10 Approximate filtering and smoothing

10.1 Introduction

In this chapter we consider approximate filtering and smoothing for data generated by a variety of non-Gaussian and nonlinear models such as those exemplified in Chapter 9. For the purpose of filtering, we assume that new observations y_t come in one at a time and that we wish to estimate functions of the state vector sequentially at each time point t taking account of all the observations up to and including time t. In the case of smoothing, we wish to estimate functions of the state vector for a given set of observations $y_1, \ldots, y_n$. Motivations and illustrations of filtering and smoothing for linear Gaussian models are discussed in Part I of this book. These motivations and illustrations are not intrinsically different in nonlinear non-Gaussian cases but the expressions for filtering and smoothing are not available in an analytical form. We rely on approximations or numerical solutions. Approximations to linear estimation and to Bayesian analysis can be obtained from Lemmas 2, 3 and 4 of Chapter 4.

We begin by considering two approximate filters, *the extended Kalman filter* in Section 10.2 and *the unscented Kalman filter* in Section 10.3. The ideas underlying the two approaches to nonlinear filtering are presented and their numerical performances are compared. Next, in Section 10.4 we consider nonlinear smoothing and show how approximate smoothing recursions can be derived for the two approximate filters. Section 10.5 argues that approximate solutions for filtering and smoothing can also be obtained when the data are transformed in an appropriate way. Two illustrations are presented as examples. In Sections 10.6 and 10.7 we discuss methods for computing the mode estimate of the state and signal vectors. The mode estimates are computed exactly. However, the analysis often focuses on mean, variance and possibly higher moments of the density and therefore provides an approximation in practice. Different treatments for models with heavy-tailed errors are collected and presented in Section 10.8.

10.2 The extended Kalman filter

The *extended Kalman filter* (EKF) is based on the idea of linearising the observation and state equations and then applying the Kalman filter straightforwardly to the resulting linearised model. We start with the following special case of the

Time Series Analysis by State Space Methods, Second Edition, J Durbin, S J Koopman © J Durbin
S J Koopman 2012. Published in 2012 by Oxford University Press

non-Gaussian nonlinear model (9.36) and (9.37) but where the disturbances are not necessarily normally distributed, that is

$$y_t = Z_t(\alpha_t) + \varepsilon_t, \qquad \alpha_{t+1} = T_t(\alpha_t) + R_t(\alpha_t)\eta_t, \tag{10.1}$$

for $t = 1, \ldots, n$, where $Z_t(\alpha_t)$, $T_t(\alpha_t)$ and $R_t(\alpha_t)$ are differentiable functions of α_t and where the random disturbances ε_t and η_t are serially and mutually uncorrelated with mean zero and variance matrices $H_t(\alpha_t)$ and $Q_t(\alpha_t)$, respectively. The initial state vector α_1 is random with mean a_1 and variance matrix P_1, and is uncorrelated with all disturbances.

We adopt the definitions of the predicted and filtered state vectors as used in Chapter 4, that is $a_t = \mathrm{E}(\alpha_t|Y_{t-1})$ and $a_{t|t} = \mathrm{E}(\alpha_t|Y_t)$, respectively. Define

$$\dot{Z}_t = \left.\frac{\partial Z_t(\alpha_t)}{\partial \alpha_t'}\right|_{\alpha_t = a_t}, \qquad \dot{T}_t = \left.\frac{\partial T_t(\alpha_t)}{\partial \alpha_t'}\right|_{\alpha_t = a_{t|t}}, \tag{10.2}$$

where we emphasise that $\dot{Z}_t$ is evaluated at time $t-1$ since a_t is a function of $y_1, \ldots, y_{t-1}$ and $\dot{T}_t$ is evaluated at time t since $a_{t|t}$ depends on $y_1, \ldots, y_t$. Expanding the matrix functions of (10.1) in Taylor series, based on the appropriate fixed values of a_t and $a_{t|t}$, gives

$$\begin{aligned}
Z_t(\alpha_t) &= Z_t(a_t) + \dot{Z}_t\,(\alpha_t - a_t) + \ldots,\\
T_t(\alpha_t) &= T_t(a_{t|t}) + \dot{T}_t\,(\alpha_t - a_{t|t}) + \ldots,\\
R_t(\alpha_t) &= R_t(a_{t|t}) + \ldots,\\
H_t(\alpha_t) &= H_t(a_t) + \ldots,\\
Q_t(\alpha_t) &= Q_t(a_{t|t}) + \ldots.
\end{aligned}$$

The matrix function $Z_t(\alpha_t)$ in the observation equation is expanded based on a_t while the matrix functions $T_t(\alpha_t)$ and $R_t(\alpha_t)$ are expanded based on $a_{t|t}$ since y_t is available for the state equation at time $t+1$. Substituting these expressions in (10.1), neglecting higher-order terms and assuming knowledge of a_t and $a_{t|t}$ gives

$$y_t = \dot{Z}_t\alpha_t + d_t + \varepsilon_t, \qquad \alpha_{t+1} = \dot{T}_t\alpha_t + c_t + R_t(a_{t|t})\eta_t, \tag{10.3}$$

where

$$d_t = Z_t(a_t) - \dot{Z}_t a_t, \qquad c_t = T_t(a_{t|t}) - \dot{T}_t a_{t|t},$$

and

$$\varepsilon_t \sim [0, H_t(a_t)], \qquad \eta_t \sim \big[0, Q_t(a_{t|t})\big].$$

Using the minimum variance matrix property of the Kalman filter discussed in Sections 4.2 and 4.3 as justification, we apply the Kalman filter with mean adjustments of Subsection 4.3.3 to the linearised model (10.1). We have

$$\begin{aligned} v_t &= y_t - \dot{Z} a_t - d_t \\ &= y_t - Z_t(a_t), \\ a_{t|t} &= a_t + P_t \dot{Z}_t' F_t^{-1} v_t, \\ a_{t+1} &= \dot{T}_t a_t + K_t v_t + c_t \\ &= \dot{T}_t a_t + K_t v_t + T_t(a_{t|t}) - \dot{T}_t [a_t + P_t \dot{Z}_t' F_t^{-1} v_t] \\ &= T_t(a_{t|t}), \end{aligned}$$

where $F_t = \dot{Z}_t P_t \dot{Z}_t' + H_t(a_t)$ and $K_t = \dot{T}_t P_t \dot{Z}_t' F_t^{-1}$. Putting these formulae together with the other equations from Subsection 4.3.3, we obtain the following recursion for calculating a_{t+1} and P_{t+1} given a_t and P_t,

$$\begin{aligned} v_t &= y_t - Z_t(a_t), & F_t &= \dot{Z}_t P_t \dot{Z}_t' + H_t(a_t), \\ a_{t|t} &= a_t + P_t \dot{Z}_t' F_t^{-1} v_t, & P_{t|t} &= P_t - P_t \dot{Z}_t' F_t^{-1} \dot{Z}_t P_t, \\ a_{t+1} &= T_t(a_{t|t}), & P_{t+1} &= \dot{T}_t P_{t|t} \dot{T}_t' + R_t(a_{t|t}) Q_t(a_{t|t}) R_t(a_{t|t})', \end{aligned} \tag{10.4}$$

for $t = 1, \ldots, n$. This recursion, together with the initial values a_1 and P_1, is called the *extended Kalman filter*. It is essentially given in this form by the equations (2.4) to (2.8) of Anderson and Moore (1979, §8.2). Earlier versions of the extended Kalman filter have appeared in Jazwinski (1970, §8.3). Comparisons of its performance with that of other filters will be presented in Subsection 10.3.4.

The extended Kalman filter is developed to accomodate nonlinear effects in the state space model. In case the densities of the disturbances ε_t and η_t in (10.1) are non-Gaussian, the extended Kalman filter does not change. We assume that the mean of the disturbances are zero and we set $H_t(\alpha_t)$ and $Q_t(\alpha_t)$ equal to the variance matrices of ε_t and η_t, respectively. In case elements of the variance matrices depend on the state vector, α_t is replaced by a_t or $a_{t|t}$.

10.2.1 A multiplicative trend-cycle decomposition

Consider the partly multiplicative model for the univariate observation y_t given by

$$y_t = \mu_t \times c_t + \varepsilon_t,$$

for $t = 1, \ldots, n$, where μ_t is the trend component as modelled by the random walk process $\mu_{t+1} = \mu_t + \xi_t$ and c_t is the unobserved cycle component (3.13) as discussed in Section 3.2. Similar multiplicative models are discussed in Section 9.7. The 3×1 state vector for this model is given by $\alpha_t = (\mu_t\,,\, c_t\,,\, c_t^*)'$ where c_t^* is implicitly defined in equation (3.13). The 3×1 disturbance vector is given by $\eta_t = (\xi_t\,,\, \tilde{\omega}_t\,,\, \tilde{\omega}_t^*)'$ where the cycle disturbances $\tilde{\omega}_t$ and $\tilde{\omega}_t^*$ are defined in equation (3.13). The three disturbances are independent normal variables. The state equation of (10.1) is linear for the multiplicative trend-cycle model. In particular, we have $T_t(\alpha_t) = T_t \times \alpha_t$, $R_t(\alpha_t) = R_t$ and $Q_t(\alpha_t) = Q_t$ with

$$T_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \rho\cos\lambda_c & \rho\sin\lambda_c \\ 0 & -\rho\sin\lambda_c & \rho\cos\lambda_c \end{bmatrix}, \qquad R_t = I_3, \qquad Q_t = \operatorname{diag}(\sigma_\xi^2\,,\,\sigma_\omega^2\,,\,\sigma_\omega^2),$$

where ρ is a discounting factor with $0 < \rho < 1$, λ_c is the frequency of the cycle c_t, σ_ξ^2 is the variance of disturbance ξ_t and σ_ω^2 is the variance of both disturbances $\tilde{\omega}_t$ and $\tilde{\omega}_t^*$. The multiplicative decomposition is represented by the nonlinear function $Z_t(\alpha_t)$ in the observation equation of (10.1) with the variance matrix $H_t(\alpha_t)$ and are given by

$$Z_t(\alpha_t) = \alpha_{1t}\alpha_{2t}, \qquad H_t(\alpha_t) = \sigma_\varepsilon^2,$$

where α_{jt} is the jth element of α_t and σ_ε^2 is the variance of the irregular or error term ε_t. We notice that $\alpha_{1t} = \mu_t$ and $\alpha_{2t} = c_t$ as required.

To apply the extended Kalman filter (10.4) we require in addition the variables $\dot{Z}_t$ and $\dot{T}_t$ and they are given by

$$\dot{Z}_t = (\alpha_{2t}\,,\,\alpha_{1t}\,,\,0)', \qquad \dot{T}_t = T_t.$$

Given these variables, we can carry out the computations for the extended Kalman filter (10.4) which needs to be initialised by

$$a_1 = 0, \qquad P_1 = \operatorname{diag}\left[\kappa\,,\,(1-\rho^2)^{-1}\sigma_\omega^2\,,\,(1-\rho^2)^{-1}\sigma_\omega^2\right],$$

where $\kappa \to \infty$. The initialisation of the extended Kalman filter needs to be modified since $\kappa \to \infty$ in P_1. The exact initialisation of the extended Kalman filter can be developed by following the treatment presented in Chapter 5 for the linear Gaussian case. We do not discuss these matters here further but refer to Koopman and Lee (2009) for a detailed derivation and discussion. An approximation to initialisation is to replace κ by a large value, say 10^7.

10.2.2 Power growth model

An alternative to the local linear trend model (3.2) discussed in Section 3.2 is the power growth model as given by

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim \mathrm{N}\big(0, \sigma_\varepsilon^2\big), \\ \mu_{t+1} &= \mu_t^{1+\nu_t} + \xi_t, & \xi_t &\sim \mathrm{N}\big(0, \sigma_\xi^2\big), \\ \nu_{t+1} &= \rho\nu_t + \zeta_t, & \zeta_t &\sim \mathrm{N}\big(0, \sigma_\zeta^2\big), \end{aligned} \tag{10.5}$$

where ρ is a discounting factor with $0 < \rho < 1$ for observation y_t and trend component μ_t with a stationary slope term ν_t. For this model we have a linear observation equation and a partly nonlinear state equation. The nonlinear state space model (10.1), with state vector $\alpha_t = (\mu_t\,,\,\nu_t)'$, has system matrices

$$Z_t(\alpha_t) = (1, 0), \qquad T_t(\alpha_t) = (\mu_t^{1+\nu_t}\,,\,\rho\nu_t)', \qquad Q_t(\alpha_t) = \operatorname{diag}(\sigma_\xi^2\,,\,\sigma_\zeta^2),$$

and with $H_t(\alpha_t) = \sigma_\varepsilon^2$ and $R_t(\alpha_t) = I_2$. The extended Kalman filter then relies only on $\dot{Z}_t$ and $\dot{T}_t$ which are given by

$$\dot{Z}_t = Z_t, \qquad \dot{T}_t = \begin{pmatrix} (\nu_t + 1)\mu_t^{\nu_t} & \mu_t^{1+\nu_t} \log \mu_t \\ 0 & \rho \end{pmatrix}.$$

The initial condition for the filter is given by

$$a_1 = 0, \qquad P_1 = \text{diag}\left[\kappa\,,\, (1-\rho^2)^{-1}\sigma_\nu^2\right],$$

where $\kappa \to \infty$.

