The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2016-11

Kalman Filtering and Sequential Bayesian Analysis

Refik Soyer
Department of Decision Sciences
The George Washington University, USA

Article type: Advanced Review

Kalman Filtering and Sequential Bayesian Analysis **Article ID**

Refik Soyer

The George Washington University

Keywords

State Space Modeling, Smoothing, Prediction, Dynamic Linear Models, Particle Filtering.

Abstract

In this paper we present an overview of the state of the art in Kalman filtering and dynamic Bayesian linear and nonlinear models. We present some of the basic results including the derivation of Kalman filtering equations as well as recent advances in Kalman filter models and their extensions including non-Gaussian state-space models. In so doing, we take a Bayesian perspective and discuss parameter learning in state-space models which typically involves Markov chain Monte Carlo and sequential Monte Carlo methods. We present particle filtering and Bayesian particle learning techniques for state space models and discuss recent advances.

Introduction and Overview

In his keynote speech to the 1985 American Control Conference, Arthur Gelb mentioned that during the 1960-1984 period about 3000 papers had been published on Kalman filtering; see Gelb (1986). Today, a search in Google with keyword "Kalman filter" gives about 3,370,000 hits and the search in Google Scholar returns more than 735,000 articles and books on the Kalman filter. Since the publication of the original paper of Kalman (1960) and Kalman and Bucy (1961) which primarily focused on problems in control and communication engineering, Kalman filtering and the associated state space models have found applications in many diverse areas. These include industrial applications in different branches of engineering [Auger et al. (2013)], applications in medical sciences and medical imaging [Artemiev et al. (2001) and Prado, West and Krystal (2001)], in meteorology [Gauthier, Courtier, and Moll (1993)] and oceanography [Evensen (1992) and Bertino, Evensen, and Wackernage (2003)] as well as applications in economics and finance as in Schneider (1988), Babbs and Nowman (1999), and Johannes, Korteweg, and Polson (2014).

Applications in oceanography have contributed to ensemble Kalman filter, see Evensen (1992) for a review. Interest in real time forecasting and volatility modeling in financial econometrics contributed to novel applications of Kalman filtering as well as to new

methodology development; see Lopes and Tsay (2011) and Lopes and Polson (2015). Most of this work is related to dynamic linear models and Bayesian forecasting in the sense of Pole, West and Harrison (1994) and West and Harrison (1997). A more recent discussion on these can be found in Smith and Freeman (2011).

There are several reviews of Kalman filtering and the state-space models. An excellent compilation of earlier work is the edited volume by Sorenson (1985) which includes the original papers by Kalman (1960) and Kalman and Bucy (1961) as well as the historical overview paper by Sorenson (1970). A more technical review of the earlier work on linear filtering is given by Kailath (1974). More recent reviews can be found in Chen (2003) which focuses on technical aspects and in Singpurwalla, Polson and Soyer (2018) which provides a more historical perspective on filtering.

In this review, we will focus on issues that are of interest to statisticians and applied probabilists in general and emphasize computational aspects of Kalman filtering as well. In so doing, we first introduce the Gaussian Kalman filter model and present a Bayesian derivation of Kalman filtering. Forecast distributions as well as smoothing results are also discussed. We next focus on extensions of the ordinary Kalman filter model including variance learning in state-space models, Bayesian analysis of state space models using Markov chain Monte Carlo methods, and Kalman filter with Student t-errors. This will be followed by a discussion of recent advances in non-Gaussian state space models and their Bayesian analysis. Use of linear Bayesian methods and Markov chain Monte Carlo techniques are considered. The final section of our overview considers use of particle filtering methods and Bayesian particle learning for static parameters in state space models. The overview ends with some concluding remarks.

Ordinary Kalman Filter Model

We consider the discrete time version of the state space model of Kalman (1960). More specifically, we let Y_t denote the measurement vector (or scalar) at discrete time point t, and let θ_t be the state vector (or scalar) at time t. Then the *observation* (or the *measurement*) *equation* of the state-space model is given by

$$Y_t = F_t\theta_t + v_t, \tag{1}$$

where F_t is a known quantity which is sometimes referred to as the *design matrix* and v_t is the zero-mean observation (or measurement) error with specified variance-covariance matrix V_t. The state vector θ_t follows a Markovian evolution as implied by the *state* (or *system*) *equation*

$$\theta_t = G_t\theta_{t-1} + w_t, \tag{2}$$

where G_t is a known *evolution* (or *transition*) matrix and w_t is the zero-mean state (or system) error with specified variance-covariance matrix W_t. Both v_t and w_t are assumed to be white noise processes and furthermore v_t's and w_t's are assumed to

be uncorrelated with each other. The latter assumption can be relaxed without much difficulty, see Shumway and Stoffer (2011, p. 354).

As pointed out by Meinhold and Singpurwalla (1983), (F_t, G_t, V_t, W_t), the known components of the model, may or may not change over time. We will refer to the model defined by (1) and (2) as the *ordinary Kalman filter* (OKF) model. The Kalman filter estimation involves obtaining the estimator of the conditional mean of the state vector θ_t given observations $D_t = (Y_1, Y_2, \ldots, Y_t)$ available at time t. We denote the conditional mean by m_t, that is, $E[\theta_t|D_t] = m_t$. Without any distributional assumptions on v_t's and w_t's the estimator of m_t can be obtained as the best linear minimum mean square error estimator as done in Kalman (1960). Kalman's derivation of the state estimator of m_t involves using the projection theorem; see Shumway and Stoffer (2011, pp. 528). An alternative approach involving stochastic differential equations was considered by Wegman (1982).

As noted by Harrison and Stevens (1976), the OKF model given by (1) and (2) can be used to describe many well known linear models such as the dynamic regression models of Cooley and Prescott (1973) and nonhomogeneous autoregressive processes of Anderson (1978). Once the observation and state error distributions are specified as Gaussian, it can be shown that the Kalman filter equations can be obtained by using standard Bayesian prior to posterior updates as discussed in Meinhold and Singpurwalla (1983) as well as originally pointed out by Harrison and Stevens (1978).

In what follows, we will look at the OKF and its extensions from a Bayesian perspective and present recent advances in sequential Bayesian analysis. We start our discussion with the Bayesian analysis of the OKF model.

Bayesian Derivation of the Kalman Filter Results

We consider the OKF model with observation and state equations, (1) and (2), respectively. We assume that both the observation and state errors are normally distributed as $v_t \sim \mathcal{N}(0, V_t)$ and $w_t \sim \mathcal{N}(0, W_t)$.

If we assume that at time $t-1$, given D_{t-1}, the state vector θ_{t-1} has a normal distribution with mean m_{t-1} and covariance matrix C_{t-1}, denoted as

$$\theta_{t-1}|D_{t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1}), \tag{3}$$

then via the state equation (2) we can obtain

$$\theta_t|D_{t-1} \sim \mathcal{N}(G_t m_{t-1}, R_t), \tag{4}$$

where $R_t = G_t C_{t-1} G_t' + W_t$. Note that using the observation equation, the conditional distribution of Y_t can be written as

$$Y_t|\theta_t \sim \mathcal{N}(F_t\theta_t, V_t). \tag{5}$$

Prior to observing Y_t, we can obtain the one-step ahead *forecast* (or *predictive*) distribution of Y_t as

$$Y_t|D_{t-1} \sim \mathcal{N}(F_t G_t m_{t-1}, Q_t), \tag{6}$$

where $Q_t = F_t R_t F_t' + V_t$.

Given the above set up, the posterior distribution of state vector θ_t can be obtained by using the fact that the joint distribution of Y_t and θ_t given D_{t-1} is Gaussian as done by Meinhold and Singpurwalla (1983). Using standard multivariate normal distribution results [see for example, Anderson (1984)], the posterior distribution of the state vector is given by a Gaussian form as

$$\theta_t | D_t \sim \mathcal{N}(m_t, C_t), \tag{7}$$

where $D_t = (Y_t, D_{t-1})$,

$$m_t = G_t m_{t-1} + R_t F_t' Q_t^{-1} (Y_t - F_t G_t m_{t-1}), \tag{8}$$

and

$$C_t = R_t - R_t F_t' Q_t^{-1} F_t R_t. \tag{9}$$

Alternatively, the above posterior distribution results can be derived by using the Bayes' law as

$$p(\theta_t | D_t) \propto p(Y_t | \theta_t) p(\theta_t | D_{t-1}) \tag{10}$$

where the right-hand side terms are given by the the likelihood (5) and the prior (4); see West and Harrison (1997, pages 104-105) for details. Initial distribution of the OKF model is specified at time 0 as $\theta_0 | D_0 \sim \mathcal{N}(m_0, C_0)$.

