

Sequential Monte Carlo Sampling for State Space Models

Mario V. Wüthrich

Abstract The aim of these notes is to revisit sequential Monte Carlo (SMC) sampling. SMC sampling is a powerful simulation tool for solving non-linear and/or non-Gaussian state space models. We illustrate this with several examples.

1 Introduction

In these notes we revisit sequential Monte Carlo (SMC) sampling for (non-linear and non-Gaussian) state space models in discrete time. SMC sampling and non-linear particle filters were introduced in the 1990s by Gordon et al. [7] and Del Moral [3]. Meanwhile there is a vast literature on SMC sampling and there are excellent (overview) contributions such as Del Moral et al. [4, 5], Johansen and Evers [9], Doucet and Johansen [6] and Creal [2]. In fact, we have learned SMC methods from these references, in particular, from Doucet and Johansen [6]. The reason for writing these notes is that we had to prepare for a tutorial lecture on SMC sampling. For this purpose it is always advantageous to develop and implement own examples to understand and back-test the algorithms. These own examples and their implementation are probably our only real contributions here, but nevertheless they might be helpful to a wider audience who wants to get familiar with SMC sampling.

Organization. We start by giving three explicit examples of state space models in Sects. 2 and 3, the sampling algorithms are only presented later in Sect. 4. In Sect. 2 we give two examples of *linear* state space models: (1) a *Gaussian linear* state space model and (2) a *non-Gaussian linear* state space model. These models can be solved with the Kalman filter technique that is exact in the former case and that is an approximation in the latter case. In Sect. 3 we consider (3) a *non-Gaussian and non-linear* state space model. Moreover, we present the corresponding densities of all three models. Section 4 is devoted to the sampling algorithms. We start with importance sampling, then discuss sequential importance sampling (SIS) and the last

M.V. Wüthrich (✉)

RiskLab, Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland
e-mail: mario.wuethrich@math.ethz.ch

algorithm presented is a SMC sampling one. These algorithms are useful to solve the three models introduced above. This is demonstrated in the examples Sect. 5. In this section we also provide another practical example that corresponds to a stochastic volatility model (that is inspired by the Heston [8] model) and we describe backward smoothing of the resulting estimates.

2 Linear State Space Models

In this section we present two explicit examples of *linear* state space models: (1) a *Gaussian* one in Sect. 2.1 and (2) a *non-Gaussian* one in Sect. 2.2. Moreover, we present the Kalman filter technique that solves these models.

2.1 Gaussian Linear State Space Models and the Kalman Filter

In many situations Gaussian linear state space models are studied. These are either exact or used as an approximation to the full problem. Therefore, we start by describing Gaussian linear state space models. Such models typically consist of two processes: (i) a *transition system*, which describes the latent risk factor process, and (ii) a *measurement system*, that describes the observable process.

The following structure gives a (one-dimensional) Gaussian linear state space model:

- (i) The *transition system* is described by a process $(\Theta_t)_{t \in \mathbb{N}_0}$ with $\Theta_0 = \theta_0$ and for $t \geq 1$

$$\Theta_t = a + b\Theta_{t-1} + \tau\eta_t, \quad (1)$$

for $a, b \in \mathbb{R}$, $\tau > 0$ and $(\eta_t)_{t \geq 1}$ being i.i.d. standard Gaussian distributed.

- (ii) The *measurement system* is described by a process $(X_t)_{t \in \mathbb{N}}$ with for $t \geq 1$

$$X_t = \Theta_t + \sigma\varepsilon_t, \quad (2)$$

for $\sigma > 0$ and $(\varepsilon_t)_{t \geq 1}$ being i.i.d. standard Gaussian distributed and being independent of the process $(\eta_t)_{t \geq 1}$.

For given parameters $\theta_0, a, b, \tau, \sigma$ one aims at inferring the (unobservable, latent) vector $\Theta_{1:t} = (\Theta_1, \dots, \Theta_t)$ from given observations $X_{1:t} = (X_1, \dots, X_t)$. The technique usually used is the so-called Kalman filter [10] which can be interpreted as an exact linear credibility estimator, see Sect. 9.5 in Bühlmann and Gisler [1]. The Kalman filter provides the following algorithm:

Step 1 (anchoring). Initialize

$$\theta_{1|0} = \mathbb{E}[\Theta_1 | X_{1:0}] = a + b\theta_0 \quad \text{and} \quad \tau_{1|0}^2 = \text{Var}(\Theta_1 | X_{1:0}) = \tau^2,$$

where the empty vector $X_{1:0}$ is assumed to generate the trivial σ -field, leading to $\mathbb{E}[\Theta_1 | X_{1:0}] = \mathbb{E}[\Theta_1]$ and $\text{Var}(\Theta_1 | X_{1:0}) = \text{Var}(\Theta_1)$.

Step 2 (forecasting the measurement system). At time $t \geq 1$ we obtain forecast

$$x_{t|t-1} = \mathbb{E}[X_t | X_{1:t-1}] = \mathbb{E}[\Theta_t | X_{1:t-1}] = \theta_{t|t-1},$$

and prediction variance

$$s_{t|t-1}^2 = \text{Var}(X_t | X_{1:t-1}) = \text{Var}(\Theta_t | X_{1:t-1}) + \sigma^2 = \tau_{t|t-1}^2 + \sigma^2.$$

One period later, for given observation X_t , we receive (observable) prediction error

$$\zeta_t = X_t - \mathbb{E}[X_t | X_{1:t-1}] = X_t - x_{t|t-1}.$$

Step 3 (Bayesian inference of the transition system). This prediction error ζ_t is used to update the transition system at time t . Using inference we obtain Bayesian estimate

$$\theta_{t|t} = \mathbb{E}[\Theta_t | X_{1:t}] = \mathbb{E}[\Theta_t | X_{1:t-1}] + K_t \zeta_t = \theta_{t|t-1} + K_t \zeta_t,$$

with the so-called Kalman gain matrix (credibility weight)

$$K_t = \text{Var}(\Theta_t | X_{1:t-1}) \text{Var}(X_t | X_{1:t-1})^{-1} = \tau_{t|t-1}^2 / s_{t|t-1}^2,$$

and the variance $\tau_{t|t-1}^2$ is updated by

$$\tau_{t|t}^2 = \text{Var}(\Theta_t | X_{1:t}) = (1 - K_t) \text{Var}(\Theta_t | X_{1:t-1}) = (1 - K_t) \tau_{t|t-1}^2.$$

Step 4 (forecasting the transition system). For the latent risk factor we obtain forecast

$$\theta_{t+1|t} = \mathbb{E}[\Theta_{t+1} | X_{1:t}] = a + b\mathbb{E}[\Theta_t | X_{1:t}] = a + b\theta_{t|t},$$

and prediction variance

$$\tau_{t+1|t}^2 = \text{Var}(\Theta_{t+1} | X_{1:t}) = b^2 \text{Var}(\Theta_t | X_{1:t}) + \tau^2 = b^2 \tau_{t|t}^2 + \tau^2.$$

Remark We emphasize the distinguished meanings of $\theta_{t|t-1}$, $x_{t|t-1}$ and $\theta_{t|t}$. The former two $\theta_{t|t-1}$ and $x_{t|t-1}$ are *predictors* to forecast Θ_t and X_t based on the

information $X_{1:t-1}$; the latter $\theta_{t|t}$ is an *estimator* for the latent Θ_t based on the information $X_{1:t}$. These predictors and estimators are exact and optimal (in a Bayesian way) for Gaussian innovations in linear state space models. In fact, we obtain the following exact credibility formula in Step 3 (weighted average between observation X_t and (prior) forecast $\theta_{t|t-1}$):

$$\theta_{t|t} = \mathbb{E}[\Theta_t | X_{1:t}] = \theta_{t|t-1} + K_t \zeta_t = K_t X_t + (1 - K_t) \theta_{t|t-1},$$

with credibility weight (Kalman gain matrix)

$$K_t = \frac{\tau_{t|t-1}^2}{\tau_{t|t-1}^2 + \sigma^2} = \frac{1}{1 + \sigma^2 / \tau_{t|t-1}^2} \in (0, 1).$$

This credibility estimator $\theta_{t|t}$ is exact in the Gaussian linear state space model and it can be used as best linear approximation (for the quadratic loss function) for other state space models, see Chap. 9 in Bühlmann and Gisler [1].