10.3 The unscented Kalman filter

The second approximate filter we consider for non-Gaussian nonlinear models is the *unscented Kalman filter* (UKF). It is based on a radically different idea from the linearisation used for the EKF. The idea can be most easily understood by considering its application to a simpler problem than the state space filtering problem. We therefore consider first in Subsection 10.3.1 the *unscented transformation* for vector functions of random vectors. The derivation of the UKF is given in Subsection 10.3.2. Further improvements of the basic unscented transformation are discussed in Subsection 10.3.3. The accuracy of the improved UKF in comparison with the EKF and standard UKF is investigated in Subsection 10.3.4.

10.3.1 The unscented transformation

Assume we have a $p \times 1$ random vector y which is a known nonlinear function

$$y = f(x), \tag{10.6}$$

of an $m \times 1$ random vector x with density $x \sim \text{N}(\bar{x}, P_{xx})$ and that we wish to find an approximation to the density of y. Julier and Uhlmann (1997) suggest that we proceed as follows: choose a set of *sigma points* denoted by $x_0, x_1, \ldots, x_{2m+1}$ with associated *sigma weights* denoted by $w_0, w_1, \ldots, w_{2m+1}$ where each $w_i > 0$ such that

$$\sum_{i=0}^{2m} w_i = 1, \qquad \sum_{i=0}^{2m} w_i x_i = \bar{x}, \qquad \sum_{i=0}^{2m} w_i (x_i - \bar{x})(x_i - \bar{x})' = P_{xx}. \tag{10.7}$$

In effect, we are approximating the continuous density $f(x)$ by a discrete density at points $x_0, x_1, \ldots, x_{2m+1}$ whose mean vector and variance matrix are the same as those of density $f(x)$. We then define $y_i = f(x_i)$, for $i = 0, \ldots, 2m$ and take

$$\bar{y} = \sum_{i=0}^{2m} w_i y_i, \qquad P_{yy} = \sum_{i=0}^{2m} w_i (y_i - \bar{y})(y_i - \bar{y})', \tag{10.8}$$

as our estimates of $\mathrm{E}(y)$ and $\mathrm{Var}(y)$, respectively. It is shown in Julier and Uhlmann (1997) that the lower order terms of the Taylor expansions of (10.6) can be made equal to the corresponding terms of the Taylor expansions of the true moments $\mathrm{E}(y)$ and $\mathrm{Var}(y)$ for any smooth function $f(\cdot)$. If more moments of x are known, for instance by assuming normality, it is possible to approximate more moments of y and to approximate the lower moments with greater precision.

There are many ways in which the sigma points and weights could be chosen subject to the constraints (10.7); Julier and Uhlmann (1997) suggest the very simple form

$$x_0 = \bar{x}, \qquad x_i = \bar{x} + \lambda\sqrt{P_{xx,i}}\,, \qquad x_{i+m} = \bar{x} - \lambda\sqrt{P_{xx,i}}\,,$$

with weights w_0 and

$$w_i = w_{i+m} = \frac{1-w_0}{2m}, \qquad i = 1, \ldots, m,$$

where λ is a scalar and $\sqrt{P_{xx,i}}$ is the ith column of a matrix square root of P_{xx} that can be obtained by, for example, the Cholesky decomposition of a symmetric matrix. The constant λ is determined by constraints as follows. The constraints

$$\sum_{i=0}^{2m} w_i = 1, \qquad \sum_{i=0}^{2m} w_i x_i = \bar{x},$$

are obviously satisfied. Substituting the x_i's and w_i's in the third item of (10.7) gives

$$\frac{1-w_0}{2m}\sum_{i=1}^{2m} \lambda^2 \left(\sqrt{P_{xx,i}}\right)\left(\sqrt{P_{xx,i}}\right)' = P_{xx},$$

from which we deduce that

$$\lambda^2 = \frac{m}{1-w_0}.$$

By taking $w_0 = k \,/\, (m+k)$ for some value k, we obtain $\lambda^2 = m + k$ and

$$\begin{aligned} x_0 &= \bar{x}, & w_0 &= k \,/\, (m+k), \\ x_i &= \bar{x} + \sqrt{m+k}\sqrt{P_{xx,i}}, & w_i &= 1 \,/\, 2(m+k), \\ x_{i+m} &= \bar{x} - \sqrt{m+k}\sqrt{P_{xx,i}}, & w_{i+m} &= 1 \,/\, 2(m+k), \end{aligned} \tag{10.9}$$

for $i = 1, \ldots, m$. Julier and Uhlmann (1997) argue that 'k provides an extra degree of freedom to "fine tune" the higher order moments of the approximation, and can be used to reduce the overall prediction errors. When x is assumed Gaussian, a useful heuristic is to select $m + k = 3$.' While there are many possible alternatives we shall adopt this proposal for applications of the UKF to our state space model.

Julier and Uhlmann (1997) call this *the unscented transformation*. By examining Taylor expansions, they show that under appropriate conditions the unscented transformation estimates of the mean vector and the variance matrix of y are accurate to the second order of approximation.

10.3.2 Derivation of the unscented Kalman filter

We now construct a filter based on the unscented transformation. As for the EKF, we suppose that we have a series of observation vectors, $y_1, \ldots, y_n$, with corresponding state vectors, $\alpha_1, \ldots, \alpha_n$, which we assume are generated by the nonlinear non-Gaussian state space model (10.1). We carry out the construction in two stages, first the *updating stage* in which a new observation y_t has arrived and we wish to estimate $a_{t|t} = \mathrm{E}(\alpha_t|Y_t)$ and $P_{t|t} = \mathrm{Var}(\alpha_t|Y_t)$ given $a_t = \mathrm{E}(\alpha_t|Y_{t-1})$ and $P_t = \mathrm{Var}(\alpha_t|Y_{t-1})$, and secondly the *prediction stage* in which we wish to estimate a_{t+1} and P_{t+1} given $a_{t|t}$ and $P_{t|t}$, for $t = 1, \ldots, n$. In order to keep the notation simple we shall use the same symbols for estimates and the quantities being estimated. We shall apply the unscented transformation separately in the two stages.

For the updating stage, let $\bar{y}_t = \mathrm{E}(y_t|Y_{t-1})$ and $v_t = y_t - \bar{y}_t$. The application of Lemma 1 in Section 4.2 is exact for the linear Gaussian state space model of Part I. In the case of non-Gaussian and nonlinear models Lemma 1 can still be applied but it provides only approximate relations. We use equation (4.2) of Lemma 1 in Section 4.2 to provide the approximate relation

$$\mathrm{E}(\alpha_t|Y_t) = \mathrm{E}(\alpha_t|Y_{t-1}) + P_{\alpha v,t} P_{vv,t}^{-1} v_t,$$

that is,

$$a_{t|t} = a_t + P_{\alpha v,t} P_{vv,t}^{-1} v_t, \tag{10.10}$$

where $P_{\alpha v,t} = \mathrm{Cov}(\alpha_t, v_t)$ and $P_{vv,t} = \mathrm{Var}(v_t)$ in the conditional joint distribution of α_t and v_t given Y_{t-1}, for $t = 1, \ldots, n$. Similarly, we use equation (4.3) of Lemma 1 to provide the approximate relation

$$\mathrm{Var}(\alpha_t|Y_t) = \mathrm{Var}(\alpha_t|Y_{t-1}) - P_{\alpha v,t} P_{vv,t}^{-1} P_{\alpha v,t}',$$

that is,

$$P_{t|t} = P_t - P_{\alpha v,t} P_{vv,t}^{-1} P_{\alpha v,t}'. \tag{10.11}$$

Moreover, since the observation equation of model (10.1) is $y_t = Z(\alpha_t) + \varepsilon_t$, we have

$$\bar{y}_t = E\left[Z_t(\alpha_t)|Y_{t-1}\right]. \tag{10.12}$$

We proceed to estimate $\bar{y}_t$, $P_{\alpha v,t}$ and $P_{vv,t}$ by the unscented transformation. Define the sigma points and weights as follows,

$$\begin{aligned} x_{t,0} &= a_t, & w_0 &= k \;/\; (m+k), \\ x_{t,i} &= a_t + \sqrt{m+k} P_{t,i}^*, & w_i &= 1 \;/\; 2(m+k), \\ x_{t,i+m} &= a_t - \sqrt{m+k} P_{t,i}^*, & w_{i+m} &= 1 \;/\; 2(m+k), \end{aligned} \tag{10.13}$$

where m is the dimensionality of the state vector α_t and $P^*_{t,i}$ is the ith column of the square root matrix P^*_t obtained from a Cholesky decomposition $P_t = P^*_t P^{*\prime}_t$ for $i = 1, \ldots, m$. We then take

$$\begin{aligned} \bar{y}_t &= \textstyle\sum_{i=0}^{2m} w_i Z_t(x_{t,i}), \\ P_{\alpha v,t} &= \textstyle\sum_{i=0}^{2m} w_i (x_{t,i} - a_t) \left[Z_t(x_{t,i}) - \bar{y}_t \right], \\ P_{vv,t} &= \textstyle\sum_{i=0}^{2m} w_i \left[Z_t(x_{t,i}) - \bar{y}_t \right] \left[Z_t(x_{t,i}) - \bar{y}_t \right]' + H_t(x_{t,i}), \end{aligned} \tag{10.14}$$

for $t = 1, \ldots, n$. Taking $v_t = y_t - \bar{y}_t$, these are substituted in (10.10) and (10.11) to give $a_{t|t}$ and $P_{t|t}$, respectively.

To implement the prediction stage of the filter we first notice that the state equation of model (10.1) is $\alpha_{t+1} = T_t(\alpha_t) + R_t(\alpha_t)\eta_t$ where η_t is independent of α_t with vector mean zero and variance matrix $Q_t(\alpha_t)$. We then have

$$a_{t+1} = \mathrm{E}(\alpha_{t+1}|Y_t) = \mathrm{E}\left[T_t(\alpha_t)|Y_t) \right], \tag{10.15}$$

and

$$P_{t+1} = \mathrm{Var}(\alpha_{t+1}|Y_t) = \mathrm{Var}\left[T_t(\alpha_t)|Y_t) \right] + \mathrm{E}\left[R_t(\alpha_t) Q_t(\alpha_t) R_t(\alpha_t)'|Y_t) \right], \tag{10.16}$$

for $t = 1, \ldots, n$. Define new $x_{t,0}, \ldots, x_{t,2m}$ by relations (10.13) with a_t replaced by $a_{t|t}$ and P_t replaced by $P_{t|t}$. From these x_i's and the values for $w_0, \ldots, w_{2m}$ from (10.13), we take

$$\begin{aligned} a_{t+1} &= \textstyle\sum_{i=0}^{2m} w_i T_t(x_{t,i}), \\ P_{t+1} &= \textstyle\sum_{i=0}^{2m} w_i \left[T_t(x_{t,i}) - a_{t+1} \right] \left[T_t(x_{t,i}) - a_{t+1} \right]' \\ &\quad + \textstyle\sum_{i=0}^{2m} w_i R_t(x_{t,i}) Q_t(x_{t,i}) R_t(x_{t,i})'. \end{aligned} \tag{10.17}$$

The filter defined by the relations (10.10), (10.11), (10.14) and (10.17) is called the *unscented Kalman filter* (UKF).

In terms of Taylor expansions, the unscented transform is accurate to the second order for the mean estimate. This result applies therefore to the estimate of the mean of the state vector by the UKF. The approximation of the mean by the EKF estimate is only accurate to the first order. However, the variance matrix is estimated to the second order of approximation by both the UKF and the EKF.