The equations (8) and (9) are the *Kalman filtering* results obtained via the best linear mean square error argument of Kalman (1960). It is important to note that the Kalman filtering results naturally arise as a result of using calculus of probability in the Bayesian paradigm. The posterior distribution of the state vector at time t given by (7) is usually referred to as the *filtering distribution*.

The posterior mean (8) of this distribution can be rewritten as

$$m_t = a_t + R_t F_t' Q_t^{-1} e_t, \tag{11}$$

where $a_t = G_t m_{t-1}$ is the prior mean of state vector θ_t at time $t-1$ and $e_t = (Y_t - F_t a_t)$ is the one-step ahead forecast error after Y_t is observed at time t. The multiplier $R_t F_t' Q_t^{-1}$ of e_t in (11) is referred to as the *Kalman gain* matrix. As pointed out by Meinhold and Singpurwalla (1983), Kalman filtering provides an updating of the state vector θ_t by first forming a prior guess and then revising this guess by adding a correction term based on the one-step ahead forecast error. It is important to note that as a result of the Gaussian assumption on the error terms, the posterior variance C_t in (9) is not affected by data. Thus, if (F_t, G_t, V_t, W_t) are known quantities for any time period, then C_t will be known for all t.

Smoothing and Forecasting in the OKF Model

Smoothing Distributions

The filtering distribution (7) describes uncertainty about state vector θ_t based on observed data at time t. Uncertainty about state vectors for periods $(t-1), (t-2), \ldots, 1$

can also be revised at time t to obtain the *smoothing distributions* $p(\theta_{t-k}|D_t)$ for $k > 1$. This type of retrospective analysis is sometimes referred to as *backward filtering*; see for example Pole et al. (1994).

For $k = 1$ the smoothing distribution $p(\theta_{t-1}|D_t)$ is given by

$$p(\theta_{t-1}|D_t) = \int p(\theta_{t-1}|\theta_t, D_t)p(\theta_t|D_t)\, d\theta_t, \tag{12}$$

where $p(\theta_{t-1}|\theta_t, D_t) = p(\theta_{t-1}|\theta_t, D_{t-1})$ can be obtained via Bayes law as

$$(\theta_{t-1}|\theta_t, D_{t-1}) \sim \mathcal{N}(h_{t-1}, H_{t-1}), \tag{13}$$

where

$$h_{t-1} = m_{t-1} + C_{t-1}G_t'R_t^{-1}(\theta_t - G_t m_{t-1}), \tag{14}$$

$$H_{t-1} = C_{t-1} - C_{t-1}G_t'R_t^{-1}G_t C_{t-1}. \tag{15}$$

Using (12) we can obtain the smoothing distribution of θ_{t-1} as

$$(\theta_{t-1}|D_t) \sim \mathcal{N}[a_t(-1), R_t(-1)], \tag{16}$$

where, following notation of Pole et al. (1994),

$$a_t(-1) = m_{t-1} - B_{t-1}(a_t - m_t),$$

$$R_t(-1) = C_{t-1} - B_{t-1}(R_t - C_t)B_{t-1}',$$

and $B_t = C_t G_{t+1}' R_{t+1}^{-1}$.

As shown by West and Harrison (1997), smoothing distributions for any period $(t - k), k < t$ can be obtained as

$$(\theta_{t-k}|D_t) \sim \mathcal{N}[a_t(-k), R_t(-k)], \tag{17}$$

where the mean vector and covariance matrix are given by

$$a_t(-k) = m_{t-k} - B_{t-k}[a_{t-k+1} - a_t(-k+1)], \tag{18}$$

$$R_t(-k) = C_{t-k} - B_{t-k}[R_{t-k+1} - R_t(-k+1)]B_{t-k}', \tag{19}$$

respectively, and $a_t(0) = m_t$ and $R_t(0) = C_t$.

Forecasting Distributions

At time t, given D_t, k-step ahead forecasts for observables and the state vectors can be obtained if $(F_{t+k}, G_{t+k}, V_{t+k}, W_{t+k})$ are known quantities. For example, at time t the one step ahead forecast distribution of Y_{t+1} can be easily obtained using (6) as

$$Y_{t+1}|D_t \sim \mathcal{N}(F_{t+1}G_{t+1}m_t, Q_{t+1}),$$

where $Q_{t+1} = F_{t+1}R_tF'_{t+1} + V_{t+1}$.

For the general case, we can use the observation and state equations

$$Y_{t+k} = F_{t+k}\theta_{t+k} + v_{t+k},$$
$$\theta_{t+k} = G_{t+k}\theta_{t+k-1} + w_{t+k}$$

for time period $(t + k)$. Given D_t we can show that the forecast distribution of state vector θ_{t+k} is

$$\theta_{t+k}|D_t \sim \mathcal{N}[a_t(k), R_t(k)] \tag{20}$$

where

$$a_t(k) = G_{t+k}a_t(k-1)$$

and

$$R_t(k) = G_{t+1}R_t(k-1)G'_{t+1} + W_{t+k}$$

with $a_t(0) = m_t$ and $R_t(0) = C_t$ as before.

It follows from the observation equation that

$$Y_{t+k}|D_t \sim \mathcal{N}[F_{t+k}a_t(k), Q_t(k)], \tag{21}$$

where $Q_t(k) = F_{t+k}R_t(k)F_{t+1} + V_{t+k}$. We refer to (21) as the k-step ahead forecast distribution of the OKF. Additional details about smoothing and forecasting can be found in West and Harrison (1997).

Extensions of the Ordinary Kalman Filter Model

In this section, we consider extensions of the OKF model with observation and state equations (1) and (2). These include extensions such as variance learning and estimation of other components of the OKF model as well as Student t-distributed observation and state errors.

Bayesian Parameter Learning in the OKF Model

The OKF model and the associated results presented in the last section are based on the assumption that (F_t, G_t, V_t, W_t) are known quantities. Earlier Bayesian work on the estimation dates back to work of Magill (1965) who considered discrete priors for the unknown components of the OKF model. Other work involved use of multiprocess models as in Smith and West (1983) and approximate Bayesian methods as in Singpurwalla and Soyer (1992) who used Laplace type approximations of Lindley (1980) and Tierney and Kadane (1986) in estimation.

Since the design matrix F_t can be specified in many commonly used dynamic models, most of the Bayesian literature focused on estimation of the observation and state covariance matrices V_t, and W_t; see for example, the earlier work by West (1981) who used conjugate priors for observation covariance matrix. West, Harrison and Migon (1985) proposed the concept of "discount matrix" for specification of the state covariance matrix W_t.

Variance Learning in the OKF

Following West and Harrison (1997) we consider the case where the observation covariance matrix is given by V_t^*/ϕ where V_t^* is the specified covariance matrix scaled by unknown precision scalar ϕ. Following their development we also scale the state covariance similary as W_t^*/ϕ where W_t^* is specified and assume that at $t = 0$, uncertainty about ϕ is specified by a gamma prior denoted as $\phi|D_0 \sim \mathcal{G}(r_0/2, d_0/2)$. We also specify the prior for state vector θ_0 at $t = 0$ as

$$(\theta_0|D_0, \phi) \sim \mathcal{N}(m_0, C_0^*/\phi).$$

Assuming that we have $(\theta_{t-1}|D_{t-1}, \phi) \sim \mathcal{N}(m_{t-1}, C_{t-1}^*/\phi)$ at time $t-1$, we can obtain

$$(\theta_t|D_{t-1}, \phi) \sim \mathcal{N}(a_t, R_t^*/\phi),$$

where $R_t^* = G_t C_{t-1}^* G_t' + W_t^*$. The forecast distribution for Y_t can also be obtained as

$$(Y_t|D_{t-1}, \phi) \sim \mathcal{N}(F_t a_t, Q_t^*/\phi),$$

where $Q_t^* = F_t R_t^* F_t' + V_t^*$. As before using multivariate normal distribution theory we can obtain the conditional posterior distribution as

$$(\theta_t|D_t, \phi) \sim \mathcal{N}(m_t, C_t^*/\phi), \tag{22}$$

where

$$m_t = a_t + R_t^* F_t' (Q_t^*)^{-1}(Y_t - F_t a_t), \tag{23}$$

and

$$C_t^* = R_t^* - R_t^* F_t' (Q_t^*)^{-1} F_t R_t^*. \tag{24}$$

The posterior distribution of ϕ can be updated as a gamma distribution,

$$(\phi|D_t) \sim \mathcal{G}(r_t/2, d_t/2), \tag{25}$$

where $r_t = r_{t-1} + 1$ and

$$d_t = d_{t-1} + (Y_t - F_t a_t)'(Q_t^*)^{-1}(Y_t - F_t a_t).$$

The unconditional distribution of θ_t can be obtained as

$$p(\theta_t|D_t) = \int p(\theta_t|D_t, \phi)\, p(\phi|D_t) d\phi$$

giving a Student-t density with r_t degrees of freedom

$$(\theta_t|D_t) \sim \mathcal{T}_{p,r_t}(m_t, C_t^* d_t/r_t), \tag{26}$$

where d_t/r_t can be considered as a posterior point estimate of $1/\phi$. Similarly

$$(\theta_t|D_{t-1}) \sim \mathcal{T}_{p,r_{t-1}}(a_t, R_t^* d_{t-1}/r_{t-1}) \tag{27}$$

and

$$(Y_t|D_{t-1}) \sim \mathcal{T}_{m,r_{t-1}}(F_t a_t, Q_t^* d_{t-1}/r_{t-1}). \tag{28}$$

