally we would take $g(\alpha_t|\alpha_{t-1}, y_t) = p(\alpha_t|\alpha_{t-1}, y_t)$ but this is normally not available in analytical form and so we look for approximations to it. The bootstrap filter takes the importance density as

$$g(\alpha_t|\alpha_{1:t-1}, y_t) = p(\alpha_t|\alpha_{t-1}). \tag{12.16}$$

At first sight this looks crude since it neglects relevant information in y_t but when used with resampling and with N large enough, it can work well in many cases of interest and it is widely used in practice. For the bootstrap filter, the recursion (12.13) therefore reduces to the simple form

$$\tilde{w}_t^{(i)} = \tilde{w}_{t-1}^{(i)} p(y_t|\alpha_t^{(i)}).$$

We employ resampling at each time $t = 1, \ldots, n$. Since the weights are reset after the resampling of $\alpha_{t-1}^{(i)}$ at $w_{t-1}^{(i)} = 1/N$, the normalised weight becomes simply

$$w_t^{(i)} = \frac{p(y_t|\alpha_t^{(i)})}{\sum_{j=1}^N p(y_t|\alpha_t^{(j)})}, \qquad i = 1, \ldots, N. \tag{12.17}$$

The observation density $p(y_t|\alpha_t^{(i)})$ is evaluated straightforwardly for given $\alpha_t^{(i)}$.

12.4.3 Algorithm for bootstrap filter

Suppose that we have a *start up sample* $y_1, \ldots, y_\tau$ and that we intend to begin particle filtering at time $\tau + 1$. The algorithm for the recursive implementation of the filter then proceeds as follows, for all $i = 1, \ldots, N$ and at fixed time t.

(i) Sample α_t : draw N values $\tilde{\alpha}_t^{(i)}$ from $p(\alpha_t|\alpha_{t-1}^{(i)})$.

(ii) Weights : compute the corresponding weights $\tilde{w}_t^{(i)}$

$$\tilde{w}_t^{(i)} = p(y_t|\tilde{\alpha}_t^{(i)}), \qquad i = 1, \ldots, N,$$

and normalise the weights to obtain $w_t^{(i)}$ as in (12.14).

(iii) Compute the variable of interest x_t: given the set of particles $\left\{\tilde{\alpha}_t^{(1)}, \ldots, \tilde{\alpha}_t^{(N)}\right\}$, compute

$$\hat{x}_t = \sum_{i=1}^{N} w_t^{(i)} x_t(\tilde{\alpha}_t^{(i)}).$$

(iv) Resample : draw N new independent particles $\alpha_t^{(i)}$ from $\left\{\tilde{\alpha}_t^{(1)}, \ldots, \tilde{\alpha}_t^{(N)}\right\}$ with replacement and with corresponding probabilities $\left\{w_t^{(1)}, \ldots, w_t^{(N)}\right\}$.

We repeat these step for $t = \tau+1, \tau+2, \ldots$. The advantages of this filter are that it is quick and easy to operate and requires very little storage. The drawback is that if the likelihood $p(y_t|\alpha_t)$ is sharply peaked relative to the density $p(\alpha_t|\alpha_{t-1})$ there may be many repetitions of the likelier particles in step (iv), thus reducing the effective number of particles under consideration. A further weakness of the bootstrap filter arises from the fact that value of y_t is not taken into account in the selection of $\alpha_t^{(i)}$ in step (i) of the algorithm. Ways of taking y_t into account are proposed in the next subsections.

12.4.4 Illustration: local level model for Nile data

To illustrate the accuracy of the bootstrap filter for the local level model (2.3) and to show the importance of resampling, we have considered the Nile time series from Chapter 2. The local level model is $y_t = \alpha_t + \varepsilon_t$ with $\alpha_{t+1} = \alpha_t + \eta_t$ and $\alpha_1 \sim \mathrm{N}(a_1, P_1)$ for $t = 1, \ldots, n$. The disturbances $\varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2)$ and $\eta_t \sim \mathrm{N}(0, \sigma_\eta^2)$ are mutually and serially independent. The filtered estimate of α_t using observations $y_1, \ldots, y_t$ together with its error variance, $a_{t|t}$ and $P_{t|t}$, respectively, are routinely computed by the Kalman filter (2.15). Alternatively, the filtered state estimate and its error variance can be computed by the bootstrap filter and we can compare its accuracy relative to the output of the Kalman filter.

We first implement the bootstrap filter without the resampling step (iv) and with $N = 10,000$. The variances σ_ε^2 and σ_η^2 of the local level model are set equal to their maximum likelihood estimates $15,099$ and 1469.1 from Subsection 2.10.3, respectively. The remaining first three steps remain. The bootstrap filter for the local level model is given by

(i) Draw N values $\tilde{\alpha}_t^{(i)} \sim \mathrm{N}(\alpha_{t-1}^{(i)}, \sigma_\eta^2)$.

(ii) Compute the corresponding weights $\tilde{w}_t^{(i)}$

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \exp\left(-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma_\varepsilon^2 - \frac{1}{2}\sigma_\varepsilon^{-2}(y_t - \tilde{\alpha}_t^{(i)})\right), \quad i = 1, \ldots, N,$$

and normalise the weights to obtain $w_t^{(i)}$ as in (12.14).

(iii) Compute

$$\hat{a}_{t|t} = \sum_{i=1}^{N} w_t^{(i)} \tilde{\alpha}_t^{(i)}, \qquad \hat{P}_{t|t} = \sum_{i=1}^{N} w_t^{(i)} \tilde{\alpha}_t^{(i)\,2} - \hat{a}_{t|t}^2.$$

(iv) Set $\alpha_t^{(i)} = \tilde{\alpha}_t^{(i)}$ for $i = 1, \ldots, N$ (without resampling).