10.3.3 Further developments of the unscented transform

The original transformation of Julier and Uhlmann (1997) based on (10.9) is simple in the sense that we only take the two sigma points

$$\bar{x} \pm \sqrt{m + k} \sqrt{P_{xx,i}},$$

for each ith element of the vector x. We now consider whether we can improve efficiency by increasing the number of sigma points for each i and also by allocating relatively higher weight to sigma points that correspond to higher densities. For this purpose we suggest taking $2\,q$ sigma points for each i by considering

$$\begin{aligned} x_{ij} &= \bar{x} + \lambda\xi_j\sqrt{P_{xx,i}} \qquad i = 1,\ldots,m, \\ x_{ij} &= \bar{x} - \lambda\xi_j\sqrt{P_{xx,i}} \qquad i = m+1,\ldots,2m, \end{aligned} \qquad j = 1,\ldots,q, \tag{10.18}$$

together with $x_0 = \bar{x}$. For example, for $q = 2$, we could take $\xi_1 = 1$ and $\xi_2 = 2$ or $\xi_1 = \frac{1}{2}$ and $\xi_2 = \frac{3}{2}$, and for $q = 4$, we could take $\xi_1 = \frac{1}{2}$, $\xi_2 = 2$, $\xi_3 = \frac{3}{2}$ and $\xi_4 = 2$. Associated with these sigma points, we take the weights

$$w_{ij} = w_{m+i,j} = w\varphi(\xi_j), \qquad i = 1,\ldots,m, \qquad j = 1,\ldots,q,$$

together with w_0, where w and w_0 are to be determined and where $\varphi(\cdot)$ is the standard normal density. The constraint that the weights must sum to one,

$$w_0 + 2\sum_{i=1}^{m}\sum_{j=1}^{q} w_{ij} = 1,$$

leads to

$$w_0 + 2m\,w\sum_{j=1}^{q}\varphi(\xi_j) = 1,$$

which gives

$$w = \frac{1 - w_0}{2m\sum_{j=1}^{q}\varphi(\xi_j)}. \tag{10.19}$$

It also follows that

$$\sum_{i=0}^{2m}\sum_{j=1}^{q} w_{ij}x_{ij} = \left[w_0 + 2m\,w\sum_{j=1}^{q}\varphi(\xi_j)\right]\bar{x} = \bar{x}.$$

The constraint

$$2\sum_{i=1}^{m}\sum_{j=1}^{q} w_{ij}(x_{ij} - \bar{x})(x_{ij} - \bar{x})' = 2\sum_{i=1}^{m}\sum_{j=1}^{q} w_{ij}\lambda^2\xi_j^2 P^*_{xx,i}P^{*\prime}_{xx,i} = P_{xx},$$

where

$$w_{ij} = \frac{(1 - w_0)\varphi(\xi_j)}{2m\sum_{l=1}^{q}\varphi(\xi_l)}$$

gives

$$\frac{2(1 - w_0)\lambda^2}{2m\sum_{l=1}^{q}\varphi(\xi_l)}m\sum_{j=1}^{q}\varphi(\xi_j)\xi_j^2 = 1, \tag{10.20}$$

and hence

$$\lambda^2 = \frac{\sum_{j=1}^{q} \varphi(\xi_j)}{(1-w_0)\sum_{j=1}^{q} \varphi(\xi_j)\xi_j^2}. \tag{10.21}$$

We now seek a further constraint by equalising fourth moments in order to provide a value for w_0. Denote the scalar elements of the vector x by x_ℓ^* for $\ell = 1, \ldots, m$ so that $x = (x_1^*, \ldots, x_m^*)'$. Let $\bar{x}_\ell^* = \mathrm{E}(x_\ell^*)$ for $\ell = 1, \ldots, m$. For simplicity, take the case

$$P_{xx} = \mathrm{Var}(x) = \mathrm{diag}\left(\sigma_1^2, \ldots, \sigma_m^2\right),$$

and assume that x is normally distributed. Thus $x_\ell^* \sim \mathrm{N}(\bar{x}_\ell^*, \sigma_\ell^2)$ for $\ell = 1, \ldots, m$. Denote the ℓth element of x_{ij} by $x_{ij\ell}$ and assign to it the weight w_{ij}. From (10.18) we have

$$\begin{aligned} x_{ij\ell} &= \bar{x}_\ell^* + \lambda\xi_j\sigma_\ell, \qquad i = 1, \ldots, m, \\ &= \bar{x}_\ell^* - \lambda\xi_j\sigma_\ell, \qquad i = m+1, \ldots, 2m, \end{aligned} \tag{10.22}$$

for $j = 1, \ldots, q$ and $\ell = 1, \ldots, m$. Let us impose the fourth moment constraint

$$\begin{aligned} \mathrm{E}\left[\sum_{\ell=1}^{m} (x_\ell^* - \bar{x}_\ell^*)^4\right] &= 2\sum_{\ell=1}^{m}\sum_{i=1}^{m}\sum_{j=1}^{q} w_{ij}\mathrm{E}\left[(x_\ell^* - \bar{x}_\ell^*)^4\right] \\ &= 2\sum_{\ell=1}^{m}\sum_{i=1}^{m}\sum_{j=1}^{q} w\varphi(\xi_j)\left(\lambda\xi_j\sigma_\ell\right)^4 \\ &= 2m\, w\lambda^4 \sum_{\ell=1}^{m}\sum_{j=1}^{q} \varphi(\xi_j)\xi_j^4\sigma_\ell^4. \end{aligned}$$

Since the fourth moment of $\mathrm{N}(0, \sigma_\ell^2)$ is $3\sigma_\ell^4$, and using (10.19) and (10.21), we have

$$3\sum_{\ell=1}^{m}\sigma_\ell^4 = \frac{1-w_0}{\sum_{j=1}^{q}\varphi(\xi_j)} \frac{\left[\sum_{j=1}^{q}\varphi(\xi_j)\right]^2 \sum_{j=1}^{q}\varphi(\xi_j)\xi_j^4 \sum_{\ell=1}^{m}\sigma_\ell^4}{(1-w_0)^2[\sum_{j=1}^{q}\varphi(\xi_j)\xi_j^2]^2}, \tag{10.23}$$

giving

$$3 = \frac{\sum_{j=1}^{q}\varphi(\xi_j)\sum_{j=1}^{q}\varphi(\xi_j)\xi_j^4}{(1-w_0)\left[\sum_{j=1}^{q}\varphi(\xi_j)\xi_j^2\right]^2}, \tag{10.24}$$

and hence

$$w_0 = 1 - \frac{\sum_{j=1}^{q}\varphi(\xi_j)\sum_{j=1}^{q}\varphi(\xi_j)\xi_j^4}{3\left[\sum_{j=1}^{q}\varphi(\xi_j)\xi_j^2\right]^2}. \tag{10.25}$$

We use this value as an approximation when P_{xx} is not diagonal.

We have thus developed an expanded set of sigma points spread over a wider area than (10.9) over the support of the distribution, while maintaining the correspondence between moments of the distribution over the sigma points and the population moments.

10.3.4 Comparisons between EKF and UKF

In this section we provide a comparison of the estimation performance of three nonlinear filtering approaches. We generate 10,000 replications from the (partly) nonlinear state space model (10.1) with a stationary autoregressive process for the univariate state α_t, that is

$$y_t = Z_t(\alpha_t) + \varepsilon_t, \qquad \alpha_{t+1} = 0.95\alpha_t + \eta_t,$$

with error terms $\varepsilon_t \sim \mathrm{N}(0, 0.01)$ and $\eta_t \sim \mathrm{N}(0, 0.01)$ for $t = 1, \ldots, 100$. We consider different nonlinear transformations for $Z_t(\alpha_t)$. Table 10.1 shows the mean of the sum of squared prediction errors for different choices of the $Z(\alpha_t)$ function and for the filters EKF, UKF and the modified UKF. The results show that the unscented filters generally provide more accurate state predictions than the extended filter. When we increase the order of the polynomial transformation, the accuracies diverge: at lower orders the two unscented filters have a similar performance, while at high orders the modified UKF performs significantly better. We achieve similar improvements with the $\exp(\alpha_t)$ and $\sin(\alpha_t) + 1.1$ transformations although they are small and comparable with transformations such as α_t^2 and α_t^3. The $\log(\alpha_t + 6)$ transformation can be appropriately handled by the EKF since the UKF does not provide much improvement.

Table 10.1 Mean squared prediction errors for three nonlinear filtering methods.

$Z(\alpha_t)$	EKF	UKF	MUKF
α_t^2	55.584	30.440	30.440
α_t^3	16.383	10.691	10.496
α_t^4	46.048	43.484	28.179
α_t^5	107.23	17.019	13.457
α_t^6	147.62	92.216	30.651
α_t^7	1468.1	85.311	24.175
α_t^8	1641.1	347.84	39.996
$\exp(\alpha_t)$	13.177	11.917	11.866
$\sin(\alpha_t) + 1.1$	32.530	24.351	23.514
$\log(\alpha_t + 6)$	38.743	38.695	38.695

The EKF refers to the extended Kalman filter of Section 10.2, UKF refers to the unscented Kalman filter of Subsection 10.3.1 and MUKF refers to the modified UKF method of Subsection 10.3.3 with $q = 4$.

10.4 Nonlinear smoothing

Smoothing is extensively discussed in Chapter 4 for the linear Gaussian model. Approximate filtering methods for nonlinear models can be based on the extended Kalman filter and on the unscented Kalman filter. Since both approximations are associated with the Kalman filter recursions, it can be anticipated that methods of smoothing for the linear Gaussian model can be adapted similarly. We discuss some details for both approaches.

10.4.1 Extended smoothing

The linearisation technique that we employed in Section 10.2 to provide the extended Kalman filter can be used to obtain an approximate state smoother. We simply take the linearised form (10.3) of the nonlinear model (10.1) and apply the EKF (10.4). The system matrices of the linearised state space form vary over time and depend on predicted and filtered estimates of the state vector at time t. The backward smoothing methods as described in Section 4.4 can be used as approximations to smoothing. More specifically, the recursions for the extended state smoother are given by

$$r_{t-1} = \dot{Z}_t' F_t^{-1} v_t + L_t' r_t, \qquad \hat{\alpha}_t = a_t + P_t r_{t-1}, \qquad t = n, \ldots, 1, \quad (10.26)$$

where F_t, v_t,, L_t and P_t are obtained from the extended Kalman filter (10.4) while the recursion is initiated with $r_n = 0$.

For the sole purpose of smoothing, the linearised form (10.3) can be considered once more, in a second round of Kalman filtering and smoothing. We then evaluate $\dot{T}_t$ and $\dot{Z}_t$ in (10.2) once more but now at $\alpha_t = \hat{\alpha}_t$, for $t = 1, \ldots, n$. The Kalman filter and smoother equations are used to obtain the new estimates $\hat{\alpha}_1, \ldots, \hat{\alpha}_n$. We expect that the linear approximation based on $\hat{\alpha}_t$ is more accurate than the approximation based on a_t and $a_{t|t}$.

10.4.2 Unscented smoothing

The unscented filtering method can be modified in different ways to obtain an unscented smoothing method. For example, the two filter formula method of Subsection 4.6.4 can be advocated. The UKF is applied to the time series once and the UKF is also applied to the same time series in reverse order. The UKF estimates are then combined as in (4.74) to obtain the smoothed estimates. This approach only leads to an approximation of $\mathrm{E}(\alpha_t|Y_n)$ for a nonlinear state space model.

An alternative and our preferred approximation to a smoothing algorithm based on the unscented transformation can be derived from Lemma 1 of Chapter 4. Its development is similar to the classical fixed interval smoother of Section 4.4. Assume that the α_t and α_{t+1} conditional on Y_t are jointly normally distributed with

$$\mathrm{E}(\alpha_t|Y_t) = a_{t|t}, \qquad \mathrm{E}(\alpha_{t+1}|Y_t) = a_{t+1}, \quad (10.27)$$

$$\mathrm{Var}(\alpha_t|Y_t) = P_{t|t}, \qquad \mathrm{Var}(\alpha_{t+1}|Y_t) = P_{t+1}, \tag{10.28}$$

and

$$\mathrm{Cov}(\alpha_t, \alpha_{t+1}|Y_t) = C_{t+1}. \tag{10.29}$$

From Lemma 1, we obtain the approximation

$$\begin{aligned} \hat{\alpha_t} &= \mathrm{E}(\alpha_t|Y_n) \\ &= \mathrm{E}(\alpha_t|\alpha_{t+1}, Y_t) \\ &= a_{t|t} + C_{t+1}P_{t+1}^{-1}\big(\hat{\alpha}_{t+1} - a_{t+1}\big), \end{aligned} \tag{10.30}$$

for $t = 1, \ldots, n$, where a_{t+1}, $a_{t|t}$ and P_{t+1} have been computed by the UKF, and

$$C_{t+1} = \sum_{i=0}^{2m} w_i\big(x_{t,i} - a_t\big)\big[T_t(x_{t,i}) - a_{t+1}\big]'. \tag{10.31}$$

These results imply a backward recursive algorithm for smoothing. Using the same arguments, we can compute the approximated smoothed state variance. A recent discussion of unscented smoothing is provided by Särkkä and Hartikainen (2010).

10.5 Approximation via data transformation

Nonlinear and non-Gaussian models can sometimes be successfully approximated by transforming observations into forms for which a linear Gaussian model can be employed as an approximation. This approach of approximation via data transformation is usually part of an *ad hoc* solution. In case of the nonlinear model (9.36), that is

$$y_t = Z_t(\alpha_t) + \varepsilon_t, \qquad \alpha_{t+1} = T_t\alpha_t + R_t\eta_t,$$

for $t = 1, \ldots, n$, we can consider a function $Z_t^-(\alpha_t)$ such that

$$Z_t^-\left[Z_t(\alpha_t)\right] = c_t^a + Z_t^a\alpha_t,$$

where vectors c_t^a and matrix Z_t^a have appropriate dimensions. The approximation is then achieved by applying the transformation $Z_t^-(\alpha_t)$ since

$$Z_t^-\left[Z_t(\alpha_t) + \varepsilon_t\right] \approx c_t^a + Z_t^a\alpha_t + u_t,$$

where u_t is an error term. A typical univariate example is to take $Z_t(\alpha_t) = \exp(\alpha_t)$ such that $Z_t^-(\alpha_t) = \log(\alpha_t)$.