West and Harrison (1997) propose to specify W_t^* using the concept discounting. More specifically, in $R_t^* = G_t C_{t-1}^* G_t' + W_t^*$, W_t^* represents the additional uncertainty in the state vector going from time $(t-1)$ to t, that is, the addition of w_t to the state equation represents increase in uncertainty about the state vector. This is easier to see when G_t is an identity matrix, that is, when $R_t^* = C_{t-1}^* + W_t^*$. In this case W_t^* represents increase in uncertainty from time $(t-1)$ to t as reflected by the difference of R_t^* and C_{t-1}^*. If we can specify this increase in uncertainty then we can specify the state noise matrix. More specifically, if $R_t^* = C_{t-1}^*/\delta$,where $0 < \delta \leq 1$, then this implies that

$$W_t^* = C_{t-1}^*(\frac{1-\delta}{\delta}) \tag{29}$$

and the increase in uncertainty is $\lambda = (\frac{1-\delta}{\delta})\,\%$.

Bayesian Learning via MCMC Methods

Bayesian methods presented in the last section are quite limited since analytical tractability of the posterior and forecast distributions cannot be maintained except in few cases. An alternative approach is use of Monte Carlo methods and more specifically MCMC methods that allow us to draw samples from the posterior distributions when analytical forms are not available; see for example, Gelfand and Smith (1990).

In what follows, we consider the OKF model with static observation and state covariance matrices, that is, we assume $V_t = V$ and $W_t = W$ for all t in (1) and (2). For both covariance matrices V and W we will consider independent inverse Wishart priors. More specifically, for the $m \times m$ precision matrix $\Phi_v = V^{-1}$ for observation errors, we assume a Wishart distribution

$$p(\Phi_v) \propto |\Phi_v|^{(r_v-m-1)/2} exp\Big[-\frac{1}{2} tr\Big\{\Sigma_v \Phi_v\Big\}\Big] \tag{30}$$

where scale matrix Σ_v and degrees of freedom $r_v > m$ are known quantities and $tr\{\}$ denotes the trace of the matrix. We denote the Wishart distribution (30) as $(\Phi_v|\Sigma_v, r_v) \sim Wish(\Sigma_v, r_v)$. Similary, for the $p \times p$ precision matrix $\Phi_w = W^{-1}$ for state errors, we assume a Wishart distribution $(\Phi_w|\Sigma_w, r_w) \sim Wish(\Sigma_w, r_w)$ with scale Σ_w and degrees of freedom $r_w > p$.

Given data D_T at time T, we are interested in the joint posterior distribution of all unknown quantities, that is, $p(\Theta_T, \Phi_v, \Phi_w|D_T)$, where $\Theta_T = (\theta_1, \theta_2, \ldots, \theta_T)$. Since the distribution is not analytically available, we can use a Gibbs sampler to draw samples from the joint posterior. The Gibbs sampler requires drawing samples iteratively from full conditional posterior distributions of the unknown quantities. As discussed by Reis, Salazar and Gamerman (2006), there are alternative strategies in drawing from the full conditionals. One can draw θ_t's individually by using full conditionals

$$p(\theta_t|D_T, \Theta_T^{(-t)}, \Phi_v, \Phi_w) \propto p(Y_t|\theta_t, \Phi_v) p(\theta_t|\theta_{t-1}, \Phi_w) p(\theta_{t+1}|\theta_t, \Phi_w)$$

where $\Theta_T^{(-t)} = \{\theta_s | s \neq t\}$. This can be achieved quite easily since $p(\theta_t | D_T, \Theta_T^{(-t)}, \Phi_v, \Phi_w)$ is a normal distribution and the full conditionals $p(\Phi_v | D_T, \Phi_w, \Theta_T)$ and $p(\Phi_w | D_T, \Phi_v, \Theta_T)$ are available as Wishart distributions. As noted by Reis et al. (2006) due to the correlation between θ_t's this strategy my turn out inefficient. The alternative is to draw the elements of Θ_T jointly. We can use the *forward filtering backward sampling* (FFBS) algorithm of Fruhwirth-Schnatter (1994) and Carter and Kohn (1994) to draw from $p(\Theta_T | D_T, \Phi_v, \Phi_w)$.

Using the Markov structure of the OKF model we can write $p(\Theta_T | D_T, \Phi_v, \Phi_w)$ as

$$p(\theta_T | D_T, \Phi_v, \Phi_w)\, p(\theta_{T-1} | \theta_T, D_{T-1}, \Phi_v, \Phi_w) \cdots\cdots p(\theta_1 | \theta_2, D_1, \Phi_v, \Phi_w), \quad (31)$$

where the first term $p(\theta_T | D_T, \Phi_v, \Phi_w)$ is available from standard OKF updating. We can start the sampling from θ_T and then sequentially sample $\theta_{T-1}, \ldots, \theta_1$ using densities $p(\theta_{t-1} | \theta_t, D_{t-1}, \Phi_v, \Phi_w)$ for $t = T-1, \ldots, 2$. The required distributions are given by the smoothing results given by (13), (14), and (15) which are all conditional on precision matrices Φ_v and Φ_w.

The full conditional $p(\Phi_v | D_T, \Phi_w, \Theta_T)$ can be obtained as proportional to

$$|\Phi_v|^{(r_v+T-m-1)/2} exp\Big[-\frac{1}{2} tr\Big\{\Big(\Sigma_v + \sum_{t=1}^{T}(Y_t - F_t\theta_t)(Y_t - F_t\theta_t)'\Big)\Phi_v\Big\}\Big] \quad (32)$$

which is again a Wishart density with degrees of freedom, $(r_v + T)$, and scale matrix

$$\Big(\Sigma_v + \sum_{t=1}^{T}(Y_t - F_t\theta_t)(Y_t - F_t\theta_t)'\Big).$$

Similarly, the full conditional $p(\Phi_w | D_T, \Phi_v, \Theta_T)$ is given by

$$|\Phi_w|^{(r_w+T-p-1)/2} exp\Big[-\frac{1}{2} tr\Big\{\Big(\Sigma_w + \sum_{t=1}^{T}(\theta_t - G_t\theta_{t-1})(\theta_t - G_t\theta_{t-1})'\Big)\Phi_w\Big\}\Big] \quad (33)$$

which is again a Wishart density with degrees of freedom $(r_w + T)$, and scale matrix $\Big(\Sigma_w + \sum_{t=1}^{T}(\theta_t - G_t\theta_{t-1})(\theta_t - G_t\theta_{t-1})'\Big)$.

Since, one can directly draw samples from the full conditionals (31), (32) and (33), implementation of the Gibbs sampler is quite straightforward.

Alternative MCMC approaches include blocking strategies of Gamerman and Moreira (2002) and reparameterization idea of Gamerman (1998) which was implemented for Bayesian analysis of dynamic generalized linear models. An empirical comparison of different MCMC strategies for dynamic linear models and discussion of computational issues can be found in Reis et al. (2006).

It is important to note that the Gibbs sampler will provide us draws from the smoothing distributions of $\theta_t, t = 1, \ldots, T-1$, at time T. To be able to obtain samples from the filtering distributions θ_t's, one needs to rerun the Gibbs sampler at each time period t. This undesirable feature of the Gibbs sampler and other MCMC methods motivated the development of sequential Monte Carlo methods such as particle filtering; see Gordon, Salmond, and Smith (1993), Liu and Chen (1998) and Pitt and Shephard (1999).

Kalman Filter with Student t-Errors

As previously mentioned, the OKF model with Gaussian errors provides the Gaussian filtering distribution (7) for the state vector where the mean m_t depends on the observed data D_t, but the variance C_t does not. Furthermore, as pointed out by Meinhold and Singpurwalla (1989), the posterior mean (8) is an "unbounded function" of the forecast error e_t implying a *nonrobust* model. These characteristics of the Gaussian OKF model have resulted in efforts for robustifying the model by using alternative error distributions as in West (1981) and Meinhold and Singpurwalla (1989).