In Fig. 12.1 we present the Nile data y_t, the filtered state estimate $\hat{a}_{t|t}$ and the corresponding confidence interval based on $\hat{P}_{t|t}$ for $t = 1, \ldots, n$. The panels (ii)

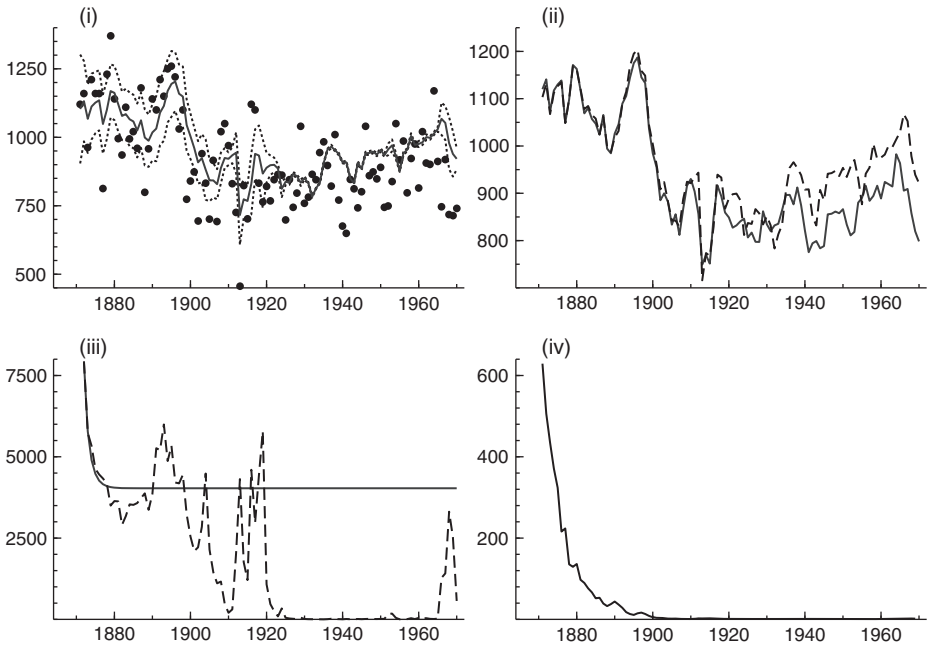

Fig. 12.1 Bootstrap filter with $N = 10,000$ **without** resampling for Nile data: (i) data (dots), filtered estimate $\hat{a}_{t|t}$ and its 90% confidence intervals; (ii) $\hat{a}_{t|t}$ (dashed) and $a_{t|t}$ (solid) from the Kalman filter; (iii) $\hat{P}_{t|t}$ and $P_{t|t}$; (iv) effective sample size ESS.

and (iii) present $a_{t|t}$ from the Kalman filter with $\hat{a}_{t|t}$ and $P_{t|t}$ from the Kalman filter with $\hat{P}_{t|t}$ for $t = 1, \ldots, n$. It is clear that the differences are large and hence the bootstrap filter without resampling is inaccurate. The ESS time series in panel (iv) confirms this finding.

In Fig. 12.2 we present the same output as in Fig. 12.1 but now for the bootstrap filter with step (iv) replaced by the stratified resampling step proposed by Kitagawa (1996), that is

(iv) Select N new independent particles $\alpha_t^{(i)}$ using stratified sampling.

Panel (i) of Fig. 12.2 presents the filtered state estimate and the 90% confidence interval. From panel (ii) we learn that the filtered estimate $\hat{a}_{t|t}$ is virtually the same as $a_{t|t}$ from the Kalman filter. Also the variances in panel (iii) are very close to each other. We can conclude that resampling takes a crucial role in the bootstrap filter. The ESS time series in panel (iv) confirms this finding. The number of relevant particles remains high throughout the time series.

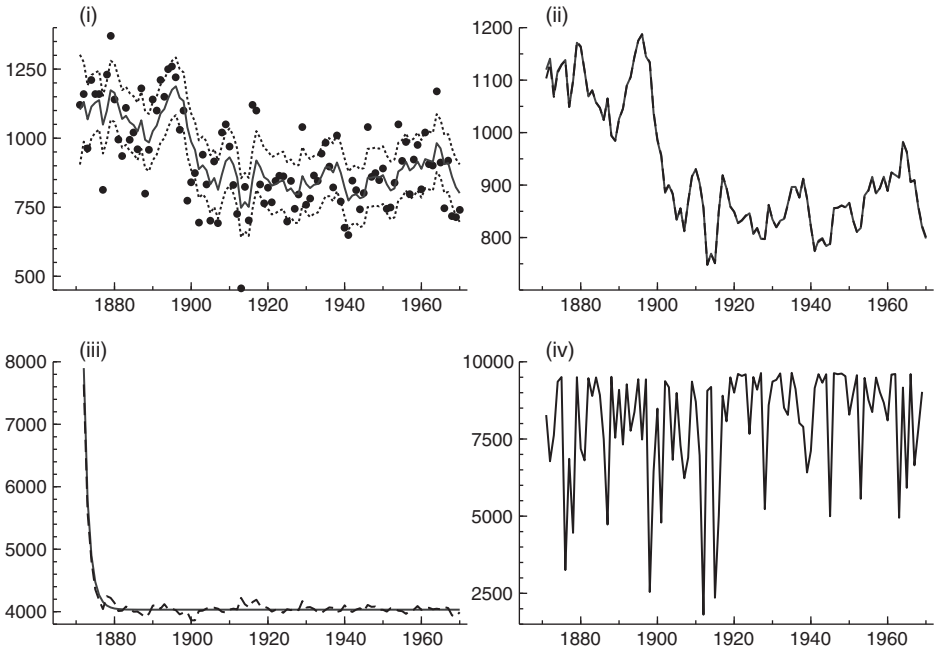

Fig. 12.2 Bootstrap filter with $N = 10,000$ **with** resampling for Nile data: (i) data (dots), filtered estimate $\hat{a}_{t|t}$ and its 90% confidence intervals; (ii) $\hat{a}_{t|t}$ (dashed) and $a_{t|t}$ (solid) from the Kalman filter; (iii) $\hat{P}_{t|t}$ and $P_{t|t}$; (iv) effective sample size ESS.

12.5 The auxiliary particle filter

The auxiliary particle filter was proposed by Pitt and Shephard (1999) as a development of the bootstrap filter. They gave a general treatment of the method whereas we will present only a simplified version below. The aim of their proposal is to reduce the number of repetitions in the final stage by introducing an additional selection stage involving the value of y_t. A modification of the auxiliary particle filter is considered by Johansen and Doucet (2008).

12.5.1 Algorithm for auxiliary filter

The primary modification of the auxiliary particle filter is based on finding an effective importance density at time t that also considers the knowledge of y_t for selecting new particles. We assume that the particles $\alpha_{1:t-1}^{(i)}$ are available at time $t-1$ with weights $w_{t-1}^{(i)} = 1/N$ for $i = 1, \ldots, N$. The auxiliary particle filter method proceeds as follows, for all $i = 1, \ldots, N$ and at fixed time t.

(i) Predict α_t: for each value $\alpha_{t-1}^{(i)}$, predict the corresponding value for α_t using a deterministic function (no sampling) and denote the prediction by $\alpha_t^{*(i)}$.

(ii) Intermediate weights: compute weights corresponding to prediction $\alpha_t^{*(i)}$,

$$\tilde{w}_t^{*(i)} = g(y_t|\alpha_t^{*(i)}) w_{t-1}^{(i)},$$

and normalise to obtain $w_t^{*(i)}$.