The main question we would like to address here is: how can we optimally infer Θ_t in non-Gaussian and non-linear state space models? Before addressing this question we briefly consider a non-Gaussian linear state space model.

2.2 Non-Gaussian Linear State Space Models

We present for illustration one example of a non-Gaussian linear state space model. Therefore, we replace (1)–(2) by the following structure:

- (i) The *transition system* is described by a process $(\Theta_t)_{t \in \mathbb{N}_0}$ with $\Theta_0 = \theta_0$ and for $t \geq 1$

$$\Theta_t = b\Theta_{t-1} + \eta_t, \quad (3)$$

for $b \in \mathbb{R}$ and $(\eta_t)_{t \geq 1}$ being i.i.d. gamma distributed with $\mathbb{E}[\eta_t] = a$ and $\text{Var}(\eta_t) = \tau^2$.

- (ii) The *measurement system* is described by a process $(X_t)_{t \in \mathbb{N}}$ with for $t \geq 1$

$$X_t = \Theta_t + \sigma \varepsilon_t, \quad (4)$$

for $\sigma > 0$ and $(\varepsilon_t)_{t \geq 1}$ being i.i.d. standard Gaussian distributed and being independent of process $(\eta_t)_{t \geq 1}$.

Observe that the measurement systems (2) and (4) are identical, conditionally given $(\Theta_t)_{t \in \mathbb{N}_0}$. The transition systems (1) and (3) differ, but not in the first two moments, that is,

$$\mathbb{E}[\Theta_t | \Theta_{t-1}] = a + b\Theta_{t-1} \quad \text{and} \quad \text{Var}(\Theta_t | \Theta_{t-1}) = \tau^2. \quad (5)$$

This implies that linear credibility filtering provides the same Kalman filter results in both models, see Chap. 9 in Bühlmann and Gisler [1].

3 Non-Gaussian and Non-linear State Space Models

3.1 Illustrative Example

We consider the following non-Gaussian and non-linear state space model:

- (i) The *transition system* is described by a process $(\Theta_t)_{t \in \mathbb{N}_0}$ with $\Theta_0 = \theta_0 = 1$ and for $t \geq 1$

$$\Theta_t = b\Theta_{t-1} + \sqrt{\Theta_{t-1}}\eta_t, \quad (6)$$

for $b > 0$ and $(\eta_t)_{t \geq 1}$ being i.i.d. gamma distributed with $\mathbb{E}[\eta_t] = a$ and $\text{Var}(\eta_t) = \tau^2$.

- (ii) The *measurement system* is described by a process $(X_t)_{t \in \mathbb{N}}$ with for $t \geq 1$

$$X_t = \Theta_t + \sigma\varepsilon_t, \quad (7)$$

for $\sigma > 0$ and $(\varepsilon_t)_{t \geq 1}$ being i.i.d. standard Gaussian distributed and being independent of process $(\eta_t)_{t \geq 1}$.

Note that in all three examples we consider the same measurement system (2), (4) and (7), but the three transition systems (1), (3) and (6) differ. Below, we will choose $\theta_0 = 1$ and $b = 1 - a \in (0, 1)$, see (21). These parameter choices imply for the first two *linear* state space models (1) and (3), see (5),

$$\mathbb{E}[\Theta_t] = 1 \quad \text{and} \quad \text{Var}(\Theta_t) = \tau^2 \frac{1 - b^{2t}}{1 - b^2} \leq \tau^2 \frac{1}{1 - b^2}.$$

The *non-linear* model (6) is mean reverting in the following sense, assume $b = 1 - a \in (0, 1)$,

$$\mathbb{E}[\Theta_t | \Theta_{t-1}] = (1 - a)\Theta_{t-1} + \sqrt{\Theta_{t-1}}a \quad \begin{cases} < \Theta_{t-1} & \text{if } \Theta_{t-1} > 1, \\ = \Theta_{t-1} & \text{if } \Theta_{t-1} = 1, \\ > \Theta_{t-1} & \text{if } \Theta_{t-1} < 1. \end{cases}$$

For $t = 1$ and $\Theta_0 = \theta_0 = 1$ we obtain

$$\mathbb{E}[\Theta_1 | \Theta_0] = (1 - a)\Theta_0 + a\sqrt{\Theta_0} = 1.$$

By induction, using the Markov property of $(\Theta_t)_{t \in \mathbb{N}_0}$, the tower property of conditional expectations and applying Jensen's inequality, we obtain for $t \geq 2$

$$\begin{aligned}\mathbb{E} [\Theta_t | \Theta_0] &= \mathbb{E} [\mathbb{E} [\Theta_t | \Theta_{t-1}] | \Theta_0] = \mathbb{E} \left[(1-a) \Theta_{t-1} + a\sqrt{\Theta_{t-1}} \middle| \Theta_0 \right] \\ &< (1-a) \mathbb{E} [\Theta_{t-1} | \Theta_0] + a\mathbb{E} [\Theta_{t-1} | \Theta_0]^{1/2} \leq 1.\end{aligned}$$

3.2 Bayesian Inference of the Transition System

The state space models introduced above can be interpreted as Bayesian models. This is highlighted next. In general, we will use letter π to denote (conditional) densities that belong to the transition system and letter f for (conditional) densities that belong to the measurement system.

We start with the non-Gaussian and non-linear state space model (6)–(7). Choose parameters $\gamma, c > 0$ such that $\mathbb{E}[\eta_t] = \gamma/c = a$ and $\text{Var}(\eta_t) = \gamma/c^2 = \tau^2$. Given the transition system $\Theta_{1:t}$, the observations $X_{1:t}$ have the following joint (product) density

$$\begin{aligned}X_{1:t} | \{\Theta_{1:t} = \theta_{1:t}\} &\sim f(x_{1:t} | \theta_{1:t}) = \prod_{s=1}^t f(x_s | \theta_s) \\ &= \prod_{s=1}^t \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x_s - \theta_s)^2}{2\sigma^2} \right\}.\end{aligned}$$

The joint (prior) density of the vector $\Theta_{1:t}$ is given by (for later purposes we indicate θ_0 in the notation)

$$\begin{aligned}\Theta_{1:t} | \theta_0 &\sim \pi(\theta_{1:t} | \theta_0) = \prod_{s=1}^t \pi(\theta_s | \theta_{s-1}) \\ &= \prod_{s=1}^t \frac{1}{\sqrt{\theta_{s-1}}} \frac{c^\gamma}{\Gamma(\gamma)} \left(\frac{\theta_s - b\theta_{s-1}}{\sqrt{\theta_{s-1}}} \right)^{\gamma-1} \exp \left\{ -c \left(\frac{\theta_s - b\theta_{s-1}}{\sqrt{\theta_{s-1}}} \right) \right\} \\ &\quad \times 1_{\{\theta_s \geq b\theta_{s-1}\}}.\end{aligned}$$

This implies that the posterior density of $\Theta_{1:t}$, conditionally given $(X_{1:t}, \theta_0)$, satisfies

$$\begin{aligned}\pi(\theta_{1:t} | X_{1:t}, \theta_0) &\propto f(X_{1:t} | \theta_{1:t}) \pi(\theta_{1:t} | \theta_0) = \prod_{s=1}^t f(X_s | \theta_s) \pi(\theta_s | \theta_{s-1}) \\ &\propto \prod_{s=1}^t \frac{1}{\sqrt{\theta_{s-1}}} \left(\frac{\theta_s - b\theta_{s-1}}{\sqrt{\theta_{s-1}}} \right)^{\gamma-1} \exp \left\{ -\frac{(X_s - \theta_s)^2}{2\sigma^2} - c \left(\frac{\theta_s - b\theta_{s-1}}{\sqrt{\theta_{s-1}}} \right) \right\} \\ &\quad \times 1_{\{\theta_s \geq b\theta_{s-1}\}}.\end{aligned} \tag{8}$$

Thus, we can determine the posterior density $\pi(\theta_{1:t}|X_{1:t}, \theta_0)$ in model (6)–(7) up to the normalizing constant, but we do not immediately recognize that it comes from a well-understood (multivariate) distribution function. Therefore, we determine the posterior distribution numerically.