In the OKF model of (1) and (2), we assume that Y_t is a $m \times 1$ observation vector and the state vector θ_t has dimension $p \times 1$. Thus, F_t and G_t are $m \times p$ and $p \times p$ matrices, respectively. An alternative to the Gaussian OKF model is a Kalman filter model with both the observation and state error vectors having Student t-distributions. More specifically, we assume that the observation error vector v_t has an m-dimensional t distribution with degrees of freedom r, mean vector 0 and $m \times m$ *scale-matrix* V_t denoted as

$$v_t \sim \mathcal{T}_{m,r}(0, V_t). \tag{34}$$

Similarly, the state error vector w_t is assumed to have a p-dimensional Student t-distribution with r degrees of freedom, mean vector 0 and $p \times p$ scale-matrix W_t denoted as

$$w_t \sim \mathcal{T}_{p,r}(0, W_t). \tag{35}$$

Following Meinhold and Singpurwalla (1989), at time $t-1$, given D_{t-1}, we assume a p-dimensional Student t-distribution with degrees of freedom $r + (t-1)m$, mean m_{t-1} and scale matrix C_{t-1} denoted as

$$\theta_{t-1}|D_{t-1} \sim \mathcal{T}_{p,r+(t-1)m}(m_{t-1}, C_{t-1}).$$

Using properties of multivariate Student t-distribution [see for example, Anderson (1984)] and the state equation (2), it can be shown that the prior of state vector θ_t at time $t-1$ is given by the p-dimensional Student t distribution

$$\theta_t|D_{t-1} \sim \mathcal{T}_{p,r+(t-1)m}(G_t m_{t-1}, c(D_{t-1})R_t), \tag{36}$$

where $c(D_{t-1})$ is a function of the observed data at time $t-1$ as will be discussed in the sequel. It was shown in Meinhold and Singpurwalla (1989), that the posterior distribution of θ_t at time t is obtained as a Student t with $r + tm$ degrees of freedom as

$$\theta_t|D_t \sim \mathcal{T}_{p,r+tm}(m_t, C_t), \tag{37}$$

where

$$m_t = G_t m_{t-1} + R_t F_t' Q_t^{-1}(Y_t - F_t G_t m_{t-1}),$$

and

$$C_t = c(D_t)(R_t - R_t F_t' Q_t^{-1} F_t R_t). \tag{38}$$

Note that the posterior mean is identical to the one from the Gaussian OKF model. However, the scale matrix of θ_t is different and it is a function of data through the $c(D_t)$ term given by

$$c(D_t) = \frac{r + \sum_{s=1}^{t}(Y_s - f_s)' Q_s^{-1}(Y_s - f_s)}{(r + tm)} \tag{39}$$

where $f_t = F_t G_t m_{t-1}$. Thus, the posterior covariance matrix of θ_t which is proportional to scale matrix C_t is increasing quadratically in forecast error e_t. This behavior is referred to as *outlier confusion* by Meinhold and Singpurwalla (1989).

Non-Gaussian State-space Models

Extension of state space models to non-Gaussian observation models has been given considerable attention in the literature during the last three decades. A generalization of univariate Gaussian OKF model was introduced in West, Harrison and Migon (1985) by assuming that the observations were generated from an exponential family of distributions. The authors used standard link functions of general linear models (GLMs) where the regression coefficients were assumed to follow a state equation as in the OKF model. The resulting models were referred to as dynamic general linear models (DGLMs). State vector updates in the DGLM set up were obtained only for the first and second moments using linear Bayesian methods of Hartigan (1969). Miller and Smith (1986) introduced non-Gaussian state space models using results from Bather (1965). Harvey and Fernandes (1989) considered related dynamic models for time-series of counts. Most of these work prior to 1990s emphasized analytical tractability for posterior distributions. Following the development of MCMC methods, a Bayesian analysis of DGLMs was presented in Gamerman (1998). In what follows, we present two examples of non-Gaussian state space models.

Poisson State Space Models

In a recent paper, Soyer, Aktekin and Kim (2015) considered time-series of count data. Such type of data can arise in numerous fields such as engineering, business, economics or epidemiology. For instance, observations under study can be the number of arrivals to a call center during every five minutes [Aktekin and Soyer (2011)], number of shopping trips of households in a week [Aktekin, Polson and Soyer (2018)], number of mortgages defaulted from a particular pool in a given month [see Aktekin, Soyer and Xu (2013)], number of accidents in a given time interval [Serhiyenko et al. (2014)] or the number of deaths from a specific disease in a given year [Schmidt and Pereira (2011)].

Let N_t be the number of occurrences of an event during time interval t, where $t = 1, 2, \ldots$, and let θ_t be the corresponding latent Poisson rate during the same time. Following Soyer et al. (2015) the number of occurrences during period t is described by

the Poisson model

$$p(N_t|\theta_t) = \frac{\theta_t^{N_t} e^{-\theta_t}}{N_t!}. \tag{40}$$

We refer to (40) as the *observation equation* of the Poisson state space model and assume that N_ts are conditionally independent given θ_ts. Time evolution of θ_ts is assumed to follow a Markovian model given by

$$\theta_t = \frac{\theta_{t-1}}{\gamma}\epsilon_t, \tag{41}$$

where $(\epsilon_t|D_{t-1}) \sim Beta[\gamma\alpha_{t-1}, (1-\gamma)\alpha_{t-1}]$ with $\alpha_{t-1} > 0, 0 < \gamma < 1$, and $D_{t-1} = \{N_1, \cdots, N_{t-1}\}$. We refer to (41) as the *state equation* of the Poisson state space model where γ acts like a discount term. It follows from (41) that $\theta_t < \frac{\theta_{t-1}}{\gamma}$ and the conditional distributions of consecutive rates are all scaled Beta densities,

$$p(\theta_t|\theta_{t-1}, D_{t-1}) = \frac{\Gamma(\alpha_{t-1})}{\Gamma(\gamma\alpha_{t-1})\Gamma(\{1-\gamma\}\alpha_{t-1})}\Big(\frac{\gamma}{\theta_{t-1}}\Big)^{\alpha_{t-1}-1}\theta_t^{\gamma\alpha_{t-1}-1}\Big(\frac{\theta_{t-1}}{\gamma}-\theta_t\Big)^{(1-\gamma)\alpha_{t-1}-1} \tag{42}$$

denoted as $(\theta_t|\theta_{t-1}, D_{t-1}) \sim Beta[\gamma\alpha_{t-1}, (1-\gamma)\alpha_{t-1}; (0, \frac{\theta_{t-1}}{\gamma})]$. The state equation (41) also implies that $E(\theta_t|\theta_{t-1}, D_{t-1}) = \theta_{t-1}$, in other words a random walk type of evolution in the expectation of the Poisson rates.

Based on the measurement and state equations, if we assume that at time 0, $(\theta_0|D_0)$ is a gamma distribution as

$$(\theta_0|D_0) \sim \mathcal{G}(\alpha_0, \beta_0), \tag{43}$$

then it is possible to develop an analytically tractable sequential updating for the model. Given the inductive hypothesis

$$(\theta_{t-1}|D_{t-1}) \sim \mathcal{G}(\alpha_{t-1}, \beta_{t-1}), \tag{44}$$

using (42) and (44), we can obtain the distribution of θ_t given D_{t-1} as

$$(\theta_t|D_{t-1}) \sim \mathcal{G}(\gamma\alpha_{t-1}, \gamma\beta_{t-1}). \tag{45}$$

It follows from the above that $E(\theta_t|D_{t-1}) = E(\theta_{t-1}|D_{t-1})$, and $V(\theta_t|D_{t-1}) = V(\theta_{t-1}|D_{t-1})/\gamma$ implying an increase in variance.

We can obtain the filtering distribution of $(\theta_t|D_t)$ using the Bayes' Rule as

$$p(\theta_t|D_t) \propto p(N_t|\theta_t)p(\theta_t|D_{t-1}), \tag{46}$$

implying that

$$p(\theta_t|D_t) \propto \theta_t^{\gamma\alpha_{t-1}+N_t-1}e^{-(\gamma\beta_{t-1}+1)\theta_t}.$$

Thus, the filtering distribution of the Poisson rate at time t is a gamma density

$$(\theta_t|D_t) \sim \mathcal{G}(\alpha_t, \beta_t), \tag{47}$$

where the recursive updating of model parameters is given by $\alpha_t = \gamma\alpha_{t-1} + N_t$ and $\beta_t = \gamma\beta_{t-1} + 1$.