(iii) Resample: draw N new particles $\tilde{\alpha}_{t-1}^{(i)}$, with replacement, from $\alpha_{t-1}^{(1)}, \ldots, \alpha_{t-1}^{(N)}$ with probabilities $w_t^{*(1)}, \ldots, w_t^{*(N)}$.

(iv) Sample α_t: draw N values $\alpha_t^{(i)}$ from $g(\alpha_t|\tilde{\alpha}_{t-1}^{(i)})$.

(v) Weights: compute the corresponding weights $\tilde{w}_t^{(i)}$

$$\tilde{w}_t^{(i)} = \frac{p(y_t|\alpha_t^{(i)}) p(\alpha_t^{(i)}|\tilde{\alpha}_{t-1}^{(i)})}{g(y_t|\alpha_t^{*(i)}) g(\alpha_t^{(i)}|\tilde{\alpha}_{t-1}^{(i)})}, \qquad i = 1, \ldots, N,$$

and normalise the weights to obtain $w_t^{(i)}$.

(vi) Compute the variable of interest x_t : given the set of particles $\left\{\alpha_t^{(1)}, \ldots, \alpha_t^{(N)}\right\}$, compute

$$\hat{x}_t = \sum_{i=1}^{N} w_t^{(i)} x_t(\alpha_t^{(i)}).$$

In step (i) we need to predict the state vector α_t given α_{t-1}. For the nonlinear state equation $\alpha_{t+1} = T_t(\alpha_t) + R_t(\eta_t)$, we can simply take $\alpha_t^{*(i)} = T_{t-1}(\alpha_{t-1}^{(i)})$ as

our state prediction for $i = 1, \ldots, N$. Pitt and Shephard (1999) originally proposed the above initial prediction; Johansen and Doucet (2008) have considered other prediction choices and their statistical performances. The steps (iv), (v) and (vi) are essentially the same as the steps (i), (ii) and (iii) of the bootstrap filter algorithm, respectively.

Since the importance density $g(\alpha_t|\alpha_{t-1})$ is continuous there will, with probability one, be no omissions or repetitions in the values of $\alpha_t^{\sigma(i)}$ chosen in step (iv). Since the distribution of weights $w_t^{\sigma(i)}$ can be expected to be relatively lightly skewed, there should be fewer omissions and repetitions of values of $\alpha_t^{(i)}$ in the particles obtained from the auxiliary particle filter, as compared with those obtained from bootstrap filter. On the other hand, because of the way the $\alpha_{t-1}^{(i)}$'s are selected in step (iii), the distribution of the components $\alpha_{t-1}^{(i)}$ for the particles $\alpha_t^{(i)}$ should be comparable with those for the bootstrap filter.

12.5.2 Illustration: local level model for Nile data

To illustrate the possible gains in accuracy of the auxiliary particle filter in comparison with the bootstrap filter, we continue our illustration for the local level model (2.3) applied to the Nile time series from Chapter 2. The local level model is $y_t = \alpha_t + \varepsilon_t$ with $\alpha_{t+1} = \alpha_t + \eta_t$ and $\alpha_1 \sim \mathrm{N}(a_1, P_1)$ for $t = 1, \ldots, n$. The disturbances $\varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2)$ and $\eta_t \sim \mathrm{N}(0, \sigma_\eta^2)$ are mutually and serially independent. We aim to compute the filtered estimate of α_t using observations Y_t together with its error variance. The estimates of σ_ε^2 and σ_η^2 are obtained from the maximum likelihood method as reported in Subsection 2.10.3. The purpose is to compare the performances of the bootstrap and the auxiliary particle filters in computing the filtered estimates of α_t. In panels (i) and (iii) of Fig. 12.3 we present the output of the bootstrap and auxiliary filters, respectively. We can conclude that differences in the graphical output cannot be detected. In panels (ii) and (iv) the ESS for each time period of the bootstrap and auxiliary filters are presented, respectively. We have shown that the effective sample size of the auxiliary particle filter is higher for all time periods. In time periods where the ESS is relatively low for the bootstrap filter, the ESS for the auxiliary filter is at least twice as high as the ESS for the bootstrap filter.

12.6 Other implementations of particle filtering

We investigate other strategies for the selection of $\alpha_t^{(i)}$ given $\alpha_{1:t-1}^{(i)}$ and Y_{t-1} next.

12.6.1 Importance density from extended or unscented filter

The importance density $g(\alpha_t|\alpha_{t-1}, Y_t)$ can also be obtained from the extended Kalman filter or the unscented Kalman filter. Different ways of incorporating the ideas behind the extended or unscented filter on the one hand and particle

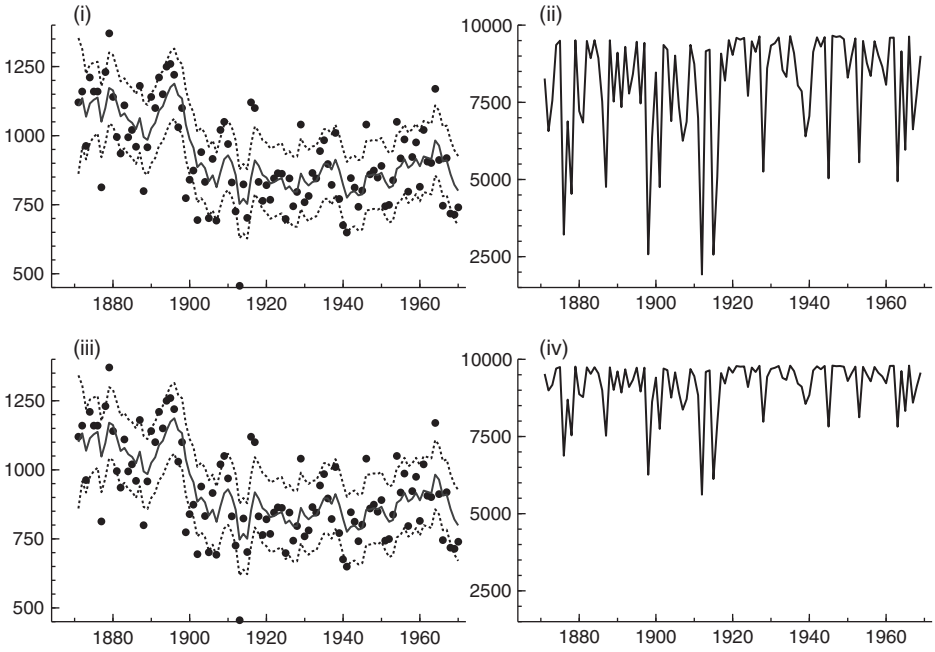

Fig. 12.3 Bootstrap filter versus auxiliary particle filter with $N = 10,000$ for Nile data: (i) data (dots), filtered estimate $\hat{a}_{t|t}$ and its 90% confidence intervals from bootstrap filter; (ii) effective sample size ESS for bootstrap filter; (iii) data (dots), filtered estimate $\hat{a}_{t|t}$ and its 90% confidence intervals from auxiliary filter; (iv) effective sample size ESS for auxiliary filter.

filtering on the other hand can be considered; see van der Merwe, Doucet and de Freitas (2000) for the first of such implementations.