For completeness we also provide the posterior distribution in the linear state space models of Sect. 2. In model (1)–(2) it is derived as follows. Given the vector $\Theta_{1:t}$ the observations $X_{1:t}$, have the following joint (product) density

$$X_{1:t} | \{\Theta_{1:t} = \theta_{1:t}\} \sim f(x_{1:t} | \theta_{1:t}) = \prod_{s=1}^t \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x_s - \theta_s)^2}{2\sigma^2} \right\}.$$

The joint (prior) density of the vector $\Theta_{1:t}$ is given by

$$\Theta_{1:t} | \theta_0 \sim \pi(\theta_{1:t} | \theta_0) = \prod_{s=1}^t \frac{1}{\sqrt{2\pi}\tau} \exp \left\{ -\frac{(\theta_s - a - b\theta_{s-1})^2}{2\tau^2} \right\}.$$

This implies that the posterior density of $\Theta_{1:t}$, conditionally given $(X_{1:t}, \theta_0)$, satisfies

$$\begin{aligned} \pi(\theta_{1:t} | X_{1:t}, \theta_0) &\propto f(X_{1:t} | \theta_{1:t}) \pi(\theta_{1:t} | \theta_0) = \prod_{s=1}^t f(X_s | \theta_s) \pi(\theta_s | \theta_{s-1}) \\ &\propto \exp \left\{ -\sum_{s=1}^t \frac{(X_s - \theta_s)^2}{2\sigma^2} + \frac{(\theta_s - a - b\theta_{s-1})^2}{2\tau^2} \right\}. \end{aligned} \quad (9)$$

From this we see that the posterior of $\Theta_{1:t}$, given $(X_{1:t}, \theta_0)$, is a multivariate Gaussian distribution (with known parameters) and any problem can directly be solved from this knowledge. Observe that this slightly differs from the Kalman filter of Sect. 2. In the Kalman filter we were estimating (the next) Θ_t based on observations $X_{1:t}$, which provided Bayesian estimate $\theta_{t|t}$. The full posterior $\pi(\theta_{1:t} | X_{1:t}, \theta_0)$ now also allows us for backward smoothing, that is, we can study the Bayesian estimator of Θ_s for any earlier time point $s = 1, \dots, t$ given by

$$\theta_{s|t} = \mathbb{E}[\Theta_s | X_{1:t}].$$

The posterior distribution in model (3)–(4) is given by

$$\begin{aligned} \pi(\theta_{1:t} | X_{1:t}, \theta_0) &\propto f(X_{1:t} | \theta_{1:t}) \pi(\theta_{1:t} | \theta_0) \\ &\propto \prod_{s=1}^t (\theta_s - b\theta_{s-1})^{\gamma-1} \exp \left\{ -\frac{(X_s - \theta_s)^2}{2\sigma^2} - c(\theta_s - b\theta_{s-1}) \right\} 1_{\{\theta_s \geq b\theta_{s-1}\}}. \end{aligned} \quad (10)$$

The aim in the next section is to describe algorithms that allow us to simulate directly from the posterior densities (8)–(10), respectively.

4 Sequential Monte Carlo Sampling

In this section we follow Sect. 3 of Doucet and Johansen [6]. Throughout we assume that all terms considered are well-defined, for instance, concerning integrability. The aim is to sample from the (posterior) densities $\pi(\theta_{1:t}|X_{1:t}, \theta_0)$ that are known up to the normalizing constants, that is,

$$\pi(\theta_{1:t}|X_{1:t}, \theta_0) \propto f(X_{1:t}|\theta_{1:t}) \pi(\theta_{1:t}|\theta_0) =: \gamma_t(\theta_{1:t}),$$

where the last identity defines γ_t which describes the functional form of the (posterior) density up to a normalizing constant Z_t that is given by

$$Z_t = \int \gamma_t(\theta_{1:t}) d\theta_{1:t}.$$

In particular, this implies that we have (posterior) density

$$\pi(\theta_{1:t}|X_{1:t}, \theta_0) = Z_t^{-1} \gamma_t(\theta_{1:t}).$$

We remark that the following algorithms are quite general. The careful reader will notice that they require much less structure than the three models introduced above possess.

4.1 Importance Sampling

A general way to obtain samples from a density $\pi(\theta_{1:t}|X_{1:t}, \theta_0)$ that is only known up to a normalizing constant is to apply importance sampling. Assume h is a well-behaved measurable function and we aim at calculating the (posterior) mean

$$\mathbb{E}[h(\Theta_{1:t})|X_{1:t}] = \int h(\theta_{1:t}) \pi(\theta_{1:t}|X_{1:t}, \theta_0) d\theta_{1:t} = \frac{\int h(\theta_{1:t}) \gamma_t(\theta_{1:t}) d\theta_{1:t}}{\int \gamma_t(\theta_{1:t}) d\theta_{1:t}}.$$

For importance sampling we choose an importance density q_t that has at least the same support as γ_t and from which we can (easily) sample. The latter is important because otherwise the problem will not be solved. Using this importance density q_t and assuming that $\tilde{\Theta}_{1:t} \sim q_t$ we can rewrite the above (posterior) mean as follows

$$\mathbb{E}[h(\Theta_{1:t})|X_{1:t}] = \frac{\int h(\theta_{1:t}) \frac{\gamma_t(\theta_{1:t})}{q_t(\theta_{1:t})} q_t(\theta_{1:t}) d\theta_{1:t}}{\int \frac{\gamma_t(\theta_{1:t})}{q_t(\theta_{1:t})} q_t(\theta_{1:t}) d\theta_{1:t}} = \frac{\mathbb{E}[h(\tilde{\Theta}_{1:t}) w_t(\tilde{\Theta}_{1:t}) | X_{1:t}]}{\mathbb{E}[w_t(\tilde{\Theta}_{1:t}) | X_{1:t}]}, \quad (11)$$

where we have defined the unnormalized importance weights

$$w_t(\theta_{1:t}) = \frac{\gamma_t(\theta_{1:t})}{q_t(\theta_{1:t})}.$$

Remarks

- Identity (11) says that we can sample from a tractable density $\tilde{\Theta}_{1:t} \sim q_t$. To obtain samples from γ_t we simply need to re-weight these samples using the importance weights w_t . Note that this requires that $\text{supp}(\gamma_t) \subseteq \text{supp}(q_t)$.
- Efficient algorithms to evaluate (11) numerically for arbitrary functions h will consider importance densities q_t such that w_t has a small variance. This leads to a fast convergence in the normalizing constant Z_t (which is the denominator of (11)). Ideally, one also wants to have fast convergence in the numerator of (11). However, since this is not possible for arbitrary function h , one only focuses on the importance weights for the normalizing constant.
- Note that we condition on $\sigma\{X_{1:t}, \theta_0\}$ in (11) because (strictly speaking) the importance weights w_t depend on these observations (if we are aiming at calculating the posterior distributions). Moreover, also the choice of $q_t(\cdot) = q_t(\cdot|X_{1:t}, \theta_0)$ may depend on these observations.