Furthermore, the one-step ahead forecast distribution of counts at time t given D_{t-1} can be obtained as

$$p(N_t|D_{t-1}) = \int_0^\infty p(N_t|\theta_t)p(\theta_t|D_{t-1})d\theta_t, \tag{48}$$

where $(N_t|\theta_t) \sim Poisson(\theta_t)$ and $(\theta_t|D_{t-1}) \sim \mathcal{G}(\gamma\alpha_{t-1}, \gamma\beta_{t-1})$. Therefore,

$$p(N_t|D_{t-1}) = \binom{\gamma\alpha_{t-1} + N_t - 1}{N_t} \Big(1 - \frac{1}{\gamma\beta_{t-1}+1}\Big)^{\gamma\alpha_{t-1}} \Big(\frac{1}{\gamma\beta_{t-1}+1}\Big)^{N_t}. \tag{49}$$

which is a negative binomial model denoted as

$$(N_t|D_{t-1}) \sim Negbin(r_t, p_t), \tag{50}$$

where $r_t = \gamma\alpha_{t-1}$ and $p_t = \frac{\gamma\beta_{t-1}}{\gamma\beta_{t-1}+1}$. Given (50), one can carry out one step ahead predictions and forecast interval calculations in a straightforward manner.

Although the k-step ahead predictive density is not analytically available, the k-step ahead predictive means can be easily obtained. Using a standard conditional expectation argument one can obtain $E(N_{t+k}|D_t)$ as follows

$$E(N_{t+k}|D_t) = E_{\theta_{t+k}}\{E(N_{t+k}|\theta_{t+k}, D_t)\} = E(\theta_{t+k}|D_t). \tag{51}$$

Furthermore, using the state equation we have

$$E(\theta_{t+k}|D_t) = E(\theta_t|D_t) \prod_{n=t+1}^{t+k} \frac{E(\epsilon_n|D_t)}{\gamma} = E(\theta_t|D_t) = \frac{\alpha_t}{\beta_t}, \tag{52}$$

where $E(\epsilon_n|D_t) = \gamma$ for any n. Therefore, combining (51) and (52), we can write

$$E(N_{t+k}|D_t) = E(\theta_{t+k}|) = \frac{\alpha_t}{\beta_t}. \tag{53}$$

Due to the random walk type of structure introduced in (42), the above result simply indicates that k-step ahead forecasts given that we have observed counts up to time t are equal to α_t/β_t. In recognition of this feature, Aktekin et al. (2013) proposed an extension of the Poisson state space models by incorporating covariates into the state equation. The Poisson state-space model is member of the class of exact marginal likelihood models considered by Gamerman et al. (2013). A multivariate version of the Poisson-state space models are proposed in Aktekin, Polson and Soyer (2018) where authors develop particle filtering methods for Bayesian analysis.

Dynamic Probit Models

Bayesian state-space models for categorical time series have been considered in Carlin and Polson (1992), Cargnoni, Mueller and West (1997) and Gamerman (1998). As noted by Soyer and Sung (2013), Bayesian analysis of some of the models are based on

Metropolis-Hastings algorithm [see Chib and Greenberg (1995)] which requires specification of proposal densities in high dimensions. Soyer and Sung (2013) proposed a probit-type state-space models and developed Bayesian analysis using an exact Gibbs sampler. They also considered generalizations such as Student t link functions and presented marginal likelihood computations along the lines of Chib (1995).

Consider a binary time-series Y_{it} for individual i such that

$$Pr\{Y_{it} = 1|\pi_{it}\} = \pi_{it}$$

with probit link

$$\pi_{it} = \Phi(F_{it}\theta_t) \tag{54}$$

where F_{it} is a covariate vector for individual i and θ_t is a p dimensional vector of regression parameters. We define the state equation for θ_t as

$$\theta_t = G_t\theta_{t-1} + w_t,$$

where w_t's are uncorrelated multivariate normal error vectors with mean 0 and covariance matrix W_θ and G_t is known. For example, in Soyer and Sung (2013) G_t is assumed to be an identity matrix.

Following Albert and Chib (1993), the probit model can be represented by using independent latent variables Z_{it} such that

$$Y_{it} = \begin{cases} 1 & \text{if } Z_{it} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

If Z_{it}'s are normally distributed with mean $F_{it}\theta_t$ and variance 1, that is, $Z_{it} \sim \mathcal{N}(F_{it}\theta_t, 1)$, then we have the probit model $\pi_{it} = \Phi(F_{it}\theta_t)$.

Given data $D_T = \{Y_{it};\, t = 1, \ldots, T\}$, Bayesian analysis can be developed by using a Gibbs sampler with full conditionals $p(\Theta_T|D_T, Z_i^T)$ and $p(Z_i^T|D_T, \Theta_T)$ where $\Theta_T = (\theta_1, \theta_2, \ldots, \theta_T)$ and $Z_i^T = (Z_{i1}, Z_{i2}, \ldots, Z_{iT})$. For drawing samples from $p(Z_i^T|D_T, \Theta_T)$, we note that Z_{it}'s are independent random variables and use

$$(Z_{it}|\theta_t, Y_{it} = 1) \sim \mathcal{N}(F_{it}\theta_t, 1)\, I(Z_{it} > 0)$$

$$(Z_{it}|\theta_t, Y_{it} = 0) \sim \mathcal{N}(F_{it}\theta_t, 1)\, I(Z_{it} < 0).$$

We can directly draw from $p(\Theta_T|D_T, Z_i^T) = p(\Theta_T|Z_i^T)$ using the FFBS algorithm of Fruhwirth-Schnatter (1994) as in (31). We define $Z_i^t = (Z_i^{t-1}, Z_{it})$, $t = 1, \ldots, T$ and write $p(\Theta_T|Z_i^T)$ as

$$p(\theta_T|Z_i^T)\, p(\theta_{T-1}|\theta_T, Z_i^{T-1}) \cdots\cdots p(\theta_1|\theta_2, Z_i^1), \tag{55}$$

where the first term $p(\theta_T|Z_i^T)$ is the filtering distribution at time T. Each element of (55) follows a normal distribution as

$$(\theta_{t-1}|\theta_t, Z_i^{t-1}) \sim \mathcal{N}(h_{t-1}, H_{t-1})$$

where h_{t-1} and H_{t-1} can be obtained from (14) and (15) with

$$m_t = G_t m_{t-1} + R_t F_{it}'(1 + F_{it} R_t F_{it}')^{-1} e_t$$

where $e_t = Z_{it} - F_{it} G_t m_{t-1}$ is a scalar, $R_t = G_t C_{t-1} G_t' + W_\theta$, and

$$C_t = R_t - R_t F_{it}'(1 + F_{it} R_t F_{it}') F_{it} R_t.$$

Note that the above results are all conditional on W_θ. Using a Wishart prior for W_θ^{-1}

$$W_\theta^{-1} | \Sigma, r \sim Wish(\Sigma, r),$$

where $r > p$, the full conditional of W_θ can be obtained as a Wishart density with degrees of freedom, $(r + T)$, and scale matrix $\left(\Sigma + \sum_{t=1}^{T} (\theta_t - \theta_{t-1})(\theta_t - \theta_{t-1})'\right)$.

Thus, the proposed approach provides an exact Gibbs sampler for Bayesian inference. The algorithm can be easily generalized to M individuals for T time periods as discussed in Soyer and Sung (2013).

Particle Filtering and Bayesian Particle Learning

The particle filter (PF) is commonly used by engineers and statisticians in diverse areas such as signal processing, internet traffic, online marketing, etc. The PF may be easily understood when cast as a problem of Bayesian learning and employing some well-known updating rules for mixture models. The PF can also be used to perform static parameter learning. The original PF was introduced by Gordon, Salmond, and Smith (1993) for Bayesian state estimation. An excellent review of particle filtering is given in Doucet and Johansen (2011) and Bayesian methods for particle learning for static parameters are discussed by Carvalho et al. (2010).

Consider the Gaussian OKF model defined by (1) and (2) and assume that we have the filtering distribution $p(\theta_t | D_t)$ as given by (7). What is meant by a *particle distribution approximation* are the draws $\theta_t^{(i)}$, $i = 1, \ldots, N$, such that

$$p^N(\theta_t | D_t) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\theta_t^{(i)}} \tag{56}$$

where δ denotes a Dirac measure.

Particle methods will rely on the law of large numbers as $N \to \infty$ rather than ergodic distribution to guarantee convergence to an equilibrium distribution as required by the Markov chain Monte Carlo (MCMC) methods. A PF can be viewed as the evolution of N interacting stochastic processes whose marginal distributions are designed to match the required sequence of filtering distributions $p(\theta_t | D_t)$. One simply simulates the N particles of the stochastic process using a resampling and propogation step. As

$N \to \infty$, the law of large numbers will guarantee weak convergence to the appropriate filtering distribution.