10.5.1 Partly multiplicative decompositions

Consider the multiplicative model for the univariate observation y_t given by

$$y_t = \mu_t \times \gamma_t + \varepsilon_t,$$

for $t = 1, \ldots, n$, where μ_t is an unobserved trend component and γ_t is an unobserved seasonal or cycle component; see Section 3.2. By taking logs of y_t we obtain approximately

$$\log(y_t) \approx \log(\mu_t) + \log(\gamma_t) + u_t,$$

where appropriate dynamic properties are given to the components $\log(\mu_t)$ and $\log(\gamma_t)$.

10.5.2 Stochastic volatility model

Consider the basic SV model (9.26) of Section 9.5 as given by

$$y_t = \mu + \sigma \exp\left(\frac{1}{2}\theta_t\right)\varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0,1),$$

for $t = 1, \ldots, n$, where the log-volatility is modelled as an autoregressive process (9.27), that is

$$\theta_{t+1} = \phi\theta_t + \eta_t, \qquad \eta_t \sim \mathrm{N}\big(0, \sigma_\eta^2\big)$$

and where the disturbances ε_t and η_t are mutually and serially uncorrelated. To obtain an approximate solution based on a linear model, we can transform the observations y_t as follows

$$\log y_t^2 = \kappa + \theta_t + \xi_t, \qquad t = 1, \ldots, n, \tag{10.32}$$

where

$$\kappa = \log\sigma^2 + \mathrm{E}\big(\log\varepsilon_t^2\big), \qquad \xi_t = \log\varepsilon_t^2 - \mathrm{E}\big(\log\varepsilon_t^2\big). \tag{10.33}$$

The noise term ξ_t is not normally distributed but the model for $\log y_t^2$ is linear and therefore we can proceed approximately with the linear techniques of Part I. This approach is taken by Harvey, Ruiz and Shephard (1994) who call the procedure for parameter estimation based on it *quasi-maximum likelihood* (QML). Parameter estimation is done via the Kalman filter; smoothed estimates of the volatility component, θ_t, can be constructed and forecasts of volatility can be generated. One of the attractions of the QML approach is that it can be carried out straightforwardly using the readily available software program *STAMP*; see Koopman, Harvey, Doornik and Shephard (2010).

10.6 Approximation via mode estimation

The treatments for the extended and unscented filters are principally motivated by nonlinear models of the form (10.1) where the state vector is the argument of the nonlinear functions in the model. A general class of models can be formulated by

$$y_t \sim p(y_t|\alpha_t), \qquad \alpha_{t+1} = T_t(\alpha_t) + R_t(\alpha_t)\eta_t, \tag{10.34}$$

where $p(y_t|\alpha_t)$ is the observation density conditional on the state vector α_t which can evolve over time depending on nonlinear functions of the state vector but which in practice is often linear and Gaussian. In case the observation density has a possibly nonlinear mean function $Z_t(\alpha_t)$ and a possibly nonlinear variance function $H_t(\alpha_t)$, the models (10.1) and (10.34) are equivalent. The focus of model (10.34) is, however, on the non-Gaussian feature of $p(y_t|\alpha_t)$ rather than on the nonlinear functions of the model. The extended and unscented filters do not account for the non-Gaussian properties of the model other than their first two moments: the mean and variance functions. We will show that computing the mode of the state vector conditional on all observations will lead to an approximating linear Gaussian state space model.

We first discuss the case where $p(y_t|\alpha_t)$ can be any density but where α_t evolves linearly over time with a Gaussian error vector η_t. We further assume that the signal vector θ_t is a linear function of the state vector and that it is sufficient for the observation density; see Section 9.2 for a discussion on this class of models. In other words, density $p(y_t|\alpha_t)$ is equivalent to $p(y_t|\theta_t)$. We obtain the non-Gaussian model specification

$$y_t \sim p(y_t|\theta_t), \quad \theta_t = Z_t\alpha_t, \quad \alpha_{t+1} = T_t\alpha_t + R_t\eta_t, \quad \eta_t \sim \mathrm{N}(0, Q_t), \tag{10.35}$$

for $t = 1, \ldots, n$. The computation of the mode of θ_t conditional on all observations will lead to an approximating linear Gaussian state space model. The standard Kalman filtering and smoothing methods can be used for analysis.

10.6.1 Mode estimation for the linear Gaussian model

In Section 4.13 we have shown that the linear Gaussian state space model can be expressed in matrix form. Since we wish to focus on the signal vector θ, where $\theta = (\theta_1', \ldots, \theta_n')'$ with signal $\theta_t = Z_t\alpha_t$ as defined in (9.5) for $t = 1, \ldots, n$, we write the observation equation in matrix form as

$$Y_n = \theta + \varepsilon, \qquad \theta = Z\alpha, \qquad \varepsilon \sim \mathrm{N}(0, H), \tag{10.36}$$

where $\alpha = T(\alpha_1^* + R\eta)$ and matrix H is block-diagonal. All vector and matrix definitions are given in Subsection 4.13.1, including definitions for matrices Z, H, T, R and Q in the equations (4.94)–(4.99) and including equations for the mean vector and variance matrix of θ, that is

$$\mathrm{E}(\theta) = \mu = ZTa_1^*, \qquad \mathrm{Var}(\theta) = \Psi = ZT(P_1^* + RQR')T'Z',$$

where a_1^* and P_1^* are defined below (4.100). It follows that the observation equation (10.36) can also be expressed as

$$Y_n = \mu + u, \qquad u \sim \mathrm{N}(0, \Sigma), \qquad \Sigma = \Psi + H.$$

Since the mode of Gaussian densities is equal to the mean it follows from Subsection 4.13.5 that the mode of the signal can be expressed by

$$\hat{\theta} = (\Psi^{-1} + H^{-1})^{-1}(\Psi^{-1}\mu + H^{-1}Y_n), \tag{10.37}$$

where the mode $\hat{\theta}$ is the value of θ that maximises the smoothed density $p(\theta|Y_n)$ for the linear Gaussian state space model.

10.6.2 Mode estimation for model with linear Gaussian signal

Here we aim to estimate the mode for the smoothed density of the class of models represented by (10.34) where the signal is linear and Gaussian; see Section 9.2. This class of models in matrix form is

$$Y_n \sim p(Y_n|\theta), \qquad \theta \sim \mathrm{N}(\mu, \Psi),$$

where

$$p(Y_n|\theta) = \prod_{t=1}^{n} p(y_t|\theta_t) = \prod_{t=1}^{n} p(\varepsilon_t).$$

The smoothed density $p(\theta|Y_n)$ does not have an explicit expression from which we can obtain the mode analytically. Therefore, we express the smoothed logdensity by

$$\log p(\theta|Y_n) = \log p(Y_n|\theta) + \log p(\theta) - \log p(Y_n), \tag{10.38}$$

and maximise this expression with respect to θ numerically using the Newton–Raphson method; see Nocedal and Wright (1999) for a general discussion of the Newton–Raphson method. The components of the smoothed density in (10.38) dependent on θ are the observation density $p(Y_n|\theta)$ and the signal density $p(\theta)$ as given by (4.111), that is

$$\log p(\theta) = \mathrm{N}(\mu, \Psi) = \text{constant} - \frac{1}{2}\log|\Psi| - \frac{1}{2}(\theta - \mu)'\Psi^{-1}(\theta - \mu), \tag{10.39}$$

since θ is linear and Gaussian. The density $p(Y_n)$ does not depend on θ.

For a given guess of the mode, say $\tilde{\theta}$, a new guess of the mode, say $\tilde{\theta}^+$, is obtained by solving a second-order Taylor expansion of $\log p(\theta|Y_n)$ around $\theta = \tilde{\theta}$. We have

$$\tilde{\theta}^+ = \tilde{\theta} - \left[\ddot{p}(\theta|Y_n)|_{\theta=\tilde{\theta}}\right]^{-1} \dot{p}(\theta|Y_n)|_{\theta=\tilde{\theta}}\,, \tag{10.40}$$

where

$$\dot{p}(\cdot|\cdot) = \frac{\partial \log p(\cdot|\cdot)}{\partial\theta}, \qquad \ddot{p}(\cdot|\cdot) = \frac{\partial^2 \log p(\cdot|\cdot)}{\partial\theta\partial\theta'}. \tag{10.41}$$

Given these definitions and the expressions (10.38) and (10.39) for $\log p(\theta|Y_n)$ and $\log p(\theta)$, respectively, we obtain

$$\dot{p}(\theta|Y_n) = \dot{p}(Y_n|\theta) - \Psi^{-1}(\theta - \mu), \qquad \ddot{p}(\theta|Y_n) = \ddot{p}(Y_n|\theta) - \Psi^{-1}. \tag{10.42}$$

The independence assumption (9.2) implies that

$$\log p(Y_n|\theta) = \sum_{t=1}^{n} \log p(y_t|\theta_t),$$

so that matrix $\ddot{p}(Y_n|\theta)$ is block-diagonal. More specifically, we have

$$\begin{aligned} \dot{p}(Y_n|\theta) &= [\dot{p}_1(y_1|\theta_1), \ldots, \dot{p}_n(y_n|\theta_n)]', \\ \ddot{p}(Y_n|\theta) &= \operatorname{diag}[\ddot{p}_1(y_1|\theta_1), \ldots, \ddot{p}_n(y_n|\theta_n)], \end{aligned} \tag{10.43}$$

where

$$\dot{p}_t(\cdot|\cdot) = \frac{\partial \log p(\cdot|\cdot)}{\partial \theta_t}, \qquad \ddot{p}_t(\cdot|\cdot) = \frac{\partial^2 \log p(\cdot|\cdot)}{\partial \theta_t \partial \theta_t'},$$

for $t = 1, \ldots, n$.

By substitution of (10.42) into (10.40), the Newton–Raphson updating step (10.40) becomes

$$\begin{aligned} \tilde{\theta}^+ &= \tilde{\theta} - \left\{ \ddot{p}(Y_n|\theta)|_{\theta=\tilde{\theta}} - \Psi^{-1} \right\}^{-1} \left\{ \dot{p}(Y_n|\theta)|_{\theta=\tilde{\theta}} - \Psi^{-1}(\tilde{\theta} - \mu) \right\} \\ &= \left(\Psi^{-1} + A^{-1} \right)^{-1} \left(A^{-1} x + \Psi^{-1} \mu \right), \end{aligned} \tag{10.44}$$

where

$$A = -\left\{ \ddot{p}(Y_n|\theta)|_{\theta=\tilde{\theta}} \right\}^{-1}, \qquad x = \tilde{\theta} + A\, \dot{p}(Y_n|\theta)|_{\theta=\tilde{\theta}}. \tag{10.45}$$

We note the similarity between (10.44) and (10.37). In case $\ddot{p}(Y_n|\theta)$ is negative definite for all θ, it is due the block-diagonality of matrix A as implied by (10.43) that the Kalman filter and smoother of Chapter 4 can be used to compute the next guess of the mode $\tilde{\theta}^+$, for a given current guess of $\tilde{\theta}$. This Kalman filter and smoother is based on the Gaussian state space model (10.36) with $Y_n = x$ and $H = A$. When the new estimate $\tilde{\theta}^+$ is computed, we can treat it as a new guess $\tilde{\theta} = \tilde{\theta}^+$ for which yet another new guess can be computed. This process can be repeated and constitutes the Newton–Raphson method for this application. Convergence is usually fast and only around ten iterations or less are needed in many cases of interest. After convergence, we have obtained the mode $\hat{\theta}$ with Hessian matrix $G = \ddot{p}(\theta|y)|_{\theta=\hat{\theta}} = -\Psi^{-1} - A^{-1}$, where A is evaluated at $\theta = \hat{\theta}$. It is shown that for an appropriately designed linear Gaussian model, the Kalman filter and smoother is able to compute the mode for the nonlinear non-Gaussian model (10.34).

The iterative approach for computing the mode was proposed by Shephard and Pitt (1997), Durbin and Koopman (1997) and So (2003, §2). The method is clearly not valid when $\ddot{p}(y|\theta)$ is not negative definite since this will imply that the variance matrix H of the linear Gaussian model (10.36) is non-negative definite. In other words, density $p(y|\theta)$ must be logconcave in θ. In the case $p(y|\theta)$ is not logconcave, the method of this section can still be adopted but the derivation must be based on other arguments; see Jungbacker and Koopman (2007).