We now evaluate (11) empirically. Choose $I \in \mathbb{N}$ independent samples $\tilde{\Theta}_{1:t}^{(i)} \sim q_t$, $i = 1, \dots, I$. We obtain empirical estimate

$$\begin{aligned} \widehat{\mathbb{E}}^{(I)} [h(\Theta_{1:t}) | X_{1:t}] &= \frac{\frac{1}{I} \sum_{i=1}^I h(\tilde{\Theta}_{1:t}^{(i)}) w_t(\tilde{\Theta}_{1:t}^{(i)})}{\frac{1}{I} \sum_{i=1}^I w_t(\tilde{\Theta}_{1:t}^{(i)})} \\ &= \sum_{i=1}^I h(\tilde{\Theta}_{1:t}^{(i)}) \frac{w_t(\tilde{\Theta}_{1:t}^{(i)})}{\sum_{j=1}^I w_t(\tilde{\Theta}_{1:t}^{(j)})}. \end{aligned} \quad (12)$$

This importance sampling algorithm proposes to evaluate the function h under the empirical (discrete) distribution

$$\widehat{\pi}^{(I)}(\theta_{1:t} | X_{1:t}, \theta_0) = \sum_{i=1}^I W_t(\tilde{\Theta}_{1:t}^{(i)}) \delta_{\tilde{\Theta}_{1:t}^{(i)}}(\theta_{1:t}),$$

with normalized importance weights

$$W_t(\tilde{\Theta}_{1:t}^{(i)}) = \frac{w_t(\tilde{\Theta}_{1:t}^{(i)})}{\sum_{j=1}^I w_t(\tilde{\Theta}_{1:t}^{(j)})} = \frac{\gamma_t(\tilde{\Theta}_{1:t}^{(i)})/q_t(\tilde{\Theta}_{1:t}^{(i)})}{\sum_{j=1}^I \gamma_t(\tilde{\Theta}_{1:t}^{(j)})/q_t(\tilde{\Theta}_{1:t}^{(j)})}.$$

Estimate (12) is consistent satisfying the central limit theorem with asymptotic variance as $I \rightarrow \infty$, see (27) in Doucet and Johansen [6],

$$\frac{1}{I} \int \frac{\pi(\theta_{1:t} | X_{1:t}, \theta_0)^2}{q_t(\theta_{1:t})} (h(\theta_{1:t}) - \mathbb{E}[h(\Theta_{1:t}) | X_{1:t}])^2 d\theta_{1:t}.$$

Moreover, as mentioned in Doucet and Johansen [6], the asymptotic bias of this empirical estimate is of order $\mathcal{O}(1/I)$, the asymptotic variance of order $\mathcal{O}(1/I)$, and the mean squared error is asymptotically dominated by the variance term, see also Theorem 2.2 in Johansen and Evers [9].

Note that so far we have not used the sequential product structure (8)–(10) of our problems. This structure will help to control the computational complexity, this we are going to explore next.

4.2 Sequential Importance Sampling

Observe that the evaluation of the importance weights w_t can be complex if we do not benefit from the additional Markovian structure of problems (8)–(10). This can be achieved by considering a product structure for the importance density q_t , i.e. we choose (by a slight abuse of notation)

$$q_t(\theta_{1:t}) = q_t(\theta_{1:t} | X_{1:t}, \theta_0) = \prod_{s=1}^t q_s(\theta_s | X_{1:s}, \theta_{0:s-1}). \quad (13)$$

These (conditional) importance densities $q_s(\theta_s | X_{1:s}, \theta_{0:s-1})$ may also depend on $X_{1:s}$, often this is not highlighted in the notation. In the sequel we drop “conditional” in the terminology because notation already indicates this. Using (8) we calculate the importance weights recursively

$$\begin{aligned} w_t(\theta_{1:t}) &= \prod_{s=1}^t \frac{f(X_s | \theta_s) \pi(\theta_s | \theta_{s-1})}{q_s(\theta_s | X_{1:s}, \theta_{0:s-1})} \\ &= w_{t-1}(\theta_{1:t-1}) \frac{f(X_t | \theta_t) \pi(\theta_t | \theta_{t-1})}{q_t(\theta_t | X_{1:t}, \theta_{0:t-1})}, \end{aligned}$$

with initialization $w_0(\theta_{1:0}) = 1$. This allows us to define the incremental importance weights

$$\alpha_t(\theta_{1:t}) = \frac{f(X_t | \theta_t) \pi(\theta_t | \theta_{t-1})}{q_t(\theta_t | X_{1:t}, \theta_{0:t-1})},$$

and then the unnormalized importance weights under (8)–(10) and (13) are written as

$$w_t(\theta_{1:t}) = w_{t-1}(\theta_{1:t-1}) \alpha_t(\theta_{1:t}) = \prod_{s=1}^t \alpha_s(\theta_{1:s}).$$

Here, we see the sequential nature of the algorithm!

In view of the Markovian structure in (8)–(10) it makes sense to also choose a Markovian structure in (13) because the numerator of the incremental importance

weights $\alpha_t(\theta_{1:t})$ only depends on X_t, θ_t and θ_{t-1} , thus we choose importance density

$$q_t(\theta_t | X_t, \theta_{t-1}) \equiv q_t(\theta_t | X_{1:t}, \theta_{0:t-1}). \quad (14)$$

This provides incremental importance weights

$$\alpha_t(\theta_{t-1:t}) = \alpha_t(\theta_{1:t}) = \frac{f(X_t | \theta_t) \pi(\theta_t | \theta_{t-1})}{q_t(\theta_t | X_t, \theta_{t-1})}. \quad (15)$$

We arrive at the following algorithm under (8)–(10), (13) and (14).

Sequential importance sampling (SIS) algorithm.

▷ Set $\tilde{\Theta}_0^{(i)} = \theta_0$ and $w_0(\tilde{\Theta}_{1:0}^{(i)}) = 1$ for $i = 1, \dots, I$.

▷ Repeat for $s = 1, \dots, t$:

- repeat for $i = 1, \dots, I$:
 - sample $\tilde{\Theta}_s^{(i)} \sim q_s(\cdot | X_s, \tilde{\Theta}_{s-1}^{(i)})$;
 - calculate the importance weights

$$\alpha_s(\tilde{\Theta}_{s-1:s}^{(i)}) = \frac{f(X_s | \tilde{\Theta}_s^{(i)}) \pi(\tilde{\Theta}_s^{(i)} | \tilde{\Theta}_{s-1}^{(i)})}{q_s(\tilde{\Theta}_s^{(i)} | X_s, \tilde{\Theta}_{s-1}^{(i)})},$$

$$w_s(\tilde{\Theta}_{1:s}^{(i)}) = w_{s-1}(\tilde{\Theta}_{1:s-1}^{(i)}) \alpha_s(\tilde{\Theta}_{s-1:s}^{(i)});$$

- calculate for $i = 1, \dots, I$ the normalized importance weights

$$W_s(\tilde{\Theta}_{1:s}^{(i)}) = \frac{w_s(\tilde{\Theta}_{1:s}^{(i)})}{\sum_{j=1}^I w_s(\tilde{\Theta}_{1:s}^{(j)})} \propto w_s(\tilde{\Theta}_{1:s}^{(i)}).$$

This SIS algorithm provides empirical distributions for any $s = 1, \dots, t$

$$\hat{\pi}^{(I)}(\theta_{1:s} | X_{1:t}, \theta_0) = \sum_{i=1}^I W_s(\tilde{\Theta}_{1:s}^{(i)}) \delta_{\tilde{\Theta}_{1:s}^{(i)}}(\theta_{1:s}). \quad (16)$$

If we are only interested in $s = t$ we would not need to calculate $W_s(\tilde{\Theta}_{1:s}^{(i)})$ for $s < t$ in the SIS algorithm, however this is going to be important in the refinement of the SIS algorithm. Note that any marginal $\hat{\pi}^{(I)}(\theta_s | X_{1:t}, \theta_0)$ can easily be obtained, for $s = t$ this refers to Step 3 in the Kalman filter, for $s < t$ this refers to backward smoothing.