In the treatment given below, we explore an alternative implementation in which we aim to incorporate y_t into the proposal density. Since the approximate filtering equations can be regarded as Taylor expansions of a certain order of $p(\alpha_t|\alpha_{t-1}, Y_t)$, we can expect such an importance density to be accurate. The extended or unscented filters can be introduced in different ways as we will discuss below.

The set of particles $\alpha_{t-1}^{(1)}, \ldots, \alpha_{t-1}^{(N)}$ provide information about the distribution of α_{t-1} given Y_{t-1} including its mean and variance, that is

$$\bar{a}_{t-1|t-1}^{+} = N^{-1} \sum_{i=1}^{N} \alpha_{t-1}^{(i)},$$

$$\bar{P}_{t-1|t-1}^{+} = N^{-1} \sum_{i=1}^{N} \left(\alpha_{t-1}^{(i)} - \bar{a}_{t-1|t-1}^{+}\right) \left(\alpha_{t-1}^{(i)} - \bar{a}_{t-1|t-1}^{+}\right)',$$

where $\bar{a}^{+}_{t-1|t-1}$ is the particle filter estimate of $\mathrm{E}(\alpha_{t-1}|Y_{t-1})$ and $\bar{P}^{+}_{t-1|t-1}$ is the variance estimate of $\mathrm{Var}(\alpha_{t-1}|Y_{t-1})$. For an appropriate Gaussian importance density $g(\alpha_t|\alpha_{t-1}, Y_t)$ we require estimates for $\mathrm{E}(\alpha_t|Y_t)$ and $\mathrm{Var}(\alpha_t|Y_t)$. We can obtain these via the extended Kalman filter or the unscented filter in combination with estimates such as $\bar{a}^{+}_{t-1|t-1}$ and $\bar{P}^{+}_{t-1|t-1}$. When considering the extended Kalman filter (10.4), we require estimates for $K_t = M_t F_t^{-1}$ and F_t in (10.4). For the unscented filter we require similar quantities but these are denoted by $P_{\alpha v,t}$, instead of M_t, and $P_{vv,t}$, instead of F_t; see equations (10.10) and (10.11). For the set of particles $\alpha^{(i)}_{t-1}$, $i = 1, \ldots, N$, estimators for $M_t = \mathrm{Cov}(\alpha_t, y_t|Y_{t-1})$ and $F_t = \mathrm{Var}(y_t|Y_{t-1})$ are obviously given by

$$\bar{M}^{+}_t = N^{-1} \sum_{i=1}^{N} \left(a^{+(i)}_t - \bar{a}^{+}_t\right)\left(v^{+(i)}_t - \bar{v}^{+}_t\right)',$$

$$\bar{F}^{+}_t = N^{-1} \sum_{i=1}^{N} \left(v^{+(i)}_t - \bar{v}^{+}_t\right)\left(v^{+(i)}_t - \bar{v}^{+}_t\right)',$$

where

$$a^{+(i)}_t = T_{t-1}(\alpha^{(i)}_{t-1}) + R_{t-1}(\alpha^{(i)}_{t-1})\eta^{(i)}_{t-1}, \quad \eta^{(i)}_{t-1} \sim \left[0, Q_{t-1}(\alpha^{(i)}_{t-1})\right],$$

$$\bar{a}^{+}_t = N^{-1} \sum_{i=1}^{N} a^{+(i)}_t,$$

and

$$v^{+(i)}_t = y_t - \mathrm{E}[y_t|a^{*(i)}_t, \varepsilon^{(i)}_t], \quad \varepsilon^{(i)}_t \sim \left[0, H_t(a^{*(i)}_t)\right], \quad \bar{v}^{+}_t = N^{-1} \sum_{i=1}^{N} v^{+(i)}_t,$$

with $a^{*(i)}_t = T_{t-1}(\alpha^{(i)}_{t-1})$ and expectation $\mathrm{E}(y_t|\alpha_t, \varepsilon_t)$ is with respect to $p[y_t|Z_t(\alpha_t), \varepsilon_t]$ for $\alpha_t = a^{*(i)}_t$ and $\varepsilon_t = \varepsilon^{(i)}_t$, for $i = 1, \ldots, N$. The variance matrix $H_t(\alpha_t)$ is with respect to $p(\varepsilon_t)$ or is a first-order Taylor approximation to the variance of $p(\varepsilon_t)$ at $\alpha_t = a^{*(i)}_t$. For the extended Kalman filter update in the particle filter, we define

$$\bar{K}^{+}_t = \bar{M}^{+}_t \bar{F}^{+\,-1}_t,$$

and compute

$$\bar{a}^{+}_{t|t} = N^{-1} \sum_{i=1}^{N} a^{+(i)}_{t|t}, \qquad \bar{P}^{+}_{t|t} = N^{-1} \sum_{i=1}^{N} \left(a^{+(i)}_{t|t} - \bar{a}^{+}_{t|t}\right)\left(a^{+(i)}_{t|t} - \bar{a}^{+}_{t|t}\right)',$$

where

$$a_{t|t}^{+(i)} = a_t^{+(i)} + \bar{K}_t^+ v_t^{+(i)}, \qquad i = 1, \ldots, N.$$

We can set the importance density equal to

$$g(\alpha_t|\alpha_{t-1}, Y_t) = \mathrm{N}\left(a_{t|t}^{+(i)}, \bar{P}_{t|t}^+\right), \tag{12.18}$$

which we expect to be an accurate approximation of $p(\alpha_t|\alpha_{t-1}, Y_t)$. Next, the particle filter steps of the algorithm in Subsection 12.3.4 can be carried out with $g(\alpha_t|\alpha_{1:t-1}, Y_t) = g(\alpha_t|\alpha_{t-1}, Y_t)$ in step (i) given by (12.18). Alternative estimates for $\bar{a}_{t|t}^+$ and $\bar{P}_{t|t}^+$ can be obtained by considering weighted sample averages based on normalised weights $p(y_t|\alpha_t)$ evaluated at $\alpha_t = a_{t|t}^{+(i)}$ for $i = 1, \ldots, N$.