When the mode of the state vector α_t is required for (10.34), we can repeat the above derivation with $p(\theta_t|Y_n)$ replaced by $p(\alpha_t|Y_n)$ and θ_t replaced by α_t. However, it is argued in Subsection 4.5.3 that state smoothing is computationally more involved than signal smoothing. Since we need to apply the Kalman filter and smoothing algorithms repeatedly as part of the Newton–Raphson method for computing the mode, the computation time is higher for obtaining the mode of the state compared to obtaining the mode of the signal. In this respect the result of Subsection 4.13.5 is relevant. It shows that once $\hat{\theta}$ is obtained, the estimate $\hat{\alpha}$ can be computed based on a single Kalman filter and smoother applied to model (4.114). It implies that once the signal mode $\hat{\theta}$ is computed (mean and mode are the same in a linear Gaussian model), it is not necessary to repeat a Newton–Raphson method to compute the mode for the state vector. We can simply formulate the linear Gaussian state space model (4.114) where the observations are replaced by $\hat{\theta}_t$ and where the observation noise is set to zero. The Kalman filter and smoothing recursions for the state vector compute the mode $\hat{\alpha}_t$ for $t = 1, \ldots, n$.

10.6.3 Mode estimation by linearisation

The computation of the mode as described in Subsection 10.6.2 can alternatively be derived by matching the first and second derivatives of the smoothing densities $p(\theta|Y_n)$ and $g(\theta|Y_n)$ where $g(\theta|Y_n)$ refers to the smoothed approximating density of the linear Gaussian model. The logdensity $\log g(\theta|Y_n)$ can be decomposed as in (10.38), that is

$$\log g(\theta|Y_n) = \log g(Y_n|\theta) + \log g(\theta) - \log g(Y_n),$$

where $\log g(Y_n|\theta)$ is the logdensity of the observation equation and is defined by (4.105). We therefore obtain

$$\dot{g}(Y_n|\theta) = H^{-1}(Y_n - \theta), \qquad \ddot{g}(Y_n|\theta) = -H^{-1}, \tag{10.46}$$

where

$$\dot{g}(\cdot|\cdot) = \frac{\partial \log g(\cdot|\cdot)}{\partial \theta}, \qquad \ddot{g}(\cdot|\cdot) = \frac{\partial^2 \log g(\cdot|\cdot)}{\partial \theta \partial \theta'}. \tag{10.47}$$

Furthermore, since we assume a linear Gaussian signal θ, we have $g(\theta) = p(\theta)$ and is given by (4.111). Finally, $\log g(Y_n)$ is the logdensity of the observations and does not depend on θ. Matching the first two derivatives of the densities

$p(\theta|Y_n)$ and $g(\theta|Y_n)$, with respect to θ, is therefore equivalent to matching the model densities $g(Y_n|\theta)$ and $p(Y_n|\theta)$, that is

$$H^{-1}(Y_n - \theta) = \dot{p}(Y_n|\theta), \qquad -H^{-1} = \ddot{p}(Y_n|\theta). \tag{10.48}$$

Since the derivatives are functions of θ themselves, we solve the equations in (10.48) by iteration. For a given value of $\theta = \tilde{\theta}$, we equalise the derivatives in (10.48) by transforming the observation vector Y_n into x and the observation disturbance variance matrix H into A for the linear Gaussian model as

$$A = -\left\{\ddot{p}(Y_n|\theta)|_{\theta=\tilde{\theta}}\right\}^{-1}, \qquad x = \tilde{\theta} + A\,\dot{p}(Y_n|\theta)|_{\theta=\tilde{\theta}}\,;$$

compare the definitions of x and A in (10.45). The application of the Kalman filter and smoother to the linear Gaussian model with observation vector $Y_n = x$ and variance matrix $H = A$, produces a new estimate $\tilde{\theta}^+$ of θ. We can replace $\tilde{\theta}$ by $\tilde{\theta}^+$ to linearise again in the form (10.48). This process leads to an iteration from which the final linearised model is the linear Gaussian model with the same conditional mode of θ given Y_n as the non-Gaussian nonlinear model.

We have shown that matching the first and second derivatives of the smoothed densities is equivalent to maximising the smoothed density of the signal $p(\theta|Y_n)$ with respect to θ. The value of θ at its maximum is the mode of θ. In the next section we will approximate more general nonlinear non-Gaussian state space models by matching the first and second derivatives of the smoothed densities with respect to the state vector α.

In case the observations can be represented by the signal plus noise model of the form (9.7) in Section 9.2, linearisation based on matching the first derivatives only may also be appropriate. We have $y_t = \theta_t + \varepsilon_t$ with $\varepsilon_t \sim p(\varepsilon_t)$ such that $p(y_t|\theta_t) = p(\varepsilon_t)$. We shall assume that y_t is univariate because it is an important case in practice and it simplifies the treatment. By matching the first derivative between $g(Y_n|\theta)$ and $p(Y_n|\theta)$ only, we obtain

$$H^{-1}(Y_n - \theta) = \dot{p}(\varepsilon), \tag{10.49}$$

where $\dot{p}(\varepsilon) = \dot{p}(Y_n|\theta)$ as defined in (10.41). For a given value of $\theta = \tilde{\theta}$ and $\varepsilon = \tilde{\varepsilon}$ with $\tilde{\varepsilon} = Y_n - \tilde{\theta}$, we equalise the first derivatives in (10.49) and transform the observation disturbance variance matrix H into A for the linear Gaussian signal plus noise model, to obtain

$$A_t = (y_t - \tilde{\theta}_t)\dot{p}(\varepsilon_t)^{-1}\Big|_{\varepsilon_t = y_t - \tilde{\theta}_t}\,,$$

where A_t is the tth diagonal element of A. The observation does not need transformation so that $x_t = y_t$ for $t = 1, \ldots, n$.

10.6.4 Mode estimation for exponential family models

An important application of these results is to observations from the exponential family of distributions. For this class of models we can compute the mode as described in Subsection 10.6.2. For density (9.3) we have

$$\log p(y_t|\theta_t) = y_t'\theta_t - b_t(\theta_t) + c_t(y_t). \tag{10.50}$$

For a given value $\tilde{\theta} = (\tilde{\theta}_1', \ldots, \tilde{\theta}_n')'$ for θ, the resulting derivatives are given by

$$\dot{p}(y_t|\theta_t) = y_t - \dot{b}_t, \qquad \ddot{p}(y_t|\theta_t) = -\ddot{b}_t,$$

where

$$\dot{b}_t = \frac{\partial b_t(\theta_t)}{\partial \theta_t}\bigg|_{\theta_t = \tilde{\theta}_t}, \qquad \ddot{b}_t = \frac{\partial^2 b_t(\theta_t)}{\partial \theta_t \partial \theta_t'}\bigg|_{\theta_t = \tilde{\theta}_t},$$

for $t = 1, \ldots, n$. These values can be substituted in (10.45) to obtain a solution for the case where signal θ is linear and Gaussian, that is

$$A_t = \ddot{b}_t^{-1}, \qquad x_t = \tilde{\theta}_t + \ddot{b}_t^{-1} y_t - \ddot{b}_t^{-1}\dot{b}_t,$$

where A_t is the tth element of the diagonal matrix A and x_t is the tth element of vector x; (block-)diagonal matrix A and vector x are defined in (10.45). Since, as shown in Section 9.3, $\ddot{b}_t = \mathrm{Var}(y_t|\theta_t)$, it is positive definite in nondegenerate cases, so for the exponential family, the method of computing the mode can always be used.

As an example, for the Poisson distribution with density (9.12), we have

$$\log p(y_t|\theta_t) = y_t\theta_t - \exp\theta_t - \log y_t!,$$

so that $b_t(\tilde{\theta}_t) = \dot{b}_t = \ddot{b}_t = \exp(\tilde{\theta}_t)$. For computing the mode, we therefore take

$$A_t = \exp(-\tilde{\theta}_t), \qquad x_t = \tilde{\theta}_t + \exp(-\tilde{\theta}_t)y_t - 1,$$

for $t = 1, \ldots, n$. Other examples of expressions for A_t and x_t for a range of exponential family models are given in Table 10.2.

10.6.5 Mode estimation for stochastic volatility model

The mode estimation of the volatility signal in a stochastic volatility model should be based on the first two derivatives. For the basic SV model (9.26) we have

$$\log p(y_t|\theta_t) = -\frac{1}{2}\big[\log 2\pi\sigma^2 + \theta_t + z_t^2\exp(-\theta_t)\big],$$

where $z_t = (y_t - \mu)/\sigma$. It follows that

$$\dot{p}_t = -\frac{1}{2}\big[1 - z_t^2\exp(-\theta_t)\big], \qquad \ddot{p}_t = -\frac{1}{2}z_t^2\exp(-\theta_t).$$

Table 10.2 Approximating model details for exponential family models.

Distribution		
Poisson	b_t	$\exp\theta_t$
	$\dot{b}_t$	$\exp\theta_t$
	$\ddot{b}_t$	$\exp\theta_t$
	$\ddot{b}_t^{-1}\dot{b}_t$	1
binary	b_t	$\log(1+\exp\theta_t)$
	$\dot{b}_t$	$\exp\theta_t(1+\exp\theta_t)^{-1}$
	$\ddot{b}_t$	$\exp\theta_t(1+\exp\theta_t)^{-2}$
	$\ddot{b}_t^{-1}\dot{b}_t$	$1+\exp\theta_t$
binomial	b_t	$k_t\log(1+\exp\theta_t)$
	$\dot{b}_t$	$k_t\exp\theta_t(1+\exp\theta_t)^{-1}$
	$\ddot{b}_t$	$k_t\exp\theta_t(1+\exp\theta_t)^{-2}$
	$\ddot{b}_t^{-1}\dot{b}_t$	$1+\exp\theta_t$
negative binomial	b_t	$k_t\{\theta_t-\log(1-\exp\theta_t)\}$
	$\dot{b}_t$	$k_t(1-\exp\theta_t)^{-1}$
	$\ddot{b}_t$	$k_t\exp\theta_t(1-\exp\theta_t)^{-2}$
	$\ddot{b}_t^{-1}\dot{b}_t$	$\exp(-\theta_t)-1$
exponential	b_t	$-\log\theta_t$
	$\dot{b}_t$	$-\theta_t^{-1}$
	$\ddot{b}_t$	θ_t^{-2}
	$\ddot{b}_t^{-1}\dot{b}_t$	$-\theta_t$

The key variables x_t and A_t for the approximating model $x_t=\theta_t+u_t$ with $u_t\sim\mathrm{N}(0,A_t)$ are given by $x_t=\tilde{\theta}_t+\ddot{b}_t^{-1}y_t-\ddot{b}_t^{-1}\dot{b}_t$ and $A_t=\ddot{b}_t^{-1}$.

Using the definitions in (10.45), we have

$$A_t=2\exp(\tilde{\theta}_t)\,/\,z_t^2,\qquad x_t=\tilde{\theta}_t+1-\exp(\tilde{\theta}_t)\,/\,z_t^2,$$

where we note that A_t is always positive as required. The method of computing the mode based on the two derivatives proceeds as before.

Next we replace $p(\varepsilon_t)=\mathrm{N}(0,1)$ in (9.26) by the Student's t-density (9.23) with $\sigma_\varepsilon^2=1$, that is

$$\log p(\varepsilon_t)=\text{constant}-\frac{\nu+1}{2}\log\big(1+\frac{\varepsilon_t^2}{\nu-2}\big),$$

where $\nu>2$ is the number of degrees of freedom Hence we obtain the SV model with t-disturbances as discussed in Subsection 9.5.3. We can represent the resulting density for y_t by

$$\log p(y_t|\theta_t) = \text{constant} - \frac{1}{2}\big[\theta_t + (\nu+1)\,\log q_t\big], \qquad q_t = 1 + \exp(-\theta_t)\frac{z_t^2}{\nu-2},$$

for $t = 1, \ldots, n$. For estimating the mode of θ_t we require

$$\dot{p}_t = -\frac{1}{2}\big[1 - (\nu+1)(q_t^{-1}-1)\big], \qquad \ddot{p}_t = \frac{1}{2}(\nu+1)(q_t^{-1}-1)q_t^{-1}.$$

Using the definitions in (10.45), we have

$$A_t = 2(\nu+1)^{-1}(\tilde{q}_t - 1)^{-1}\tilde{q}_t^2, \qquad x_t = \tilde{\theta}_t + \tilde{q}_t^2 - \frac{1}{2}A_t,$$

where $\tilde{q}_t$ is q_t evaluated at $\tilde{\theta}_t$. Since $q_t > 1$, all A_t's are positive with probability one and we can proceed with the estimation of the mode.

Other variations of the stochastic volatility model as discussed in Section 9.5 can be considered as well. Estimation of the mode can be carried out similarly for more advanced specifications of the stochastic volatility model. For example, the details of mode estimation for the SV model with leverage effects, as discussed in Subsection 9.5.5, are given by Jungbacker and Koopman (2007).

10.7 Further advances in mode estimation

The basic ideas of mode estimation, together with illustrations, for state space models with a nonlinear non-Gaussian observation equation that depends on a linear Gaussian signal, are presented in Section 10.6. Although many nonlinear non-Gaussian models in practice belong to this class of models, we need to discuss more general cases as well for completeness. We therefore start to derive the general linearisation method based on the state vector for estimating the mode of the state vector. This method is illustrated by a set of examples within the general class of nonlinear non-Gaussian state space models. We close this section by deriving optimality properties of the mode for our class of models.