A main deficiency of the SIS algorithm is that the variance increases rapidly in the number of periods t considered, and thus a large number I of simulations is

needed in order to get accurate results, Doucet and Johansen [6] provide an example in Sect. 3.3. Therefore, variance reduction techniques should be applied and the SIS algorithm needs to be refined.

4.3 Sequential Monte Carlo with Adaptive Resampling

The SIS algorithm provides empirical distributions $\hat{\pi}^{(t)}(\theta_{1:s}|X_{1:s}, \theta_0)$ for $s = 1, \dots, t$, see (16). These empirical distributions are estimates for the true distributions $\pi(\theta_{1:s}|X_{1:s}, \theta_0)$. Resampling the particle system means that we sample from these empirical distributions $\hat{\pi}^{(t)}(\theta_{1:s}|X_{1:s}, \theta_0)$, that is, we may sample $\Theta_{1:s}^{(j)} \sim \hat{\pi}^{(t)}(\theta_{1:s}|X_{1:s}, \theta_0)$ i.i.d. for $j = 1, \dots, I$. Denote for $i = 1, \dots, I$

$$N_s^{(i)} = \sum_{j=1}^I \delta_{\tilde{\Theta}_{1:s}^{(i)}} \left(\Theta_{1:s}^{(j)} \right),$$

the number of times that $\tilde{\Theta}_{1:s}^{(i)}$ was re-chosen among the I trials $\Theta_{1:s}^{(1)}, \dots, \Theta_{1:s}^{(I)}$. This provides a second (resampled) empirical distribution

$$\bar{\pi}^{(t)}(\theta_{1:s}|X_{1:s}, \theta_0) = \sum_{i=1}^I \frac{N_s^{(i)}}{I} \delta_{\tilde{\Theta}_{1:s}^{(i)}}(\theta_{1:s}). \quad (17)$$

This resampled empirical distribution $\bar{\pi}^{(t)}(\theta_{1:s}|X_{1:s}, \theta_0)$ serves as an approximation to the empirical distribution $\hat{\pi}^{(t)}(\theta_{1:s}|X_{1:s}, \theta_0)$ and henceforth to $\pi(\theta_{1:s}|X_{1:s}, \theta_0)$.

The important remark here is that this resampling does not necessarily reduce the variance, but it may remove particles $\tilde{\Theta}_{1:s}^{(i)}$ that have low weights $W_s(\tilde{\Theta}_{1:s}^{(i)})$ (are in an unlikely region of the probability space) and we only work in the part of the probability space that has a sufficiently high probability mass. There are the following important remarks:

- There are more efficient resampling techniques than the i.i.d. resampling one proposed above (which in fact provides a multinomial distribution). Doucet and Johansen [6] support the systematic resampling technique. It samples $U_1 \sim \text{Uniform}[0, 1]$ and then defines $U_{i+1} = U_1 + i/I$ for $i = 1, \dots, I-1$. An unbiased resampled distribution is obtained by setting

$$N_s^{(i)} = \sum_{j=1}^I 1_{\left\{ \sum_{k=1}^{i-1} w_s(\tilde{\Theta}_{1:s}^{(k)}) \leq U_j \leq \sum_{k=1}^i w_s(\tilde{\Theta}_{1:s}^{(k)}) \right\}}. \quad (18)$$

- For convergence results we refer to the literature mentioned in Doucet and Johansen [6].

- The resampling step may lead to degeneracy of $\bar{\pi}^{(I)}(\theta_{1:s}|X_{1:s}, \theta_0)$ with positive probability. Therefore, one should always back-test whether the resulting empirical distribution is sufficiently rich for the indexes $s = 1, \dots, t$ under consideration.
- In many cases one applies adaptive resampling, i.e. the resampling step is only applied if the weights are too disperse. One way to measure dispersion is the effective sample size (ESS) defined by

$$\text{ESS}_s = \left(\sum_{i=1}^I \left(W_s(\tilde{\theta}_{1:s}^{(i)}) \right)^2 \right)^{-1} \in [1, I]. \quad (19)$$

The resampling is then only applied if the ESS is too small. Note that if all particles have the same weight $1/I$, then ESS_s is equal to I , if one particle concentrates the entire probability mass, then ESS_s is equal to 1.

This provides the following algorithm under (8)–(10), (13) and (14) and given resampling threshold $\chi \in [1, I]$.

Sequential Monte Carlo (SMC) with adaptive resampling algorithm.

▷ Set $\tilde{\theta}_0^{(i)} = \theta_0$ and $w_0(\tilde{\theta}_{1:0}^{(i)}) = 1$ for $i = 1, \dots, I$.

▷ Repeat for $s = 1, \dots, t$:

- repeat for $i = 1, \dots, I$:
 - sample $\tilde{\theta}_s^{(i)} \sim q_s(\cdot | X_s, \tilde{\theta}_{s-1}^{(i)})$;
 - calculate the importance weights

$$\alpha_s(\tilde{\theta}_{s-1:s}^{(i)}) = \frac{f(X_s | \tilde{\theta}_s^{(i)}) \pi(\tilde{\theta}_s^{(i)} | \tilde{\theta}_{s-1}^{(i)})}{q_s(\tilde{\theta}_s^{(i)} | X_s, \tilde{\theta}_{s-1}^{(i)})},$$

$$w_s(\tilde{\theta}_{1:s}^{(i)}) = w_{s-1}(\tilde{\theta}_{1:s-1}^{(i)}) \alpha_s(\tilde{\theta}_{s-1:s}^{(i)});$$

- calculate for $i = 1, \dots, I$ the normalized importance weights

$$W_s(\tilde{\theta}_{1:s}^{(i)}) = \frac{w_s(\tilde{\theta}_{1:s}^{(i)})}{\sum_{j=1}^I w_s(\tilde{\theta}_{1:s}^{(j)})} \propto w_{s-1}(\tilde{\theta}_{1:s-1}^{(i)}) \alpha_s(\tilde{\theta}_{s-1:s}^{(i)}),$$

and the corresponding ESS_s according to (19);

- if $\text{ESS}_s \leq \chi$ resample $\theta_{1:s}^{(1)}, \dots, \theta_{1:s}^{(I)}$ from (16) and set for $i = 1, \dots, I$

$$w_s(\tilde{\theta}_{1:s}^{(i)}) = 1, \quad W_s(\tilde{\theta}_{1:s}^{(i)}) = \frac{1}{I} \quad \text{and} \quad \tilde{\theta}_{1:s}^{(i)} \leftarrow \theta_{1:s}^{(i)}.$$

Operation $\tilde{\theta}_{1:s}^{(i)} \leftarrow \theta_{1:s}^{(i)}$ is an assignment in the R coding language sense.