Since the MCMC methods provide us with samples from the smoothing distribution $p(\theta_1, \theta_2, \ldots, \theta_T|D_T)$, obtaining samples sequentially from the filtering distributions $p(\theta_t|D_t), t = 1, 2, \ldots, T$, requires rerunning the MCMC each time. The PF provides us with an efficient way of drawing from the filtering distributions sequentially.

Recursive Estimation Procedure

As pointed out by Johannes and Polson (2009), updating of the state vector θ_t can be viewed in two ways by factoring the joint distribution as

$$p(Y_{t+1}, \theta_{t+1}|\theta_t) = p(Y_{t+1}|\theta_{t+1})p(\theta_{t+1}|\theta_t) \tag{57}$$

or

$$p(Y_{t+1}, \theta_{t+1}|\theta_t) = p(Y_{t+1}|\theta_t)\, p(\theta_{t+1}|\theta_t, Y_{t+1}). \tag{58}$$

The former factorization given by (57) is the more traditional Kalman filtering approach. In PF language we *propogate* with $p(\theta_{t+1}|\theta_t^{(i)})$ and *resample* with $p(Y_{t+1}|\theta_{t+1}^{(i)})$. More specifically, in obtaining the new filtering distribution $p(\theta_{t+1}|D_{t+1})$ we use the prediction for θ_{t+1}, that is,

$$p(\theta_{t+1}|D_t) = \int p(\theta_{t+1}|\theta_t)p(\theta_t|D_t)d\theta_t$$

for propogation. Given particles $\theta_t^{(i)}$ we draw $\theta_{t+1}^{(i)}$, $i = 1, \ldots, N$ and then use the Bayes' rule

$$p(\theta_{t+1}|D_{t+1}) \propto p(Y_{t+1}|\theta_{t+1})p(\theta_{t+1}|D_t)$$

for resampling $\theta_{t+1}^{k(i)}$ by drawing the index $k(i)$ from a multinomial distribution with weights

$$w_{t+1}^{(i)} = \frac{p(Y_{t+1}|\theta_{t+1}^{(i)})}{\sum_{j=1}^{N} p(Y_{t+1}|\theta_{t+1}^{(j)})}. \tag{59}$$

We then set $\theta_{t+1}^{(i)} = \theta_{t+1}^{k(i)}$. This is known as the *sample importance resampling* (SIR) approach; see for example, Liu and Chen (1998). Note that in the resampling step the likelihood $P(y_{t+1}|\theta_{t+1})$ weighs the prior $p(\theta_{t+1}|D_t)$ for updating. This is sometimes referred to as *weighted bootstrap*; see Smith and Gelfand (1992). As noted by Johannes and Polson (2009) the SIR approach in PF can lead to degeneracies in the filtering distribution since particles may collapse on few values. One way to alleviate this problem is to implement a *resample first-propagate next* strategy as in the auxiliary particle filter method of Pitt and Shephard (1999).

The latter factorization given by (58) leads to a mixture approximation to the next posterior $p(\theta_{t+1}|D_{t+1})$ given a particle representation

$$p(\theta_t|y_{1:t}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\theta_t^{(i)}}.$$

In this case, we first resample θ_t's using the smoothing distibution

$$p(\theta_t|D_{t+1}) \propto p(Y_{t+1}|\theta_t)p(\theta_t|D_t)$$

obtained by Bayes' rule. Thus, we draw $\theta_t^{k(i)}$ by drawing the index $k(i)$ from a multinomial distribution with weights

$$w_t^{(i)} = \frac{p(Y_{t+1}|\theta_t^{(i)})}{\sum_{j=1}^N p(Y_{t+1}|\theta_t^{(j)})}. \tag{60}$$

We then set $\theta_t^{(i)} = \theta_t^{k(i)}$ and propogate using

$$p(\theta_{t+1}|Y_{t+1}) = \int p(\theta_{t+1}|\theta_t, Y_{t+1})p(\theta_t|D_{t+1})d\theta_t.$$

More specifically, the mixture approximation to the posterior $p(\theta_{t+1}|D_{t+1})$ is given by

$$p(\theta_{t+1}|D_{t+1}) = \sum_{i=1}^N \frac{p(Y_{t+1}|\theta_t^{(i)})}{\sum_{j=1}^N p(Y_{t+1}|\theta_t^{(j)})} p(\theta_{t+1}|\theta_t^{(i)}, Y_{t+1}). \tag{61}$$

This is known as the *exact particle filtering* approach; see Johannes and Polson (2009) and it does not suffer from the degeneracy problem of SIR when Y_{t+1} is an outlier.

Example: PF for the Steady Model

The special case of the OKF model where Y_t and θ_t are scalars, and $F_t = G_t = 1$, for all t, is referred to as the *steady model* or *first-order polynomial model*; see West and Harrison (1997). We also specify $V_t = \sigma_v^2$, and $W_t = \sigma_w^2$ for all t in the model. In other words we can write the observation and the state equations of the model as

$$Y_t = \theta_t + v_t, \tag{62}$$

where $v_t \sim \mathcal{N}(0, \sigma_v^2)$, and

$$\theta_t = \theta_{t-1} + w_t, \tag{63}$$

where $w_t \sim \mathcal{N}(0, \sigma_w^2)$.

We first consider the case where σ_v^2 and σ_w^2 are known quantities. As previously discussed, the filtering, smoothing and forecast distributions can all be obtained analytically in this model. We next illustrate the SIR and exact PF approaches presented above.

SIR PF:

Propogation Step: Given particles $\theta_t^{(i)}$ we draw $\theta_{t+1}^{(i)}$, $i = 1, \ldots, N$ from

$$\theta_{t+1}|\theta_t^{(i)} \sim \mathcal{N}(\theta^{(i)}, \sigma_w^2).$$

Resampling Step: Given $\theta_{t+1}^{(i)}$, we resample $\theta_{t+1}^{k(i)}$ using weights proportional to normal density $\mathcal{N}(\theta_{t+1}^{(i)}, \sigma_v^2)$, that is,

$$w_{t+1}^{(i)} \propto p(Y_{t+1}|\theta_{t+1}^{(i)}) = \mathcal{N}(\theta_{t+1}^{(i)}, \sigma_v^2)$$

for $i = 1, \ldots, N$.

Exact PF:

Resampling Step: Given particles $\theta_t^{(i)}$, $i = 1, \ldots, N$, we first resample θ_t's using weights proportional to normal density $N(\theta^{(i)}, \sigma_v^2 + \sigma_w^2)$, that is,

$$w_t^{(i)} \propto p(Y_{t+1}|\theta_t^{(i)}) = \mathcal{N}(\theta^{(i)}, \sigma_v^2 + \sigma_w^2)$$

for $i = 1, \ldots, N$.

Propogation Step: Given resampled $\theta_t^{(i)}$'s, we draw $\theta_{t+1}^{(i)}$, $i = 1, \ldots, N$ from $\theta_{t+1}|\theta_t^{(i)}, Y_{t+1} \sim \mathcal{N}(m_{t+1}^{(i)}, 1/\phi)$, where $\phi = 1/\sigma_v^2 + 1/\sigma_w^2$ and

$$m_{t+1}^{(i)} = \frac{1/\sigma_v^2}{\phi} Y_{t+1} + \frac{1/\sigma_w^2}{\phi} \theta_t^{(i)}.$$

Parameter Learning and Sufficient Statistics

As previously discussed, in many filtering problems we also need to learn about other unknown parameters to obtain the filtering distributions. We denote these unknown parameters by λ. It is possible to develop a particle learning algorithm for these parameters in PF. This can be efficiently done if there exists a conditionally conjugate probability model for λ given the state variables; see for example Lopes, Polson and Carvalho (2012). In such cases, there exists a sufficient statistic s_t for λ at time t defined by the deterministic recursion $S(\cdot)$ as $s_{t+1} = S(s_t, \theta_{t+1}, Y_{t+1})$. Then, the PF algorithms can be easily extended to incorporate learning about λ by tracking particles of s_t at each period; see Lopes et al. (2011) and Lopes and Polson (2015) for some recent work on particle learning.