10.7.1 Linearisation based on the state vector

In Section 10.6 we have considered models where the observation density, conditional on a linear Gaussian signal, is non-Gaussian or nonlinear. The presented method of mode estimation is preferred when it is sufficient to condition on the signal for capturing all nonlinear non-Gaussian features of the model. These developments are also applicable when we need to condition on the state vector rather than the signal. The second-order expansion arguments are still applicable and we can pursue this approach as in the previous section. However we find the linearisation argument for obtaining the mode more insightful, especially when models become more intricate. We therefore present the general linearisation method next.

10.7.2 Linearisation for linear state equations

Consider the non-Gaussian state space model (10.34) but with a linear state equation, that is

$$y_t \sim p(y_t|\alpha_t), \qquad \alpha_{t+1} = T_t\alpha_t + R_t\eta_t, \qquad t = 1, \ldots, n,$$

where the relation between y_t and α_t in density $p(y_t|\alpha_t)$ can be nonlinear and where $p(\alpha_1)$ and $p(\eta_t)$ can be non-Gaussian. We introduce two series of variables $\bar{y}_t$ and $\bar{\alpha}_t$ for $t = 1, \ldots, n$. Let $g(\bar{\alpha}|\bar{y})$ and $g(\bar{\alpha}, \bar{y})$ be the conditional and joint densities generated by the linear Gaussian model (3.1) where observation y_t is replaced by $\bar{y}_t$ and state α_t is replaced by $\bar{\alpha}_t$. Define $\bar{y} = (\bar{y}_1', \ldots, \bar{y}_n')'$ and $\bar{\alpha} = (\bar{\alpha}_1', \ldots, \bar{\alpha}_n')'$. We use notation $\bar{y}_t$ and $\bar{\alpha}_t$ as these quantities are not necessarily equivalent to y_t and α_t, respectively, in our treatment below. All variables x related to (3.1) will be indicated by $\bar{x}$ for $x = Z_t, H_t, T_t, R_t, Q_t, \varepsilon_t, \eta_t, a_1, P_1$ in the same way as for y_t and α_t. Let $p(\alpha|Y_n)$ and $p(\alpha, Y_n)$ be the corresponding densities generated by the general model (10.34) for the observation vector Y_n.

Taking the Gaussian model first, the mode $\hat{\bar{\alpha}}$ is the solution of the vector equation $\partial \log g(\bar{\alpha}|\bar{y})/\partial\bar{\alpha} = 0$. Now $\log g(\bar{\alpha}|\bar{y}) = \log g(\bar{\alpha}, \bar{y}) - \log g(\bar{y})$. Thus the mode is also the solution of the vector equation $\partial \log g(\bar{\alpha}, \bar{y})/\partial\bar{\alpha} = 0$. This version of the equation is easier to manage since $g(\bar{\alpha}, \bar{y})$ has a simple form whereas $g(\bar{\alpha}|\bar{y})$ does not. Since R_t is the linear Gaussian model (3.1) consists of columns of I_m, $\bar{\eta}_t = \bar{R}_t'(\bar{\alpha}_{t+1} - \bar{T}_t\bar{\alpha}_t)$. Assuming that $g(\bar{\alpha}_1) = \mathrm{N}(\bar{a}_1, \bar{P}_1)$, we therefore have

$$\begin{aligned}\log g(\bar{\alpha}, \bar{y}) = \text{constant} &- \frac{1}{2}(\bar{\alpha}_1 - \bar{a}_1)'\bar{P}_1^{-1}(\bar{\alpha}_1 - \bar{a}_1) \\ &- \frac{1}{2}\sum_{t=1}^{n}\left(\bar{\alpha}_{t+1} - \bar{T}_t\bar{\alpha}_t\right)'\bar{R}_t\bar{Q}_t^{-1}\bar{R}_t'(\bar{\alpha}_{t+1} - \bar{T}_t\bar{\alpha}_t) \\ &- \frac{1}{2}\sum_{t=1}^{n}\left(\bar{y}_t - \bar{Z}_t\bar{\alpha}_t\right)'\bar{H}_t^{-1}(\bar{y}_t - \bar{Z}_t\bar{\alpha}_t). \qquad (10.51)\end{aligned}$$

Differentiating with respect to $\bar{\alpha}_t$ and equating to zero gives the equations

$$\begin{aligned}&(d_t - 1)\bar{P}_1^{-1}(\bar{\alpha}_1 - \bar{a}_1) - d_t\bar{R}_{t-1}\bar{Q}_{t-1}^{-1}\bar{R}_{t-1}'(\bar{\alpha}_t - \bar{T}_{t-1}\bar{\alpha}_{t-1}) \\ &\quad + \bar{T}_t'\bar{R}_t\bar{Q}_t^{-1}\bar{R}_t'(\bar{\alpha}_{t+1} - \bar{T}_t\bar{\alpha}_t) + \bar{Z}_t'\bar{H}_t^{-1}(\bar{y}_t - \bar{Z}_t\bar{\alpha}_t) = 0, \qquad (10.52)\end{aligned}$$

for $t = 1, \ldots, n$, where $d_1 = 0$ and $d_t = 1$ for $t = 2, \ldots, n$, together with the equation

$$\bar{R}_n\bar{Q}_n\bar{R}_n'(\bar{\alpha}_{n+1} - \bar{T}_n\bar{\alpha}_n) = 0.$$

The solution to these equations is the conditional mode $\hat{\bar{\alpha}}$. Since $g(\bar{\alpha}|\bar{y})$ is Gaussian the mode is equal to the mean so $\hat{\bar{\alpha}}$ can be routinely calculated by the Kalman filter and smoother. We conclude that linear equations of the form

(10.52) can be solved by the Kalman filter and smoother which is efficient computationally.

Assuming that the nonlinear non-Gaussian state space model (10.34) is sufficiently well behaved, the mode $\hat{\alpha}$ of $p(\alpha|Y_n)$ is the solution of the vector equation

$$\frac{\partial \log p(\alpha|Y_n)}{\partial \alpha} = 0$$

and hence, as in the Gaussian case, of the equation

$$\frac{\partial \log p(\alpha, Y_n)}{\partial \alpha} = 0,$$

where

$$\log p(\alpha, Y_n) = \text{constant} + \log p(\alpha_1) + \sum_{t=1}^{n} \left[\log p(\eta_t) + \log p(y_t|\theta_t)\right], \qquad (10.53)$$

with $\eta_t = R_t'(\alpha_{t+1} - T_t\alpha_t)$. The mode $\hat{\alpha}$ is the solution of the vector equations

$$\begin{aligned}\frac{\partial \log p(\alpha, Y_n)}{\partial \alpha_t} &= (1-d_t)\frac{\partial \log p(\alpha_1)}{\partial \alpha_1} + d_t R_{t-1}\frac{\partial \log p(\eta_{t-1})}{\partial \eta_{t-1}} \\ &\quad - T_t' R_t \frac{\partial \log p(\eta_t)}{\partial \eta_t} + \frac{\partial \log p(y_t|\alpha_t)}{\partial \alpha_t} = 0, \qquad (10.54)\end{aligned}$$

for $t = 1, \ldots, n$, where, as before, $d_1 = 0$ and $d_t = 1$ for $t = 2, \ldots, n$, together with the equation

$$R_n \frac{\partial \log p(\eta_n)}{\partial \eta_n} = 0.$$

We solve these equations by iteration, where at each step we linearise, put the result in the form (10.52) and solve by the Kalman filter and smoother. The final linearised model in the iteration is then the linear Gaussian model with the same conditional mode of $\bar{\alpha}$ given $\bar{y}$ as the non-Gaussian model with the conditional mode α given Y_n.

We first consider the linearisation of the state component in equations (10.54) for a case where the state disturbances η_t are non-Gaussian. Suppose that $\tilde{\eta} = [\tilde{\eta}_1', \ldots, \tilde{\eta}_n']'$ is a trial value of $\eta = (\eta_1', \ldots, \eta_n')'$ where $\tilde{\eta}_t = R_t'(\tilde{\alpha}_{t+1} - T_t\tilde{\alpha}_t)$. We shall confine ourselves to the situation where the elements η_{it} of η_t are mutually independent, in other words Q_t is diagonal. Then the state contribution to the conditional mode equations (10.54) is

$$d_t \sum_{i=1}^{r} R_{i,t-1}\frac{\partial \log p(\eta_{i,t-1})}{\partial \eta_{i,t-1}} - T_t' \sum_{i=1}^{r} R_{it}\frac{\partial \log p(\eta_{it})}{\partial \eta_{it}},$$

where we denote the ith column of R_t by R_{it} for $t = 1, \ldots, n$ and $i = 1, \ldots, r$. Define

$$q_{it} = \left. \frac{\partial \log p(\eta_{it})}{\partial \eta_{it}} \right|_{\eta_t = \tilde{\eta}_t}.$$

The linearised form at $\eta = \tilde{\eta}$ is then given by

$$d_t \sum_{i=1}^{r} R_{i,t-1} q_{i,t-1} \; - \; T_t' \sum_{i=1}^{r} R_{it} q_{it},$$

which has the same form as the state contribution of (10.52) when we set

$$\bar{Q}_t^{-1} = \mathrm{diag}(q_{1t}, \ldots, q_{rt}),$$

since $\eta_t = R_t'(\alpha_{t+1} - T_t \alpha_t)$ for $t = 1, \ldots, n$. All other state variables $\bar{x}$ are set equal to x for $x = Z_t, H_t, T_t, R_t, \ldots$. In the iterative estimation of $\hat{\alpha}$ the Kalman filter and smoother can be used to update the trial value $\tilde{\alpha}$ and, consequently, $\tilde{\eta}$.

10.7.3 Linearisation for nonlinear models

Consider the non-Gaussian state space model (10.34) where the density of y_t is subject to a nonlinear signal, $p(y_t|\theta_t) = p(y_t|\alpha_t)$, where

$$\theta_t = Z_t(\alpha_t), \qquad t = 1, \ldots, n. \tag{10.55}$$

Our objective is to find an approximating linear Gaussian model with the same conditional mode of $\bar{\alpha}$ given $\bar{y}$ as the nonlinear model. We do this by a technique which is slightly different, though simpler, than that used for the non-Gaussian models. The basic idea is to linearise the observation and state equations (10.34) and (10.55) directly, which immediately delivers an approximating linear Gaussian model. We then iterate to ensure that this approximating model has the same conditional mode as the original nonlinear model.

Taking first the nonlinear signal (10.55), let $\tilde{\alpha}_t$ be a trial value of α_t. Expanding about $\tilde{\alpha}_t$ gives approximately

$$Z_t(\alpha_t) = Z_t(\tilde{\alpha}_t) + \dot{Z}_t(\tilde{\alpha}_t)(\alpha_t - \tilde{\alpha}_t),$$

where $\dot{Z}(\alpha_t) = \partial Z_t(\alpha_t)/\partial \alpha_t'$. From an approximation to (10.55) we obtain

$$y_t = \bar{d}_t + \dot{Z}_t(\tilde{\alpha}_t)\alpha_t + \varepsilon_t, \qquad \bar{d}_t = Z_t(\tilde{\alpha}_t) - \dot{Z}_t(\tilde{\alpha}_t)\tilde{\alpha}_t, \tag{10.56}$$

which is the linear observation equation in the mean adjusted form; see Subsection 4.3.3. Similarly, if we expand the state updating function in (10.34) about $\tilde{\alpha}_t$ we obtain approximately

$$T_t(\alpha_t) = T_t(\tilde{\alpha}_t) + \dot{T}_t(\tilde{\alpha}_t)(\alpha_t - \tilde{\alpha}_t),$$

where $\dot{T}(\alpha_t) = \partial T_t(\alpha_t)/\partial \alpha_t'$. Thus we obtain the linearised relation

$$\alpha_{t+1} = \bar{c}_t + \dot{T}_t(\tilde{\alpha}_t)\alpha_t + R_t(\tilde{\alpha}_t)\eta_t, \qquad \bar{c}_t = T_t(\tilde{\alpha}_t) - \dot{T}_t(\tilde{\alpha}_t)\tilde{\alpha}_t. \tag{10.57}$$

We approximate this nonlinear model at its conditional mode by the linear Gaussian model with mean adjustments as discussed in Subsection 4.3.3 with the modified form

$$\begin{aligned} \bar{y}_t &= \bar{d}_t + \bar{Z}_t\alpha_t + \bar{\varepsilon}_t, & \bar{\varepsilon}_t &\sim \mathrm{N}(0, \bar{H}_t), \\ \alpha_{t+1} &= \bar{c}_t + \bar{T}_t\alpha_t + \bar{R}_t\eta_t, & \bar{\eta}_t &\sim \mathrm{N}(0, \bar{Q}_t), \end{aligned} \tag{10.58}$$

where

$$\bar{Z}_t = \dot{Z}_t(\tilde{\alpha}_t), \quad \bar{H}_t = H_t(\tilde{\alpha}_t), \quad \bar{T}_t = \dot{T}_t(\tilde{\alpha}_t), \quad \bar{R}_t = R_t(\tilde{\alpha}_t), \quad \bar{Q}_t = Q_t(\tilde{\alpha}_t),$$

for $t = 1, \ldots, n$. The Kalman filter of the form (4.25) in Subsection 4.3.3 can be applied to model (10.58). We use the output of the Kalman filter (4.26) to define a new $\tilde{\alpha}_t$ which gives a new approximating model (10.58), and we continue to iterate as in Subsection 10.6.3 until convergence is achieved. Denote the resulting value of α by $\hat{\alpha}$.