This SMC with adaptive resampling algorithm provides empirical distributions for any $s = 1, \dots, t$

$$\hat{\pi}^{(t)}(\theta_{1:s} | X_{1:s}, \theta_0) = \sum_{i=1}^I W_s(\tilde{\Theta}_{1:s}^{(i)}) \delta_{\tilde{\Theta}_{1:s}^{(i)}}(\theta_{1:s}). \quad (20)$$

Note that the weights $W_s(\tilde{\Theta}_{1:s}^{(i)})$ and particles $\tilde{\Theta}_{1:s}^{(i)}$ may now differ from the ones of the SIS algorithm (16) due to the potential resampling step that is performed whenever $\text{ESS}_u \leq \chi$ for $u \leq s$. Often one chooses resampling threshold $\chi = I/2$.

5 Examples and Backward Smoothing

In Sect. 5.1 we study the two linear state space models introduced in Sect. 2, in Sect. 5.2 we explore the non-Gaussian and non-linear state space model of Sect. 3, and in Sect. 5.3 we consider a (new) model that may serve as a stochastic volatility model for asset prices.

5.1 Linear State Space Models

We start by considering the linear state space models (1)–(2) and (3)–(4) in the Gaussian and the gamma case, respectively. As parameters we choose

$$\theta_0 = 1, \quad a = 1/10, \quad b = 9/10, \quad \tau = \sqrt{1/10}, \quad \sigma = 1/2. \quad (21)$$

The transition systems (1) and (3) have the same first two moments, but different distributional properties, in particular, the gamma one is bounded from below by zero, whereas the Gaussian one is unbounded from below. In Fig. 1 (lhs) we plot 10 sample paths $\Theta_{1:t}$, $t = 100$, in the transition system for each of the two linear models. This figure is complemented by the empirical means and the confidence bounds of 2 empirical standard deviations (of 1'000 simulations). We see that these measures coincide for the Gaussian and gamma cases, however, the sample paths look rather different in the two models. Based on two selected samples $\Theta_{1:t}$ (darker trajectories) in the transition system we draw a sample $X_{1:t}$ (for each model) in the measurement system according to (2) and (4), respectively. The ones given in Fig. 1 (rhs) are used in the further analysis in order to infer Θ_t from $X_{1:t}$, i.e. we aim at calculating the Bayesian estimate

$$\theta_{t|t} = \mathbb{E}[\Theta_t | X_{1:t}].$$

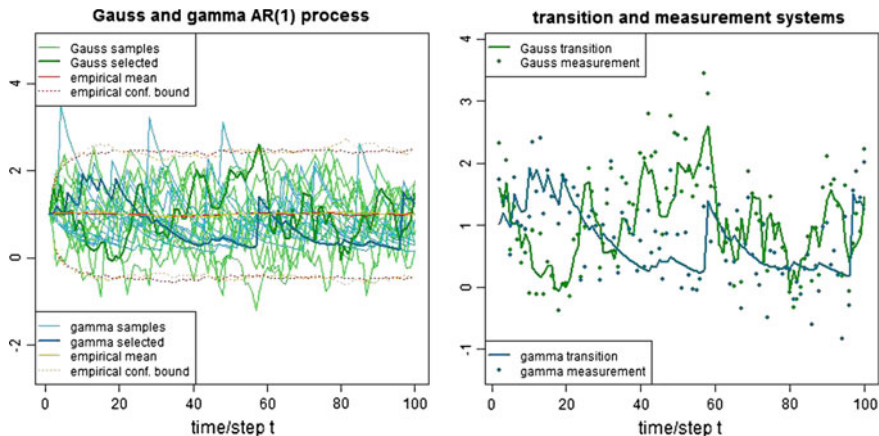


Fig. 1 (*lhs*) Simulated sample paths $\Theta_{1:100}$ of the Gaussian linear transition system (1) and the gamma linear transition system (3) for parameters (21) complemented by the empirical means and the confidence bounds of 2 empirical standard deviations; the darker sample paths were selected for the subsequent state space model analysis; (*rhs*) empirical samples $X_{1:100}$ and $\Theta_{1:100}$ in the two linear state space models (1)–(2) and (3)–(4) for parameters (21) and the selected sample paths of the (*lhs*)

In the Gaussian linear state space model we can calculate $\theta_{t|t}$ exactly, using the Kalman filter; in the gamma linear state space model the Kalman filter provides the best linear (credibility) approximation to the conditional mean of Θ_t , given $X_{1:t}$, see Chap. 9 in Bühlmann and Gisler [1]. The Kalman filter results are presented in Fig. 2. We observe that the Kalman filter achieves to estimate the true $\Theta_{1:t}$ quite accurately, however the noise in $X_{1:t}$ slightly distorts these estimates. Of course, the bigger the parameter σ the harder it becomes to infer the transition system $\Theta_{1:t}$ from the observations $X_{1:t}$.

Next we explore the SIS and the SMC algorithms and compare the results to the Kalman filter ones. We therefore need to choose importance densities q_t . A simple (non-optimal) way is to choose $q_t(\cdot|X_t, \theta_{t-1}) = \pi(\cdot|\theta_{t-1})$, see also (14). This choice provides incremental importance weights under model assumptions (2) and (4) given by

$$\alpha_t(\theta_{t-1:t}) = f(X_t|\theta_t) = \exp\left\{-\frac{(X_t - \theta_t)^2}{2\sigma^2}\right\}. \quad (22)$$

Note that these incremental importance weights are uniformly bounded. In the case of the Gaussian innovations (1) the choice of the importance density q_t could be improved because we can directly simulate from the posterior distribution (which is

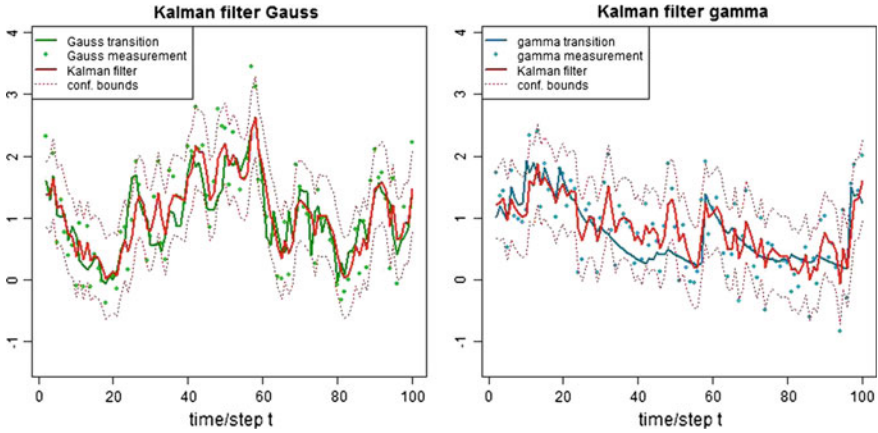


Fig. 2 (Estimated) posterior means $\theta_{t|t} = \mathbb{E}[\Theta_t | X_{1:t}]$ using the Kalman filter complemented by the (estimated) confidence bounds of 2 posterior standard deviations $\text{Var}(\Theta_t | X_{1:t})^{1/2}$: (*lhs*) (exact) Gaussian linear state space model posterior means and (*rhs*) (estimated) gamma linear state space model posterior means

a multivariate normal one). However, we refrain from doing so because our choice works in all three models introduced above and leads to identical incremental importance weights (22).