The mean, variance and covariance estimates as given above are subject to Monte Carlo error and require computational effort. An alternative is to use the extended or unscented Kalman filter approximations to obtain these estimates. In the case of the extended filter, the mean and variance estimates $a_{t|t}^+$ and $P_{t|t}^+$ can be obtained by $a_{t|t}$ and $P_{t|t}$, respectively, from (10.4) with

$$a_t = N^{-1} \sum_{i=1}^{N} T_{t-1}(\alpha_{t-1}^{(i)}),$$

and

$$P_t = N^{-1} \sum_{i=1}^{N} \left[T_{t-1}(\alpha_{t-1}^{(i)}) - a_t\right] \left[T_{t-1}(\alpha_{t-1}^{(i)}) - a_t\right]' + R_{t-1}(\alpha_{t-1}^{(i)}) Q_{t-1}(\alpha_{t-1}^{(i)}) T_{t-1}(\alpha_{t-1}^{(i)})'.$$

In the case of the unscented filter, the mean and variance estimates $a_{t|t}^+$ and $P_{t|t}^+$ can be obtained by $a_{t|t}$ from (10.10) and $P_{t|t}$ from (10.11), respectively, where a_t and P_t can be computed as above.

The importance density (12.18) can be incorporated in both algorithms of Subsections 12.3.4 and 12.5.1. Step (i) of the algorithm in Subsection 12.3.4 can directly be based on (12.18). In the algorithm in Subsection 12.5.1, $\mu_t^{(i)}$ can be redefined in step (i) as a draw from importance density (12.18). Such modified algorithms are expected to be more effective since the information of y_t is taken into account when a new state variable α_t is generated.

12.6.2 The local regression filter

This filter can be used for the class of non-Gaussian state space models of the form

$$\begin{aligned} y_t &= Z_t \alpha_t + \varepsilon_t, & \varepsilon_t &\sim p(\varepsilon_t), \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t, & \eta_t &\sim p(\eta_t), \end{aligned} \tag{12.19}$$

where $\mathrm{E}(\varepsilon_t) = \mathrm{E}(\eta_t) = 0$, $\mathrm{V}(\varepsilon_t) = H_t$ and $\mathrm{V}(\eta_t) = Q_t$ for $t = 1, 2, \ldots$; matrices Z_t, T_t and R_t are assumed known. As before we would ideally like to take the importance density $g(\alpha_t|\alpha_{1:t-1}, Y_t)$ equal to $p(\alpha_t|\alpha_{t-1}, y_t)$ but this is not normally available in an analytical form. A further complication is that for many practical cases the dimension of α_t is greater than that of η_{t-1} so the conditional density of α_t given α_{t-1} is singular. For this reason it is more convenient to work with $(\alpha_{1:t-1}, \eta_{t-1})$ than working with $\alpha_{1:t}$ throughout. The idea behind the method is to obtain an importance density by approximating the conditional density of η_{t-1} given $\alpha_{1:t-1}$ and Y_t by $\mathrm{N}(b_t, c_t)$ where $b_t = \mathrm{E}(\eta_{t-1}|\alpha_{t-1}, y_t)$ and $c_t = \mathrm{V}(\eta_{t-1}|\alpha_{t-1}, y_t)$. The approximation does not need to be highly accurate. The advantage of the approach is that it takes the values of y_t explicitly into account in the selection of α_t.

Transform from $\alpha_{1:t}$ to $(\alpha_{1:t-1}, \eta_{t-1})$ and define $x_t^*(\alpha_{1:t-1}, \eta_{t-1}) = x_t(\alpha_{1:t})$. Analogously to (12.5) and (12.6), we have

$$\begin{aligned}\bar{x}_t &= \mathrm{E}(x_t^*(\alpha_{1:t-1}, \eta_{t-1})|Y_t) \\ &= \int x_t^*(\alpha_{1:t-1}, \eta_{t-1}) p(\alpha_{1:t-1}, \eta_{t-1}|Y_t) \mathrm{d}(\alpha_{1:t-1}, \eta_{t-1}) \\ &= \frac{1}{p(Y_t)} \mathrm{E}_g[x_t^*(\alpha_{1:t-1}, \eta_{t-1}) \tilde{w}_t], \end{aligned} \tag{12.20}$$

where E_g denotes expectation with respect to importance density $g(\alpha_{1:t-1}, \eta_{t-1}|Y_t)$ and

$$\tilde{w}_t = \frac{p(\alpha_{1:t-1}, \eta_{t-1}, Y_t)}{g(\alpha_{1:t-1}, \eta_{t-1}|Y_t)}. \tag{12.21}$$

Now

$$\begin{aligned} p(\alpha_{1:t-1}, \eta_{t-1}, Y_t) &= p(\alpha_{1:t-1}, Y_{t-1}) p(\eta_{t-1}, y_t|\alpha_{1:t-1}, Y_{t-1}) \\ &= p(\alpha_{1:t-1}, Y_{t-1}) p(\eta_{t-1}) p(y_t|\alpha_t), \end{aligned}$$

with $\alpha_t = T_{t-1}\alpha_{t-1} + R_{t-1}\eta_{t-1}$, due to the Markovian structure of model (12.1) and (12.2). Analogously to (12.12), for particle filtering we must have

$$g(\alpha_{1:t-1}, \eta_{t-1}|Y_t) = g(\alpha_{1:t-1}|Y_{t-1}) g(\eta_{t-1}|\alpha_{1:t-1}, Y_t).$$

Assuming that importance densities $g(\cdot)$ have similar Markovian structure to that of model (12.1) and (12.2) we take $g(\eta_{t-1}|\alpha_{1:t-1}, Y_t) = g(\eta_{t-1}|\alpha_{t-1}, y_t)$. We therefore have from (12.21)

$$\tilde{w}_t = \frac{p(\alpha_{1:t-1}, Y_{t-1})}{g(\alpha_{1:t-1}|Y_{t-1})} \frac{p(\eta_{t-1}) p(y_t|\alpha_t)}{g(\eta_{t-1}|\alpha_{t-1}, y_t)}.$$

The Jacobians of the transformation from $\alpha_{1:t-1}, Y_{t-1}$ to $\alpha_{1:t-2}, \eta_{t-2}, Y_{t-1}$ cancel out so

$$\tilde{w}_t = \tilde{w}_{t-1} \frac{p(\eta_{t-1}) p(y_t|\alpha_t)}{g(\eta_{t-1}|\alpha_{t-1}, y_t)}, \tag{12.22}$$

with $\alpha_t = T_{t-1}\alpha_{t-1} + R_{t-1}\eta_{t-1}$.