10.7.4 Linearisation for multiplicative models

Shephard (1994b) considered the multiplicative trend and seasonal model with additive Gaussian observation noise. We consider here a simple version of this model with the trend modelled as a local level and the seasonal given by a single trigonometric term such as given in (3.6) with $s = 3$. We have

$$y_t = \mu_t\gamma_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}\big(0, \sigma_\varepsilon^2\big),$$

with

$$\alpha_{t+1} = \begin{pmatrix} \mu_{t+1} \\ \gamma_{t+1} \\ \gamma_{t+1}^* \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\lambda & \sin\lambda \\ 0 & -\sin\lambda & \cos\lambda \end{bmatrix} \alpha_t + \begin{pmatrix} \eta_t \\ \omega_t \\ \omega_t^* \end{pmatrix},$$

and $\lambda = 2\pi/3$. It follows that $Z_t(\alpha_t) = \mu_t\gamma_t$ and $\dot{Z}_t(\alpha_t) = (\gamma_t, \mu_t, 0)$ which lead us to the approximating model

$$\tilde{y}_t = (\tilde{\gamma}_t, \tilde{\mu}_t, 0)\alpha_t + \varepsilon_t,$$

where $\tilde{y}_t = y_t + \tilde{\mu}_t\tilde{\gamma}_t$.

Another example of a multiplicative model that we consider is

$$y_t = \mu_t\varepsilon_t, \qquad \mu_{t+1} = \mu_t\xi_t,$$

where ε_t and ξ_t are mutually and serially uncorrelated Gaussian disturbance terms. For the general model (9.36) and (9.37) we have $\alpha_t = (\mu_t, \varepsilon_t, \xi_t)'$, $\eta_t =$

$(\varepsilon_{t+1}, \xi_{t+1})'$ $Z_t(\alpha_t) = \mu_t \varepsilon_t$, $H_t = 0$, $T_t(\alpha_t) = (\mu_t \xi_t, 0, 0)'$, $R_t = [0, I_2]'$ and Q_t is a 2×2 diagonal matrix. It follows that

$$\dot{Z}_t(\alpha_t) = (\varepsilon_t, \mu_t, 0), \qquad \dot{T}_t(\alpha_t) = \begin{bmatrix} \xi_t & 0 & \mu_t \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The approximating model (10.56) and (10.57) reduces to

$$\tilde{y}_t = \tilde{\varepsilon}_t \mu_t + \tilde{\mu}_t \varepsilon_t, \qquad \mu_{t+1} = -\tilde{\mu}_t \tilde{\xi}_t + \tilde{\xi}_t \mu_t + \tilde{\mu}_t \xi_t,$$

with $\tilde{y}_t = y_t + \tilde{\mu}_t \tilde{\varepsilon}_t$. Thus the Kalman filter and smoother can be applied to an approximating time-varying local level model with state vector $\alpha_t = \mu_t$.

10.7.5 An optimal property for the mode

We will emphasise in the next chapter the use of the mode $\hat{\alpha}$ of $p(\alpha|Y_n)$ to obtain a linear approximating model which we use for simulation. If, however, the sole object of the investigation was to estimate α, then $\hat{\alpha}$ could be used for the purpose without recurse to simulation; indeed, this was the estimator used by Durbin and Koopman (1992) and an approximation to it was used by Fahrmeir (1992).

The property that the conditional mode is the most probable value of the state vector given the observations can be regarded as an optimality property; we now consider a further optimality property possessed by the conditional mode. To find it we examine the analogous situation in maximum likelihood estimation. The maximum likelihood estimate of a parameter ψ is the most probable value of it given the observations and is well known to be asymptotically efficient. To develop a finite-sample property analogous to asymptotic efficiency, Godambe (1960) and Durbin (1960) introduced the idea of unbiased estimating equations and Godambe showed that the maximum likelihood estimate of scalar ψ is the solution to an unbiased estimating equation which has a minimum variance property. This can be regarded as a finite-sample analogue of asymptotic efficiency. The extension to multidimensional ψ was indicated by Durbin (1960). Since that time there have been extensive developments of this basic idea, as can be seen from the collection of papers edited by Basawa, Godambe and Taylor (1997). Following Durbin (1997), we now develop a minimum-variance unbiased estimating equation property for the conditional mode estimate $\hat{\alpha}$ of the random vector α.

If α^* is the unique solution for α of the $mn \times 1$ vector equation $H(\alpha, Y_n) = 0$ and if $\mathrm{E}[H(\alpha, Y_n)] = 0$, where expectation is taken with respect to the joint density $p(\alpha, Y_n)$, we say that $H(\alpha, Y_n) = 0$ is an *unbiased estimating equation*. It is obvious that the equation can be multiplied through by an arbitrary nonsingular matrix and still give the same solution α^*. We therefore standardise $H(\alpha, Y_n)$ in the way that is usual in estimating equation theory

and multiply it by $[\mathrm{E}\{\dot{H}(\alpha, Y_n)\}]^{-1}$, where $\dot{H}(\alpha, Y_n) = \partial H(\alpha, Y_n)/\partial\alpha'$, and then seek a minimum variance property for the resulting function $h(\alpha, Y_n) = [\mathrm{E}\{\dot{H}(\alpha, Y_n)\}]^{-1} H(\alpha, Y_n)$.

Let

$$\mathrm{Var}[h(\alpha, Y_n)] = \mathrm{E}[h(\alpha, Y_n)\, h(\alpha, Y_n)'],$$
$$\mathcal{J} = \mathrm{E}\left[\frac{\partial \log p(\alpha, Y_n)}{\partial\alpha}\frac{\partial \log p(\alpha, Y_n)}{\partial\alpha'}\right].$$

Under mild conditions that are likely to be satisfied in many practical cases, Durbin (1997) showed that $\mathrm{Var}[h(\alpha, Y_n)] - \mathcal{J}^{-1}$ is non-negative definite. If this is a zero matrix we say that the corresponding equation $H(\alpha, Y_n) = 0$ is an *optimal estimating equation*. Now take $H(\alpha, Y_n) = \partial \log p(\alpha, Y_n)/\partial\alpha$. Then $\mathrm{E}[\dot{H}(\alpha, Y_n)] = -\mathcal{J}$, so $h(\alpha, Y_n) = -\mathcal{J}^{-1}\partial \log p(\alpha, Y_n)/\partial\alpha$. Thus $\mathrm{Var}[h(\alpha, Y_n)] = \mathcal{J}^{-1}$ and consequently the equation $\partial \log p(\alpha, Y_n)/\partial\alpha = 0$ is optimal. Since $\hat{\alpha}$ is the solution of this, it is the solution of an optimal estimating equation. In this sense the conditional mode has an optimality property analogous to that of maximum likelihood estimates of fixed parameters in finite samples.

We have assumed above that there is a single mode and the question arises whether multimodality will create complications. If multimodality is suspected it can be investigated by using different starting points and checking whether iterations from them converge to the same mode. In none of the cases we have examined has multimodality of $p(\alpha|Y_n)$ caused any difficulties. For this reason we regard it as unlikely that it will give rise to problems in routine time series analysis. If, however, multimodality were to occur in a particular case, we would suggest fitting a linear Gaussian model to the data at the outset and using this to define the first importance density $g_1(\eta|Y_n)$ and conditional joint density $g_1(\eta, Y_n)$. Simulation is employed to obtain a first estimate $\tilde{\eta}^{(1)}$ of $\mathrm{E}(\eta|Y_n)$ and from this a first estimate $\tilde{\theta}_t^{(1)}$ of θ_t is calculated for $t = 1, \ldots, n$. Now linearise the true densities at $\tilde{\eta}^{(1)}$ or $\tilde{\theta}_t^{(1)}$ to obtain a new approximating linear Gaussian model which defines a new $g(\eta|Y_n)$, $g_2(\eta|Y_n)$, and a new $g(\eta, Y_n)$, $g_2(\eta, Y_n)$. Simulation using these gives a new estimate $\tilde{\eta}^{(2)}$ of $\mathrm{E}(\eta|Y_n)$. This iterative process is continued until adequate convergence is achieved. We emphasise, however, that it is not necessary for the final value of α at which the model is linearised to be a precisely accurate estimate of either the mode or the mean of $p(\alpha|Y_n)$. The only way that the choice of the value of α used as the basis for the simulation affects the final estimate $\hat{x}$ is in the variances due to simulation as we shall show later. Where necessary, the simulation sample size can be increased to reduce these error variances to any required extent. It will be noted that we are basing these iterations on the mean, not the mode. Since the mean, when it exists, is unique, no question of 'multimeanality' can arise.

10.8 Treatments for heavy-tailed distributions

In this section we consider both approximate and exact treatments of linear state space models with disturbances from heavy-tailed distributions. First we apply linearisation techniques based on the first derivative only for different model specifications. This shows that the general methodology can lead to practical methods for treating outliers and breaks in time series. The methods lead to mode estimates. We further discuss simulation treatments for heavy-tailed models which are relatively simple and lead to exact estimation methods subject to simulation error.

10.8.1 Mode estimation for models with heavy-tailed densities

In a signal plus noise model, $y_t = \theta_t + \varepsilon_t$ where the signal θ_t is linear Gaussian and the noise distribution for ε_t has a heavy-tailed density, we can estimate the mode using the first two derivatives or using the first derivative only. For example, consider the logdensity of the Student's t-distribution as specified in (9.23). To estimate the mode of θ_t based on (10.44) we require expressions for A_t, the tth element of diagonal matrix A, and x_t, the tth element of vector x, where A and x are given by (10.45). The variables A_t and x_t rely on $\dot{p}(\varepsilon_t) = \dot{p}(y_t|\theta_t)$ and $\ddot{p}(\varepsilon_t) = \ddot{p}(y_t|\theta_t)$ which are given by

$$\dot{p}(\varepsilon_t) = (\nu+1)s_t^{-1}\tilde{\varepsilon}_t, \qquad \ddot{p}(\varepsilon_t) = (\nu+1)s_t^{-1}\left[2s_t^{-1}\tilde{\varepsilon}_t - 1\right],$$

where $s_t = (\nu-2)\sigma_\varepsilon^2 + \tilde{\varepsilon}_t^2$. Since we cannot rule out a positive value for $\ddot{p}(\varepsilon_t)$, the t-density $p(\varepsilon_t)$ is not logconcave in θ_t and the variance A_t can become negative. However, the method of mode estimation is still applicable in this case; see Jungbacker and Koopman (2007). When we prefer to estimate the mode using the first derivative only as in (10.49), we can adopt the linear Gaussian signal plus noise model with its observation variance given by

$$A_t = (\nu+1)^{-1}s_t,$$

for $t = 1, \ldots, n$. We proceed by applying the Kalman filter and smoother to obtain a new smooth estimate of θ_t.

Similar computations can be adopted for other heavy-tailed densities. In the case of the mixture of normals model with density (9.24) we obtain,

$$\log p(\varepsilon_t) = \log\left\{\lambda^* e_t(\sigma_\varepsilon^2) \;+\; [1-\lambda^*]\, e_t(\sigma_\varepsilon^2\chi)\right\},$$

where $e_t(z) = \exp(-\frac{1}{2}\varepsilon_t^2/z)\,/\,\sqrt{2\pi z}$, with $1-\lambda^*$ as the proportion of outliers and χ as the multiplier of the variance for the outliers. For the computation of the mode using the first derivative only, we require

$$A_t = p(\tilde{\varepsilon}_t)\left\{\lambda^*\,\sigma_\varepsilon^{-2}\,\tilde{e}_t(\sigma_\varepsilon^2) \;+\; [1-\lambda^*]\,(\sigma_\varepsilon^2\chi)^{-1}\,\tilde{e}_t(\sigma_\varepsilon^2\chi)\right\}^{-1}$$

where $\tilde{\varepsilon}_t$ is a particular value of ε_t and $\tilde{e}_t(z)$ is $e_t(z)$ evaluated at $\varepsilon_t = \tilde{\varepsilon}_t$ for any $z > 0$.

Finally, in the case of the general error density (9.25) we obtain,

$$\log p(\varepsilon_t) = \text{constant} - c(\ell) \left| \frac{\varepsilon_t}{\sigma_\varepsilon} \right|^\ell ,$$

for coefficient $1 < \ell < 2$ and with $c(\ell)$ as a known function of ℓ. For given values $\varepsilon_t = \tilde{\varepsilon}_t$, we can compute the model using the first derivative only by computing

$$A_t = \text{sign}(\tilde{\varepsilon}_t) \frac{\tilde{\varepsilon}_t \sigma_\varepsilon}{c(\ell)\,\ell} \left| \frac{\tilde{\varepsilon}_t}{\sigma_\varepsilon} \right|^{1-\ell} .$$

All A_t's are positive with probability one and we can proceed with the estimation of the mode as before.