Finally, we choose $I = 10'000$ independent samples and for the SMC with adaptive resampling algorithm we choose resampling threshold $\chi = I/2$. All parameters are now determined and we can sample from to the SIS and SMC algorithms. In Fig. 3

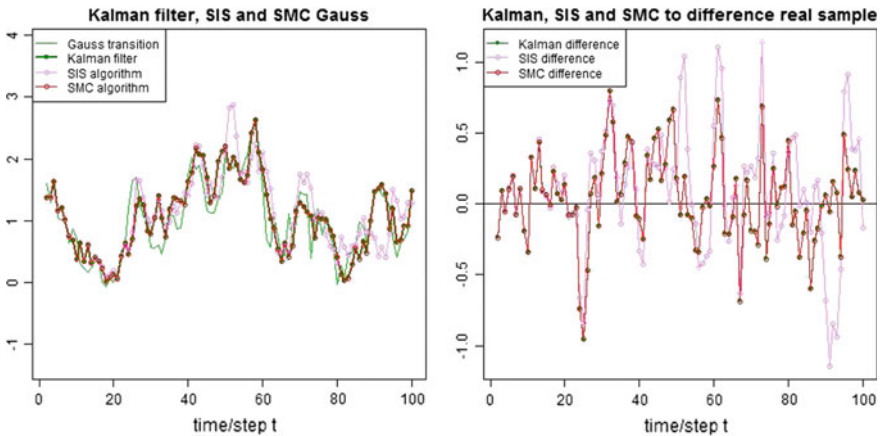


Fig. 3 Comparison between the true sample $\Theta_{1:t}$, the Kalman filter estimate $\theta_{t|t}$, the SIS estimate and the SMC estimate in the Gaussian linear state space model (1)–(2): (*lhs*) estimates and (*rhs*) resulting differences to the true sample $\Theta_{1:t}$

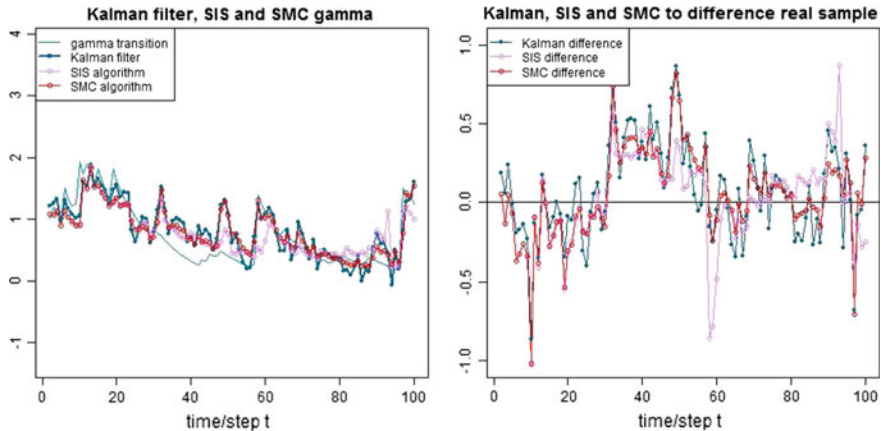


Fig. 4 Comparison between the true sample $\Theta_{1:t}$, the Kalman filter estimate $\theta_{t|t}$, the SIS estimate and the SMC estimate in the gamma linear state space model (3)–(4): (lhs) estimates and (rhs) resulting differences to the true sample $\Theta_{1:t}$

(lhs) we provide the results for the Gaussian linear state space model (1)–(2). We observe that the SMC estimates coincide almost perfectly with the (exact) Kalman filter estimates $\theta_{t|t}$. On the other hand the SIS estimates start to deviate from $\theta_{t|t}$ after roughly $t = 20$ time steps because the normalized importance weights $W_t(\tilde{\Theta}_{1:t}^{(i)})$ start to be too disperse and a resampling step should be applied. Figure 3 (rhs) shows the differences between the estimates $\theta_{t|t}$ and the true factors Θ_t . Also here we see that the SIS estimates start to have difficulties with increasing t .

Next we analyze the same plots for the gamma linear state space model (3)–(4). In this model the Kalman filter gives a best linear credibility approximation to the true posterior mean $\mathbb{E}[\Theta_t | X_{1:t}]$, and the SIS and SMC estimates should be exact up to simulation error. Also here we see that the SIS algorithm has a poor behavior for bigger t and one should prefer the SMC estimate. Interestingly, the SMC estimate clearly differs from the Kalman filter estimate because of the different distributional properties from the Gaussian ones. In particular, the Kalman filter has difficulties to cope with the tails which leads to too extreme estimates. We conclude that we should choose the SMC estimates, as soon as the number of samples I is sufficiently large (Fig. 4).

In Fig. 5 we plot the posterior standard deviations $\text{Var}(\Theta_t | X_{1:t})^{1/2}$. In the Gaussian model the SMC and the Kalman filter estimates coincide, whereas the SIS estimate has a poor volatile behavior. In the gamma model the SMC estimate is also volatile and of smaller size than the Kalman filter estimate (which is also not exact in the

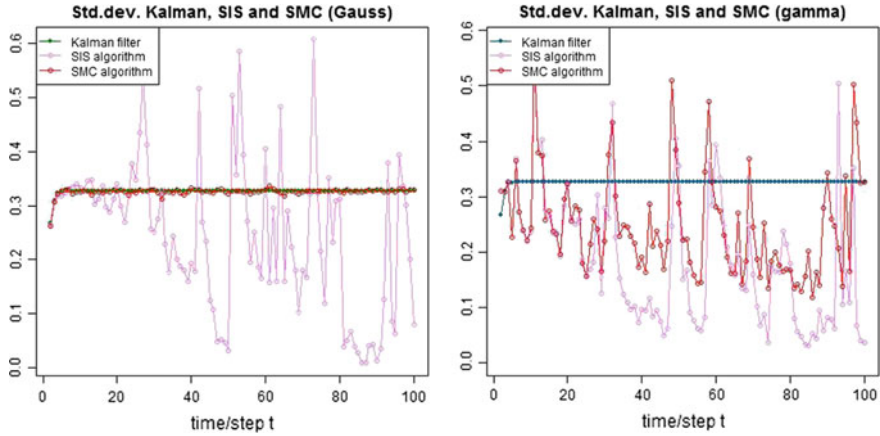


Fig. 5 (Estimated) posterior standard deviation $\text{Var}(\Theta_t|X_{1:t})^{1/2}$ from the Kalman filter, the SIS and SMC algorithms for the (*lhs*) Gaussian and the (*rhs*) the gamma linear state space models

non-Gaussian case). Here one should probably use a smoothed version of the SMC estimate because the Kalman filter over-estimates the posterior standard deviation because it cannot cope with the tail of the gamma distribution.

5.2 Non-Gaussian and Non-linear State Space Models

In this section we explore the non-linear state space model (6)–(7). The posterior density is given by (8) and under the choice $q_t(\cdot|X_t, \theta_{t-1}) = \pi(\cdot|\theta_{t-1})$ we obtain incremental importance weights (22). These are, of course, again bounded and we can apply the algorithm from before, the only change lies in the choice of the importance distribution which now has a non-linear scaling, see (8).

In Fig. 6 (rhs) we provide an explicit sample $\Theta_{1:t}$ for the transition system and a corresponding sample $X_{1:t}$ for the measurement system. Note that we provide exactly the same random samples, but with scaling (6) instead of scaling (3). Figure 7 then shows the resulting Kalman filter approximations, where for the non-linear model (6) we use first order Taylor approximation $\sqrt{\Theta_{t-1}} \approx 1$ in the Kalman filter application. Note that this could be refined by a second order Taylor approximation $\sqrt{\Theta_{t-1}} \approx 1 + (\Theta_{t-1} - 1)/2$.