In the exact PF method we can write

$$p(Y_{t+1}, \theta_{t+1}|\theta_t, \lambda = p(Y_{t+1}|\theta_t, \lambda)\, p(\theta_{t+1}|\theta_t, \lambda, Y_{t+1}) \tag{64}$$

with the filtering distribution given by

$$p(\theta_{t+1}|D_{t+1}) \propto \int p(Y_{t+1}|\theta_t, \lambda)\, p(\theta_{t+1}|\theta_t, \lambda, Y_{t+1}) p(\theta_t, \lambda|D_t) d\lambda d\theta_t. \tag{65}$$

Given particles $(\theta_t^{(i)}, \lambda^{(i)}, s_t^{(i)})$, $i = 1, \ldots, N$, $(\theta_t^{k(i)}, \lambda^{k(i)}, s_t^{k(i)})$ can be resampled with weights proportional to $p(Y_{t+1}|\theta_t^{k(i)}, \lambda^{k(i)})$ and $s_t^{k(i)} = S(s_t^{(i)}, \theta_t^{k(i)}, Y_{t+1})$. This is followed by propogation to $p(\theta_{t+1}|D_{t+1})$ by drawing $\theta_{t+1}^{(i)}$, from $p(\theta_{t+1}|\theta_t^{k(i)}, \lambda^{k(i)}, Y_{t+1})$, for $i = 1, \ldots, N$. Next the sufficient statistic can be updated as

$$s_{t+1} = S(s_t^{k(i)}, \theta_{t+1}^{(i)}, Y_{t+1}), \tag{66}$$

for $i = 1, \ldots, N$, which represents a deterministic propogation. Finally, parameter learning is completed by drawing $\lambda^{(i)}$ using $p(\lambda|s_{t+1}^{(i)})$ for $i = 1, \ldots, N$.

We show below an implementation of the particle learning algorithm by using an extension of the steady model where the state variance σ_w^2 is treated as unknown.

Particle Learning in the Steady Model

Consider an extension of the steady model given by

$$Y_t = \theta_t + v_t,\ v_t \sim \mathcal{N}(0, \sigma_v^2)$$

and

$$\theta_t = \theta_{t-1} + w_t, w_t \sim \mathcal{N}(0, 1/\lambda),$$

where $\theta_0 \sim \mathcal{N}(m_0, C_0)$, $\lambda \sim \mathcal{G}(r_0/2, s_0/2)$, and θ_0 is independent of λ at $t = 0$. Note that s_0 may also be random with density $p(s_0)$, but without loss of generality, here, we assume that s_0 is fixed.

In this example, the filtering and smoothing distributions of θ_t's as well as the posterior distribution of λ cannot be obtained analytically, but a Gibbs sampler based on the FFBS algorithm of Fruhwirth-Schnatter (1994) can be developed as previously discussed. This approach is not efficient for obtaining the filtering distributions since it requires rerunning the Gibbs sampler for each time period t.

The exact PF for this model is as follows:

Resampling Step: Given particles $(\theta_t^{(i)}, \lambda^{(i)}, s_t^{(i)})$, $i = 1, \ldots, N$ we first resample $(\theta_t^{k(i)}, \lambda^{k(i)})$ using weights proportional to normal density $\mathcal{N}(\theta^{(i)}, \sigma_v^2 + 1/\lambda^{(i)})$, and obtain resampled sufficient statistic values $s_t^{k(i)} = S(s_t^{(i)}, \theta_t^{k(i)}, Y_{t+1})$ for $i = 1, \ldots, N$.

Propogation Step: Given resampled $\theta_t^{(i)}$'s, we draw $\theta_{t+1}^{(i)}$, $i = 1, \ldots, N$ from

$$\theta_{t+1}|\theta_t^{k(i)}, \lambda^{k(i)}, Y_{t+1} \sim N(m_{t+1}^{(i)}, \sigma_v^2 + 1/\lambda^{k(i)}),$$

where

$$m_{t+1}^{(i)} = \frac{1/\sigma_v^2}{\lambda^{k(i)} + 1/\sigma_v^2} y_{t+1} + \frac{\lambda^{k(i)}}{\lambda^{k(i)} + 1/\sigma_v^2} \theta_t^{k(i)}.$$

We update the sufficient statistic via

$$s_{t+1}^{(i)} = S(s_t^{k(i)}, \theta_{t+1}^{(i)}, Y_{t+1}) = s_t^{k(i)} + (\theta_{t+1}^{(i)} - \theta_t^{k(i)})^2 \tag{67}$$

for $i = 1, \ldots, N$.

Updating Step: Given the sufficient statistic $s_{t+1}^{(i)}$, $i = 1, \ldots, N$ we draw $\lambda^{(i)}$, $i = 1, \ldots, N$ using $p(\lambda|s_{t+1}^{(i)})$ which is $\mathcal{G}(r_{t+1}/2, s_{t+1}^{(i)}/2)$ where $r_{t+1} = r_t + 1$.

It is easy to see that the approach can be performed more efficiently if we can reduce the dimension by integrating out λ. We note that in our example, initially, $p(Y_1|\theta_0, s_0)$ is a Student t-density. Thus, at $t = 0$, given s_0 and N particles $\theta_0^{(i)}$ $i = 1, \ldots, N$, from the t-density, the resampling weights are

$$w_0^{(i)} = \frac{p(Y_1|(\theta_0, s_0)^{(i)})}{\sum_{j=1}^N p(Y_1|(\theta_0, s_0)^{(j)})}, \tag{68}$$

where $(\theta_0, s_0)^{(i)} = (\theta_0^{(i)}, s_0)$. Given the weights $w_0^{(i)}$, we can resample new particles, $(\theta_0, s_0)^{k(i)}$, $i = 1, \ldots, N$ and track these at each stage in our approach.

For example, we can propogate to stage 1 by drawing $\theta_1^{(i)}$ from $p(\theta_1|(\theta_0, s_0)^{k(i)}, Y_1)$, which is Student t, and update the sufficient statistic as $s_1^{(i)} = s_0^{k(i)} + (\theta_1^{(i)} - \theta_0^{k(i)})^2$. In summary, the approach requires only keeping track of $(\theta_t, s_t)^{(i)}$, for $i = 1, \ldots, N$, for any time period t. Since the particles of the sufficient statistic s_t are available we can always draw samples from the posterior distribution of λ.

Furthermore, we note that a SIR based particle learning approach can also be developed. This starts with the propogation step where we draw θ_{t+1}, using the Student t-distribution $p(\theta_{t+1}|\theta_t^{(i)}, s_t^{(i)})$. We follow with the resampling step where we draw particles $\theta_{t+1}^{k(i)}$'s using weights proportional to $p(Y_{t+1}|\theta_{t+1}^{(i)}, s_t^{(i)})$ and update the sufficient statistic via $s_{t+1}^{(i)} = S(s_t^{(i)}, \theta_{t+1}^{k(i)}, Y_{t+1})$, for $i = 1, \ldots, N$. As before, given $s_{t+1}^{(i)}$ we can draw particles $\lambda^{(i)}$, using $p(\lambda|s_{t+1}^{(i)})$ to evaluate the posterior distribution of λ.

Concluding Remarks

In this paper we have given an overview of Kalman filtering and state-space models and discussed their role in Bayesian time-series analysis. In so doing, we have presented some key results in ordinary Gaussian Kalman filter model, considered some important extensions and discussed parameter learning and related computational issues. We have also covered some recent advances in non-Gaussian state space models and their Bayesian analysis using MCMC and sequential MCMC methods such as particle filtering.

Other recent developments in dynamic Bayesian modeling and computational aspects of sequential learning can be found in Prado and West (2010). The boom of big data

brings in new challenges both in terms of modeling high dimensional data and efficient computing for sequential Bayesian learning. Scott et al. (2016) discuss some advances in "Consensus Monte Carlo" algorithms for big data.

Acknowledgements

The author would like to thank Professor Nicholas G. Polson for many discussions on particle filtering methods and sequential Monte Carlo techniques.

References

Aktekin, T. & Soyer, R. (2011). Call center modeling: A Bayesian state space approach. *Naval Research Logistics*, **58**, 28-42.

Aktekin, T., Soyer, R. & Xu, F. (2013). Assessment of mortgage default risk via Bayesian state-space models. *Annals of Applied Statistics*, **7**, 1450-1473.

Aktekin, T., Polson, N. G. & Soyer, R. (2018). Sequential Bayesian analysis of multivariate Poisson count data. *Bayesian Analysis*, **13**, 385-409.

Albert, J. H. & Chib S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**, 669-679.

Anderson, T. W. (1978). Repeated measurements on autoregressive processes. *Journal of the American Statistical Association*, **73**, 371-378.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. 2nd Edition, Wiley.

Artemiev, V. M., Naumov, A. O. & Tillack, G. R. (2001). Recursive tomographic image reconstruction using a Kalman filter in the time domain, *Journal of Physics D: Applied Physics*, **34**, 2073-2083.

Auger, F., Hilairet, M., Guerrero, J. M., Monmasson, E., Orlowska-Kowalska, T. & Katsura, S. (2013). Industrial applications of the Kalman filter: a review. *IEEE Transactions on Industrial Electronics*, **60**, 5458-5471.

Babbs, S. H. & Nowman, K. B. (1999). Kalman filtering of generalized Vasicek term structure models. *Journal of Financial and Quantitative Analysis*, **34**, 115-130.