Putting $x^*(\alpha_{1:t-1}, \eta_{t-1}) = 1$ in (12.20) gives $p(Y_t) = \mathrm{E}_g(\tilde{w}_t)$ so we have

$$\bar{x}_t = \frac{\mathrm{E}_g[x^*(\alpha_{1:t-1}, \eta_{t-1})\tilde{w}_t]}{\mathrm{E}_g[\tilde{w}_t]}. \tag{12.23}$$

Taking a random sample $\eta_{t-1}^{(i)}$ from $g(\eta_{t-1})$ and computing $\alpha_t^{(i)} = T_{t-1}\alpha_{t-1}^{(i)} + R_{t-1}\eta_{t-1}^{(i)}$ for $i = 1, \ldots, N$, our estimate of x_t is

$$\hat{x}_t = \sum_{i=1}^{N} x_t(\alpha_{1:t}^{(i)}) w_t^{(i)},$$

as in (12.9), where

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^{N} \tilde{w}_t^{(i)}}, \qquad i = 1, \ldots, N, \tag{12.24}$$

with the recursion for $\tilde{w}_t^{(i)}$ given by

$$\tilde{w}_t^{(i)} = \tilde{w}_{t-1}^{(i)} \frac{p(\eta_{t-1}^{(i)}) p(y_t|\alpha_t^{(i)})}{g(\eta_{t-1}^{(i)}|\alpha_{t-1}^{(i)}, y_t)}.$$

To construct the importance density $g(\eta_{t-1}|\alpha_{t-1}, y_t)$ we postulate that for $\alpha_{1:t-1}$ and Y_{t-1} fixed, η_{t-1} and y_t are generated by the linear Gaussian model

$$\begin{array}{rcl@{\qquad}rcl} \eta_{t-1} & \sim & \mathrm{N}(0, Q_{t-1}), & \alpha_t & = & T_{t-1}\alpha_{t-1} + R_{t-1}\eta_{t-1}, \\ \varepsilon_t & \sim & \mathrm{N}(0, H_t), & y_t & = & Z_t\alpha_t + \varepsilon_t. \end{array}$$

We then choose η_{t-1} from the conditional distribution of η_{t-1} given α_{t-1} and y_t; this is a normal distribution which we denote by $N(b_t, c_t)$.

We have

$$\mathrm{E}(y_t|\alpha_{t-1}) = Z_t T_{t-1}\alpha_{t-1}, \qquad \mathrm{V}(y_t|\alpha_{t-1}) = W_t = Z_t R_{t-1} Q_{t-1} R_{t-1}' Z_t' + H_t.$$

Further, $\mathrm{E}(\eta_{t-1}y_t') = Q_{t-1}R_{t-1}'Z_t'$. It follows from elementary regression theory that

$$\begin{array}{rcl} b_t & = & \mathrm{E}(\eta_{t-1}|\alpha_{t-1}, y_t) = Q_{t-1}R_{t-1}'Z_t'W_t^{-1}(y_t - Z_t T_{t-1}\alpha_{t-1}), \\ c_t & = & \mathrm{V}(\eta_{t-1}|\alpha_{t-1}, y_t) = Q_{t-1} - Q_{t-1}R_{t-1}'Z_t'W_t^{-1}Z_t R_{t-1}Q_{t-1}. \end{array} \tag{12.25}$$

The simulation steps are given by the following algorithm:

(i) Draw independent values $\eta_{t-1}^{(1)}, \ldots, \eta_{t-1}^{(N)}$ from $\mathrm{N}(b_t, c_t)$ and compute $\alpha_t^{(i)} = T_{t-1}\alpha_{t-1}^{(i)} + R_{t-1}\eta_{t-1}^{(i)}$ for $i = 1, \ldots, N$.

(ii) Compute normalised weights $w_t^{(1)}, \ldots, w_t^{(N)}$ as in (12.24) and resample with replacement using these weights.

(iii) Relabelling the resampled values as $\alpha_t^{(i)}$ and resetting the weights as $w_t^{(i)} = 1/N$, compute $\bar{x}_t = \frac{1}{N}\sum_{i=1}^{N} x(\alpha_{1:t})$.

12.6.3 The mode equalisation filter

This filter is intended for models of the form (9.1). We shall aim at obtaining a more accurate Gaussian approximation to $p(\eta_{t-1}|\alpha_{t-1}, y_t)$ than was used for the local regression filter. We do this by employing the method used in Section 10.6 for smoothing, that is, we choose the Gaussian approximating density which has the same mode as $p(\eta_{t-1}|\alpha_{t-1}, y_t)$. The construction of a similar importance density based on the mode is also explored by Doucet, Godsill and Andrieu (2000) and Cappé, Moulines and Rydén (2005, Chapter 7).

Instead of an importance density of the form $g(\eta_{t-1}|\alpha_{t-1}, y_t)$ we introduce a modified version $\tilde{y}_t$ of y_t and take as our importance density $g(\eta_{t-1}|\alpha_{t-1}, \tilde{y}_t)$ where η_{t-1} and $\tilde{y}_t$ are generated by the linear Gaussian model

$$\begin{aligned} \eta_{t-1} &\sim N(0, \tilde{Q}_{t-1}), & \varepsilon_t &\sim N(0, \tilde{H}_t), \\ \alpha_t &= T_{t-1}\alpha_{t-1} + R_{t-1}\eta_{t-1}, & \tilde{y}_t &= Z_t\alpha_t + \varepsilon_t. \end{aligned}$$

We shall determine $\tilde{y}_t$, $\tilde{Q}_{t-1}$ and $\tilde{H}_t$ so that $g(\eta_{t-1}|\alpha_{t-1}, \tilde{y}_t)$ and $p(\eta_{t-1}|\alpha_{t-1}, y_t)$ have the same mode to a good enough approximation.