10.8.2 Mode estimation for state errors with t-distribution

As an illustration of mode estimation for linear models with a t-distribution, we consider the local level model (2.3) but with a t-distribution for the state error term η_t. We obtain

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t, \qquad & \varepsilon_t &\sim \mathrm{N}\big(0, \sigma_\varepsilon^2\big), \\ \alpha_{t+1} &= \alpha_t + \eta_t, \qquad & \eta_t &\sim t_\nu, \end{aligned}$$

and we assume that $\alpha_1 \sim \mathrm{N}(0, \kappa)$ with $\kappa \to \infty$. By adopting the same arguments for linearisation using one derivative only, we obtain

$$A_t^* = (\nu + 1)^{-1} s_t^*,$$

where $s_t^* = (\nu - 2)\sigma_\eta^2 + \tilde{\eta}_t^2$, for $t = 1, \ldots, n$. Starting with initial values for $\tilde{\eta}_t$, we compute A_t^* and apply the Kalman filter and disturbance smoother to the approximating Gaussian local level model with

$$y_t = \alpha_t + \varepsilon_t, \qquad \alpha_{t+1} = \alpha_t + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, A_t^*).$$

New values for the smoothed estimates $\tilde{\eta}_t$ are used to compute new values for A_t^* until convergence to $\hat{\eta}_t$. When we assume that both disturbances ε_t and η_t in the local level model (2.3) are generated by t-distributions, we can obtain the mode by computing both A_t and A_t^* and adopt them as variances for the two corresponding disturbances in the linear Gaussian local level model.

10.8.3 A simulation treatment for t-distribution model

In some cases it is possible to construct simulations by using antithetic variables without importance sampling. For example, it is well-known that if a random

variable u_t has the standard t-distribution with ν degrees of freedom then u_t has the representation

$$u_t = \frac{\nu^{1/2}\varepsilon_t^*}{c_t^{1/2}}, \qquad \varepsilon_t^* \sim \mathrm{N}(0,1), \qquad c_t \sim \chi^2(\nu), \qquad \nu > 2, \tag{10.59}$$

where ε_t^* and c_t are independent. In the case where ν is not an integer we take $\frac{1}{2}c_t$ as a gamma variable with parameter $\frac{1}{2}\nu$. It follows that if we consider the case where ε_t is univariate and we take ε_t in model (9.4) to have logdensity (9.23) then ε_t has the representation

$$\varepsilon_t = \frac{(\nu-2)^{1/2}\sigma_\varepsilon \varepsilon_t^*}{c_t^{1/2}}, \tag{10.60}$$

where ε_t^* and c_t are as in (10.59). Now take $\varepsilon_1^*, \ldots, \varepsilon_n^*$ and $c_1, \ldots, c_n$ to be mutually independent. Then conditional on $c_1, \ldots, c_n$ fixed, model (9.4) and (9.2), with $\eta_t \sim \mathrm{N}(0, Q_t)$, is a linear Gaussian model with $H_t = \mathrm{Var}(\varepsilon_t) = (\nu-2)\sigma_\varepsilon^2 c_t^{-1}$. Put $c = (c_1, \ldots, c_n)'$. We now show how to estimate the conditional means of functions of the state using simulation samples from the distribution of c.

Suppose first that α_t is generated by the linear Gaussian model $\alpha_{t+1} = T_t\alpha_t + R_t\eta_t$, $\eta_t \sim \mathrm{N}(0, Q_t)$, and that as in (11.12) we wish to estimate

$$\begin{aligned}
\bar{x} &= \mathrm{E}[x^*(\eta)|y] \\
&= \int x^*(\eta) p(c, \eta|y)\, dc\, d\eta \\
&= \int x^*(\eta) p(\eta|c, y) p(c|y)\, dc\, d\eta \\
&= \int x^*(\eta) p(\eta|c, y) p(c, y) p(y)^{-1} dc\, d\eta \\
&= p(y)^{-1} \int x^*(\eta) p(\eta|c, y) p(y|c) p(c)\, dc\, d\eta.
\end{aligned} \tag{10.61}$$

For given c, the model is linear and Gaussian. Let

$$\bar{x}(c) = \int x^*(\eta) p(\eta|c, y)\, d\eta.$$

For many cases of interest, $\bar{x}(c)$ is easily calculated by the Kalman filter and smoother, as in Chapters 4 and 5; to begin with, let us restrict attention to these cases. We have

$$p(y) = \int p(y, c)\, dc = \int p(y|c) p(c)\, dc,$$

where $p(y|c)$ is the likelihood given c which is easily calculated by the Kalman filter as in 7.2. Denote expectation with respect to density $p(c)$ by E_c Then from (10.61),

$$\bar{x} = \frac{\mathrm{E}_c[\bar{x}(c)p(y|c)]}{\mathrm{E}_c[p(y|c)]}. \tag{10.62}$$

We estimate this by simulation. Independent simulation samples $c^{(1)}, c^{(2)}, \ldots$ of c are easily obtained since c is a vector of independent χ^2_ν variables. We suggest that antithetic values of χ^2_ν are employed for each element of c, either in balanced pairs or balanced sets of four as described in Subsection 11.4.3. Suppose that values $c^{(1)}, \ldots, c^{(N)}$ have been selected. Then estimate $\bar{x}$ by

$$\hat{x} = \frac{\sum_{i=1}^N \bar{x}\big(c^{(i)}\big)p\big(y|c^{(i)}\big)}{\sum_{i=1}^N p\big(y|c^{(i)}\big)}. \tag{10.63}$$

When $\bar{x}(c)$ cannot be computed by the Kalman filter and smoother we first draw a value of $c^{(i)}$ as above and then for the associated linear Gaussian model that is obtained when this value $c^{(i)}$ is fixed we draw a simulated value $\eta^{(i)}$ of η using the simulated smoother of Section 4.9, employing antithetics independently for both $c^{(i)}$ and $\eta^{(i)}$. The value $x^*(\eta^{(i)})$ is then calculated for each $\eta^{(i)}$. If there are N pairs of values $c^{(i)}, \eta^{(i)}$ we estimate $\bar{x}$ by

$$\hat{x}^* = \frac{\sum_{i=1}^N x^*\big(\eta^{(i)}\big)p\big(y|c^{(i)}\big)}{\sum_{i=1}^N p\big(y|c^{(i)}\big)}. \tag{10.64}$$

Since we now have sampling variation arising from the drawing of values of η as well as from drawing values of c, the variance of $\hat{x}^*$ will be larger than of $\hat{x}$ for a given value of N. We present formulae (10.63) and (10.64) at this point for expository convenience in advance of the general treatment of analogous formulae in Section 11.5.

Now consider the case where the error term ε_t in the observation equation is $\mathrm{N}(0, \sigma^2_\varepsilon)$ and where the elements of the error vector η_t in the state equation are independently distributed as Student's t. For simplicity assume that the number of degrees of freedom in these t-distributions are all equal to ν, although there is no difficulty in extending the treatment to the case where some of the degrees of freedom differ or where some elements are normally distributed. Analogously to (10.60) we have the representation

$$\eta_{it} = \frac{(\nu-2)^{1/2}\sigma_{\eta i}\eta^*_{it}}{c_{it}^{1/2}}, \qquad \eta^*_{it} \sim \mathrm{N}(0,1), \qquad c_{it} \sim \chi^2_\nu, \qquad \nu > 2, \tag{10.65}$$

for $i = 1, \ldots, r$ and $t = 1, \ldots, n$, where $\sigma^2_{\eta i} = \mathrm{Var}(\eta_{it})$. Conditional on $c_{11}, \ldots, c_{rn}$ held fixed, the model is linear and Gaussian with $H_t = \sigma^2_\varepsilon$ and

$\eta_t \sim \mathrm{N}(0, Q_t)$ where $Q_t = \mathrm{diag}[(\nu-2)\sigma_{\eta 1}^2 c_{1t}^{-1}, \ldots, (\nu-2)\sigma_{\eta r}^2 c_{rt}^{-1}]$. Formulae (10.63) and (10.64) remain valid except that $c^{(i)}$ is now a vector with r elements. The extension to the case where both ε_t and elements of η_t have t-distributions is straightforward.

The idea of using representation (10.59) for dealing with disturbances with t-distributions in the local level model by means of simulation was proposed by Shephard (1994b) in the context of MCMC simulation.

10.8.4 A simulation treatment for mixture of normals model

An alternative to the t-distribution for representing error distributions with heavy tails is to employ the Gaussian mixture density (9.24), which for univariate ε_t we write in the form

$$p(\varepsilon_t) = \lambda^* \mathrm{N}\big(0, \sigma_\varepsilon^2\big) + (1-\lambda^*)\mathrm{N}\big(0, \chi\sigma_\varepsilon^2\big), \qquad 0 < \lambda^* < 1. \tag{10.66}$$

It is obvious that values of ε_t with this density can be realised by means of a two-stage process in which we first select the value of a binomial variable b_t such that $\Pr(b_t = 1) = \lambda^*$ and $\Pr(b_t = 0) = 1 - \lambda^*$, and then take $\varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2)$ if $b_t = 1$ and $\varepsilon_t \sim \mathrm{N}(0, \chi\sigma_\varepsilon^2)$ if $b_t = 0$. Assume that the state vector α_t is generated by the linear Gaussian model $\alpha_{t+1} = T_t\alpha_t + R_t\eta_t$, $\eta_t \sim \mathrm{N}(0, Q_t)$. Putting $b = (b_1, \ldots, b_n)'$, it follows that for b given, the state space model is linear and Gaussian. We can therefore employ the same approach for the mixture distribution that we used for the t-distribution in the previous subsection, giving as in (10.61),

$$\bar{x} = p(y)^{-1} M^{-1} \sum_{j=1}^{M} \int x^*(\eta) p\big(\eta | b_{(j)}, y\big) p\big(y | b_{(j)}\big) p\big(b_{(j)}\big)\, d\eta, \tag{10.67}$$

where $b_{(1)}, \ldots, b_{(M)}$ are the $M = 2^n$ possible values of b. Let

$$\bar{x}(b) = \int x^*(\eta) p(\eta | b, y)\, d\eta,$$

and consider cases where this can be calculated by the Kalman filter and smoother. Denote expectation over the distribution of b by E_b. Then

$$p(y) = M^{-1} \sum_{j=1}^{M} p\big(y | b_{(j)}\big) p\big(b_{(j)}\big) = \mathrm{E}_b[p(y|b)],$$

and analogously to (10.62) we have,

$$\bar{x} = \frac{\mathrm{E}_b[\bar{x}(b) p(y|b)]}{\mathrm{E}_b[p(y|b)]}. \tag{10.68}$$

We estimate this by simulation. A simple way to proceed is to choose a sequence $b^{(1)}, \ldots, b^{(N)}$ of random values of b and then estimate $\bar{x}$ by

$$\hat{x} = \frac{\sum_{i=1}^{N} \bar{x}\big(b^{(i)}\big) p\big(y|b^{(i)}\big)}{\sum_{i=1}^{N} p\big(y|b^{(i)}\big)}. \tag{10.69}$$

Variability in this formula arises only from the random selection of b. To construct antithetic variables for the problem we consider how this variability can be restricted while preserving correct overall probabilities. We suggest the following approach. Consider the situation where the probability $1-\lambda^*$ in (10.66) of taking $\mathrm{N}(0, \chi\sigma_\varepsilon^2)$ is small. Take $1 - \lambda^* = 1/B$ where B is an integer, say $B = 10$ or 20. Divide the simulation sample of values of b into K blocks of B, with $N = KB$. Within each block, and for each $t = 1, \ldots, n$, choose integer j randomly from 1 to B, put the jth value in the block as $b_t = 0$ and the remaining $B - 1$ values in the block as $b_t = 1$. Then take $b^{(i)} = (b_1, \ldots, b_n)'$ with $b_1, \ldots, b_n$ defined in this way for $i = 1, \ldots, N$ and use formula (10.69) to estimate $\bar{x}$. With this procedure we have ensured that for each i, $\Pr(b_t = 1) = \lambda^*$ as desired, with b_s and b_t independent for $s \neq t$, while enforcing balance in the sample by requiring that within each block b_t has exactly $B - 1$ values of 1 and one value of 0. Of course, choosing integers at random from 1 to B is a much simpler way to select a simulation sample than using the simulation smoother.

The restriction of B to integer values is not a serious drawback since the results are insensitive to relatively small variations in the value of λ^*, and in any case the value of λ^* is normally determined on a trial-and-error basis. It should be noted that for purposes of estimating mean square errors due to simulation, the numerator and denominator of (10.69) should be treated as composed of M independent values.

The idea of using the binomial representation of (10.66) in MCMC simulation for the local level model was proposed by Shephard (1994b).