Bather, J.A. (1965). Invariant conditional distributions. *Ann. Math. Statist.*, **36**, 829-46.

Bertino, L., Evensen, G. & Wackernage, H. (2003). Sequential data assimilation techniques in oceanography. International Statistical Review, **71**, 223-241.

Cargnoni C., Mueller P., & West M. (1997). Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models. *Journal of the American Statistical Association*, **92**, 640-647.

Carlin, B. P. & Polson, N. G. (1992). Monte Carlo Bayesian methods for discrete regression models and categorical time series. *Bayesian Statistics 4*, (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Eds.), 577-586, Oxford University Press.

Carter, C. K. & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, **81**, 541-553.

Carvalho, C., Johannes, M. S., Lopes, H. F. & Polson, N. G. (2010). Particle learning and smoothing, *Statistical Science*, **25**, 88-106.

Chen, Z. (2003). Bayesian filtering: From Kalman filters to particle filters, and beyond. *Statistics*, **182**, 1-69.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *J. of the Am. Statist. Assoc.*, **90**, 1313-1321.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *American Statistician*, **49**, 327-335.

Cooley, T. F. & Prescott, E. C. (1973). An adaptive regression model. *Int. Econ. Review*, 14, 364-371.

Doucet, A. and Johansen, A. M. (2011). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering*, Eds.,D. Crisan and B. Rozovsky, 656-704, Oxford University Press.

Evensen, G. (1992). Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *Jour. of Geophys. Res.*, **97**, l7905-17924.

Fruhwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models, *Journal of Time Series Analysis*, **15**,183-202.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalized linear models. *Biometrika*, **85**, 215-227.

Gamerman, D. & Moreira, A. R. B. (2002). Bayesian analysis of econometric time series models using hybrid integration rules. *Communications in Statistics*, **31**, 4972.

Gamerman, D., Dos-Santos, T. R. & Franco, G. C. (2013). A non-Gaussian family of state-space models with exact marginal likelihood. *Journal of Time Series Analysis*, **34**,625645.

Gauthier, P. Courtier, P. & Moll, P. (1993). Assimilation of simulated wind lidar data with a Kalman filter. *Mon. Weather Rev.*, **121**, 1803-1820.

Gelb, A. (1986). Dual contributions of optimal estimation theory in aerospace applications. *IEEE Control Systems Magazine*, February issue, 3-13.

Gelfand, A. E. & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities.*J. Amer. Statist. Assoc.*, **85**, 398-409.

Gordon, N., Salmond, D. & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc.* F, **140**, 107-113.

Harrison, P. J. & Stevens, C. F. (1976). Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society, Ser. B*, **38**, 205-247.

Hartigan, J. A. (1969). Linear Bayesian methods. *Journal of the Royal Statistical Society, Ser. B*, **31**, 446-454.

Harvey, A. C. & Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business and Economic Statistics*, **7**, 407417.

Johannes, M. S. & Polson, N. G. (2009). Particle filtering, *Handbook of Financial Time Series*. (Eds. T. G. Anderson, et al.), 1015-1029.

Johannes, M. S., Korteweg, A. & Polson, N. G. (2014). Sequential learning, predictability, and optimal portfolio returns. *Journal of Finance*, **69**, 611-644.

Kailath, T. (1974). A view of three decades of linear filtering theory. *IEEE Transactions on Information Theory*, **20**, 146-181.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, **82**, 35-45.

Kalman, R. E. & Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, **83**, 95-108.

Lindley, D. V. (1980). Approximate Bayesian methods. *Trabajos Estadistica*, **31**, 223-237.

Liu, J. S. & Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. J. *Amer. Statist. Assoc.*, **93**, 1032-1044.

Lopes, H. F., Carvalho, C. M., Johannes, M. S. & Polson, N. G. (2011). Particle learning for sequential Bayesian estimation (with discussion). *Bayesian Statistics 9*. (Eds. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West), 317-360.

Lopes, H. F. & Tsay, R. S. (2011). Particle filters and Bayesian inference in financial econometrics. *Journal of Forecasting*, **30**, 168-209.

Lopes, H. F., Polson, N. G. & Carvalho, C. M. (2012). Bayesian statistics with a smile: a resampling-sampling perspective. *Brazilian Journal of Probability and Statistics*, **26**, 358-371.

Lopes, H. F. & Polson, N. G. (2015). Particle learning for fat-tailed distributions. *Econometric Reviews*, **35**, 1666-1691.

Magill, D. T. (1965). Optimal adaptive estimation of sampled stochastic processes. *IEEE Transactions on Automatic Control*, **10**, 434-439.

Meinhold, R. J. & Singpurwalla, N. D. (1983). Understanding the Kalman filter. *American Statistician*, **37**, 123-127.

Meinhold, R. J. & Singpurwalla. N. D. (1989). Robustification of Kalman filter models. *Journal of the American Statistical Association*, **84**, 479-486.

Miller, J. E. & Smith, R. L. (1986). A non-Gaussian state space model and application to prediction of records. *Journal of the Royal Statistical Society, Series B*, **48**, 79-88.

Pitt, M. & Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Amer. Statist. Assoc.*, **94**, 590-599.

Pole, A., West, M. & Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis*, CRC Press.

Prado R., West M. & Krystal A. (2001). Multi-Channel EEG analyses via dynamic regression models with time-varying lag/lead structure. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **50**, 95-109.

Prado R. & West M. (2010). *Time Series: Modelling, Computation and Inference*. Chapman & Hall/CRC.

Reis, E. A., Salazar, E. & Gamerman, D. (2006). Comparison of sampling schemes for dynamic linear models. *International Statistical Review*, **74**, 203-214.

Schmidt, A. M. & Pereira, J. B. M. (2011). Modelling time series of counts in epidemiology. *International Statistical Review*, **79**, 4869.

Schneider, W. (1988). Analytical uses of Kalman filtering in econometrics A survey. *Statistical Papers*, **29**, 333.

Scott, S. L., Blocker, A. W., Bonassi, F. V., Chipman, H. A., George, E. I. & McCulloch, R. E. (2016). Bayes and big data: the consensus Monte Carlo algorithm. *International Journal of Management Science and Engineering Management*, forthcoming.

Serhiyenko, V. , Ivan, J. H., Ravishanker, N. & Islam, S. (2014). Dynamic compositional modeling of pedestrian crash counts on urban roads in Connecticut. *Accident Analysis & Prevention*, **64**, 7885.

Shumway, R. H. & Stoffer, D. S. (2011). *Time Series Analysis and Its Applications*. 3rd Edition, Springer-Verlag.

Singpurwalla, N. D. & Soyer, R. (1992). Nonhomogeneous autoregressive processes for tracking reliability growth, and their Bayesian analysis. *Journal of the Royal Statistical Society, Series B*, **54**, 145-156.

Singpurwalla, N. D., Polson, N. G. & Soyer, R. (2018). From least squares to signal processing and particle filtering. *Technometrics*, forthcoming.

Smith, A. F. M. & Gelfand, A. (1992). Bayesian statistics without tears: A sampling-resampling perspective, *American Statistician*, **46**, 84-88.

Smith, A. F. M. & West, M. (1983). Monitoring renal transplants: An application of the multiprocess Kalman filter. *Biometrics*, **39**, 867-878.

Smith, J. Q. & Freeman, G. (2011). Distributional Kalman filters for Bayesian forecasting and closed form recurrences. *Journal of Forecasting*, **30**, 210-224.

Sorenson, H. W. (1970). Least-squares estimation: from Gauss to Kalman. *IEEE Spectrum*, **7**, 63-68.

Sorenson, H. W. (1985). *Kalman Filtering: Theory and Application*. IEEE press.

Soyer, R. & Sung, M. (2013). Bayesian Dynamic probit models for analysis of longitudinal data. *Computational Statistics and Data Analysis*, **68**, 388-398.

Soyer, R., Aktekin, T. & Kim, B. (2015). Bayesian modeling of time series of counts with business applications. *Handbook of Discrete-Valued Time Series*, (R. Davis, S. Holan, R. Lund and N. Ravishanker, Eds), 245-266.

Tierney, L. & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, **81**, 82-86.

Wegman, E. J. (1982). Kalman filtering. in *Encyclopedia of Statistics*, eds. N. Johnson & S. Kotz, Wiley.

West, M. (1981). Robust sequential approximate Bayesian estimation. *Journal of the Royal Statistical Society, Ser. B*, **43**, 157-166.

West, M.., Harrison, P. J. & Migon, H. S. (1985). Dynamic generalized linear models, and Bayesian forecasting. *Journal of the American Statistical Association*, **80**, 73-97.

West, M. & Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Models*. 2nd Edition, Springer-Verlag.