The mode of density $g(\eta_{t-1}|\alpha_{t-1}, \tilde{y}_t)$ is the solution of the equation

$$\frac{\partial \log g(\eta_{t-1}|\alpha_{t-1}, \tilde{y}_t)}{\partial \eta_{t-1}} = 0.$$

Since $\log g(\eta_{t-1}|\alpha_{t-1}, \tilde{y}_t) = \log g(\eta_{t-1}, \tilde{y}_t|\alpha_{t-1}) - \log g(\tilde{y}_t|\alpha_{t-1})$, it can also be obtained as

$$\frac{\partial \log g(\eta_{t-1}, \tilde{y}_t|\alpha_{t-1})}{\partial \eta_{t-1}} = 0.$$

We have

$$\log g(\eta_{t-1}, \tilde{y}_t|\alpha_{t-1}) = \text{constant} - \frac{1}{2}[\eta_{t-1}'\tilde{Q}_t^{-1}\eta_{t-1} + (\tilde{y}_t - Z_t\alpha_t)'\tilde{H}_t^{-1}(\tilde{y}_t - Z_t\alpha_t)],$$

where $\alpha_t = T_{t-1}\alpha_{t-1} + R_{t-1}\eta_{t-1}$. Differentiating and equating to zero gives

$$-\tilde{Q}_{t-1}^{-1}\eta_{t-1} + R_t'Z_t'\tilde{H}_t^{-1}(\tilde{y}_t - Z_t\alpha_t) = 0. \tag{12.26}$$

The mode of a Gaussian density is equal to the mean. Therefore, at first sight the solution for η_{t-1} of (12.26) appears inconsistent with the expression for b_t in (12.25). For the two solutions to be equal we need to have

$$\begin{aligned}&\tilde{Q}_{t-1}R'_{t-1}Z'_t(Z_tR_{t-1}\tilde{Q}_{t-1}R'_{t-1}Z'_t+\tilde{H}_t)^{-1}\\&\quad=(\tilde{Q}_{t-1}^{-1}+R'_{t-1}Z'_t\tilde{H}_t^{-1}Z_tR_{t-1})^{-1}R'_{t-1}Z'_t\tilde{H}_t^{-1}.\end{aligned}$$

The equality can be verified by premultiplying the equation by $\tilde{Q}_{t-1}^{-1}+R'_{t-1}Z'_t\tilde{H}_t^{-1}Z_tR_{t-1}$ and postmultiplying by $Z_tR_{t-1}\tilde{Q}_{t-1}R'_{t-1}Z'_t+\tilde{H}_t$.

Denote $Z_t\alpha_t$ by θ_t. For the model

$$p(y_t|\theta_t),\qquad \theta_t=Z_t\alpha_t,\qquad \alpha_t=T_{t-1}\alpha_{t-1}+R_{t-1}\eta_{t-1},\qquad \eta_{t-1}\sim p(\eta_{t-1}),$$

the mode of η_{t-1} given α_{t-1} and y_t is the solution of the equation

$$\frac{\partial\log p(\eta_{t-1},y_t|\alpha_{t-1})}{\partial\eta_{t-1}}+\frac{\partial\log p(\eta_{t-1})}{\partial\eta_{t-1}}+R'_{t-1}Z'_t\frac{\partial\log p(y_t|\theta_t)}{\partial\theta_t}=0. \tag{12.27}$$

We want to find $\tilde{y}_t$, $\tilde{Q}_{t-1}$ and $\tilde{H}_t$ so that the solutions of (12.26) and (12.27) are the same; we do this by iteration based on linearisation. Suppose that $\tilde{\eta}_{t-1}$ is a trial value of η_{t-1}. Let

$$\tilde{\theta}_t=Z_t(T_{t-1}\alpha_{t-1}+R_{t-1}\tilde{\eta}_{t-1}),\qquad \dot{h}_t=-\left.\frac{\partial\log p(y_t|\theta_t)}{\partial\theta_t}\right|_{\theta_t=\tilde{\theta}_t},$$

$$\ddot{h}_t=-\left.\frac{\partial^2\log p(y_t|\theta_t)}{\partial\theta_t\partial\theta'_t}\right|_{\theta_t=\tilde{\theta}_t},$$

and take

$$\tilde{H}_t=\ddot{h}_t^{-1},\qquad \tilde{y}_t=\tilde{\theta}_t-\ddot{h}_t^{-1}\dot{h}_t. \tag{12.28}$$

The linearised form of

$$\frac{\partial\log p(y_t|\theta_t)}{\partial\theta_t},$$

at $\theta_t=\tilde{\theta}_t$ is $-\dot{h}_t-\ddot{h}_t(\theta_t-\tilde{\theta}_t)=\tilde{H}_t^{-1}(\tilde{y}_t-\theta_t)$. Substituting this in (12.27) and comparing the result with (12.26), we see that we have achieved the linearised form required for the observation term of (12.27).

To linearise the state term of (12.27) when $p(\eta_{t-1})$ is non-Gaussian, we proceed as in Section 11.6. We assume that $\eta_{t-1,i}$, the ith element of η_{t-1}, is independent of the other elements in η_{t-1} and that its density is a function of $\eta_{t-1,i}^2$, for $i=1,\ldots,r$. Although these assumptions appear restrictive, they enable us to deal with the important case in which the error densities have heavy tails. Let $r_{t-1,i}(\eta_{t-1,i}^2)=-2\log p(\eta_{t-1,i})$, and let

$$\dot{r}_{t-1,i}=\left.\frac{\partial r_{t-1,i}(\eta_{t-1,i}^2)}{\partial\eta_{t-1,i}^2}\right|_{\eta_{t-1}=\tilde{\eta}_{t-1}},$$

for a trial value $\tilde{\eta}_{t-1}$ of η_{t-1}. Then the linearised form of the state term (12.26) is

$$-\sum_{i=1}^{r} \dot{r}_{t-1,i}\eta_{t-1,i}. \tag{12.29}$$

Putting $\tilde{Q}_{t-1}^{-1} = \mathrm{diag}(\dot{r}_{t-1,1}, \ldots, \dot{r}_{t-1,r})$ we see that (12.29) has the same form as the state component of (12.26).

Starting with the trial value $\tilde{\eta}_{t-1}$ of η_{t-1} we use the solution

$$\eta_{t-1} = (\tilde{Q}_{t-1}^{-1} + R_{t-1}'Z_t'\tilde{H}_t^{-1}Z_tR_{t-1})^{-1}R_{t-1}'Z_t'\tilde{H}_t^{-1}(\tilde{y}_t - Z_tT_{t-1}\alpha_{t-1}),$$

of (12.26) to obtain a new value of η_{t-1} which we use as the next trial value. We repeat the process and continue until reasonable convergence is achieved.

12.7 Rao–Blackwellisation

12.7.1 Introduction

Suppose that we can partition the state vector into two components

$$\alpha_t = \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \end{pmatrix},$$

where α_{1t} and α_{2t} are such that the integration with respect to α_{1t} given α_{2t} can be performed analytically. For example, consider the univariate local level model (2.3) with a state equation variance η_t which varies over time. This model is given by

$$y_t = \mu_t + \varepsilon_t, \qquad \mu_{t+1} = \mu_t + \eta_t, \tag{12.30}$$

where $\varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2)$ and $\eta_t \sim \mathrm{N}(0, \exp h_t)$ with

$$h_{t+1} = (1-\phi)h^* + \phi h_t + \zeta_t, \qquad \zeta_t \sim \mathrm{N}(0, \sigma_\zeta^2), \tag{12.31}$$

and known fixed parameters σ_ε^2, h^*, ϕ and σ_ζ^2. The model is a variant of the local level model (2.3) in which η_t is modelled similarly to y_t of the stochastic volatility model (9.26) and (9.27) with $\mu = 0$ and $\sigma^2 \exp\theta_t = \exp h_t$. Since both μ_t and h_t evolve stochastically, we can place them both in the state vector α_t and take $\alpha_{1t} = \mu_t$ and $\alpha_{2t} = h_t$. It follows that for a given α_{2t}, this local level model reduces to a standard linear Gaussian model with known time-varying variances. The standard Kalman filter and smoothing methods of Chapter 4 can be employed for the estimation of α_{1t} given α_{2t}, or, in the wording of this section, for the analytical integration of α_{1t}.

We proceed as follows. Choose a sample $h_1^{(i)}, \ldots, h_n^{(i)}$, for $i = 1, \ldots, N$, using one of the techniques described in Section 12.4 applied to (12.31). Next, apply the Kalman filter and smoother to (12.30) with $\mathrm{Var}(\eta_t) = \exp h_t^{(i)}$ for $i = 1, \ldots, N$.

By using this approach, we only need to simulate h_t instead of having to simulate both μ_t and h_t as is required for the general analysis of nonlinear models described in Section 12.3.

This device for reducing the amount of simulation required by analytical integration of one component of a state vector for given values of another component is called *Rao-Blackwellisation*.

12.7.2 The Rao–Blackwellisation technique

The gain in efficiency from Rao–Blackwellisation arises from the following basic result in estimation theory. Suppose that a vector z is a function of the random vectors x and y with mean $\mathrm{E}(z) = \mu$. Let $z^* = \mathrm{E}(z|x)$; then $\mathrm{E}(z^*) = \mu$ and

$$\begin{aligned}\mathrm{Var}(z) &= \mathrm{E}\left[(z-\mu)(z-\mu)'\right] \\ &= \mathrm{E}\left\{\mathrm{E}\left[(z-z^*+z^*-\mu)(z-z^*+z^*-\mu)'\right]|x\right\} \\ &= E\left[\mathrm{Var}(z|x)\right] + \mathrm{Var}(z^*).\end{aligned}$$

Since $\mathrm{Var}(z|x)$ is non-negative definite, it follows that $\mathrm{Var}(z^*)$ is equal to or smaller in the matrix sense than $\mathrm{Var}(z)$. If z is regarded as an estimate of μ and $\mathrm{Var}(z^*)$ is strictly smaller than $\mathrm{Var}(z)$, then its conditional expectation z^* is an improved estimate.

Strictly speaking, the Rao–Blackwell theorem is concerned with the special case where x is a sufficient statistic for an unknown parameter; see, for example, Rao (1973, §5a.2(iii)) and Lehmann (1983, Theorem 6.4). Since in applications using the state space model, variable x is not a sufficient statistic, the use of the term 'Rao–Blackwellisation' is to some extent inappropriate; nevertheless since its use in the literature is well established we shall continue to use it here.

Denote the sets $\{\alpha_{1,1}, \ldots, \alpha_{1,t}\}$ and $\{\alpha_{2,1}, \ldots, \alpha_{2,t}\}$ by $\alpha_{1,1:t}$ and $\alpha_{2,1:t}$, respectively. As in Section 12.3, let x_t be an arbitrary function of α_t and suppose that we wish to estimate the conditional mean $\bar{x}_t$ given by

$$\begin{aligned}\bar{x}_t &= \mathrm{E}\left[x_t(\alpha_{1:t})|Y_t)\right] = \mathrm{E}\left[x_t(\alpha_{1,1:t}, \alpha_{2,1:t})|Y_t)\right] \\ &= \int x_t(\alpha_{1,1:t}, \alpha_{2,1:t}) p_t(\alpha_{1,1:t}, \alpha_{2,1:t}|Y_t)\, \mathrm{d}\alpha_{1,1:t}\, \mathrm{d}\alpha_{2,1:t}.\end{aligned}$$

Let

$$h_t(\alpha_{2,1:t}) = \int x_t(\alpha_{1,1:t}, \alpha_{2,1:t}) p_t(\alpha_{1,1:t}|\alpha_{2,1:t}, Y_t)\, \mathrm{d}\alpha_{1,1:t}. \tag{12.32}$$

We then have

$$\bar{x}_t = \int h_t(\alpha_{2,1:t}) p_t(\alpha_{2,1:t}|Y_t)\, \mathrm{d}\alpha_{2,1:t}. \tag{12.33}$$

We assume that $h_t(\alpha_{2,1:t})$ can be calculated analytically for given values of $\alpha_{2,1:t}$. Then (12.33) has the same form as (12.5) so similar methods can be used for evaluation of (12.33) as were used in Section 12.2.

Let $g(\alpha_{2,1:t}|Y_t)$ be an importance density chosen to be close to $p(\alpha_{2,1:t}|Y_t)$ and choose a random sample $\alpha_{2,1:t}^{(1)}, \ldots, \alpha_{2,1:t}^{(N)}$ from density $g(\alpha_{2,1:t}|Y_t)$. Then as in Section 12.2, $\bar{x}_t$ can be estimated by

$$\hat{x}_t = \sum_{i=1}^{N} h_t(\alpha_{2,1:t}^{(i)}) w_t^{(i)}, \tag{12.34}$$

with

$$\tilde{w}_t^{(i)} = \frac{p_t(\alpha_{2,1:t}^{(i)}, Y_t)}{q_t(\alpha_{2,1:t}^{(i)}|Y_t)}, \qquad \text{and} \qquad w_t^{(i)} = \tilde{w}_t^{(i)} \; / \; \sum_{j=1}^{N} \tilde{w}_t^{(j)}.$$

Here, appropriate particle filtering methods can be employed as in Section 12.2.

Since $h_t(\alpha_{2,1:t})$ is calculated analytically for each sample value $\alpha_{2,1:t} = \alpha_{2,1:t}^{(i)}$ for $i = 1, \ldots, N$, for example by means of the Kalman filter, computational gains are typically high. Although Rao–Blackwellisation can be used generally, it is specifically advantageous for high-dimensional models, where standard particle filtering can be computationally costly. This point is discussed by Doucet, De Freitas and Gordon (2001, pp. 499–515) who consider an example where the state vector has the huge dimension of 2^{100}. Further details are discussed in Doucet, Godsill and Andrieu (2000) and Maskell (2004).