In Fig. 8 we then compare the Kalman filter, SIS and SMC estimates of the posterior means $\mathbb{E}[\Theta_t|X_{1:t}]$. We observe that the Kalman filter receives too high peaks and should not be used for the gamma non-linear state space model. The SIS estimate

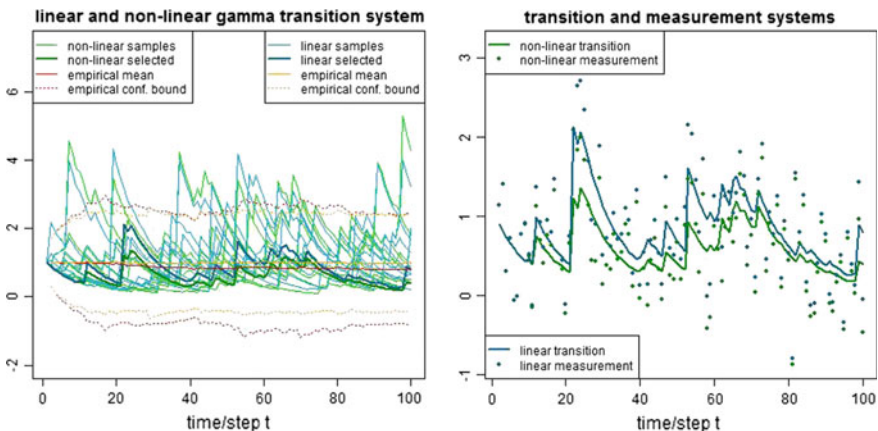


Fig. 6 (*lhs*) Simulated sample paths $\Theta_{1:100}$ of the gamma linear transition system (3) and the gamma non-linear transition system (6) for parameters (21) complemented by the empirical means and the confidence bounds of 2 empirical standard deviations; the darker sample paths were selected for the state space model analysis; (*rhs*) empirical samples $X_{1:100}$ and $\Theta_{1:100}$ in the two models (note that we give exactly the same random samples, only scaling in the transition system differs)

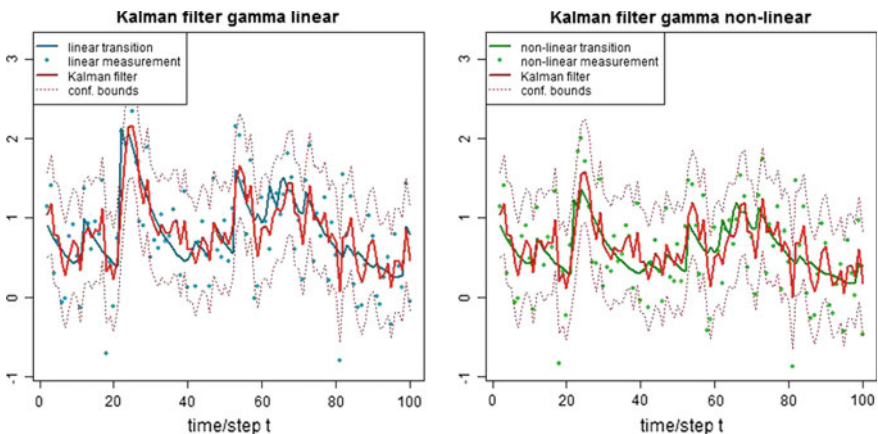


Fig. 7 Estimated posterior means $\theta_{t|t} = \mathbb{E}[\Theta_t | X_{1:t}]$ using the Kalman filter complemented by the estimated confidence bounds of 2 posterior standard deviations: (*lhs*) gamma linear and (*rhs*) gamma non-linear state space models; in the latter we approximate $\sqrt{\Theta_{t-1}} \approx 1$

becomes poorer for bigger t , hence the SMC estimate, that looks reasonable in Fig. 8 (*rhs*), should be preferred. In Fig. 9 we also see that the Kalman filter over-estimates posterior variance because it cannot cope with the gamma distribution and it cannot interpret scaling $\sqrt{\Theta_{t-1}}$ in (6). For this reason a smoothed version of the SMC posterior standard deviation estimate should be used.

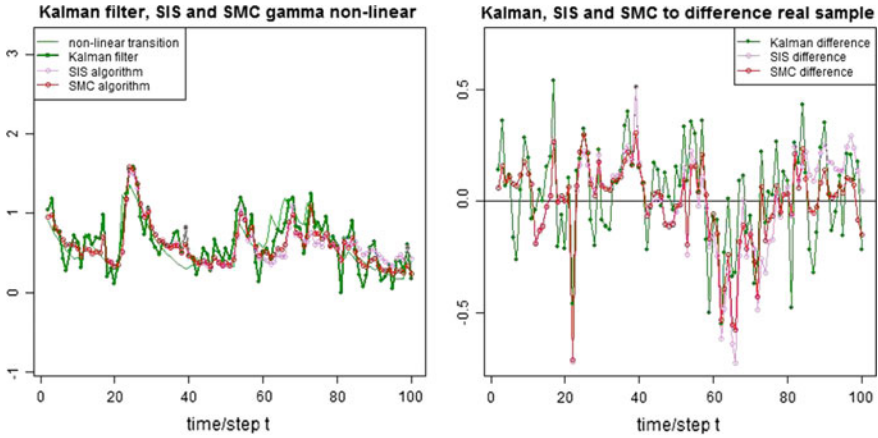
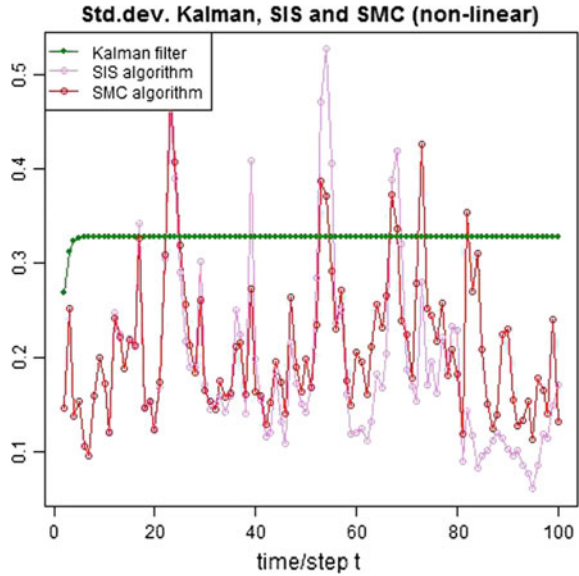


Fig. 8 Comparison between the true sample $\Theta_{1:t}$, the Kalman filter estimate $\theta_{t|t}$, the SIS estimate and the SMC estimate in the gamma non-linear state space model (6)–(7): (lhs) estimates and (rhs) resulting differences to the true sample $\Theta_{1:t}$

Fig. 9 Estimated posterior standard deviation $\text{Var}(\Theta_t | X_{1:t})^{1/2}$ from the Kalman filter, the SIS and SMC algorithms for the gamma non-linear state space model



5.3 Stochastic Volatility Model for Asset Prices

We close this section with an example that considers stochastic volatility modeling in the transition system and (de-trended) logarithmic asset prices in the measurement system. Inspired by the Heston [8] model we consider a gamma non-linear transition system for the stochastic volatility process and a log-normal model for the de-trended

asset prices. Note that de-trended asset prices means that the log-normal distribution has to have a negative mean parameter being equal to minus one half of the variance parameter. This motivates the following model:

- (i) The *transition system* is described by a process $(\Theta_t)_{t \in \mathbb{N}_0}$ with $\Theta_0 = \theta_0 = 1$ and for $t \geq 1$

$$\Theta_t = b\Theta_{t-1} + \sqrt{\Theta_{t-1}}\eta_t, \quad (23)$$

for $b \in \mathbb{R}$ and $(\eta_t)_{t \geq 1}$ being i.i.d. gamma distributed with $\mathbb{E}[\eta_t] = a$ and $\text{Var}(\eta_t) = \tau^2$.

- (ii) The *measurement system* is described by a process $(X_t)_{t \in \mathbb{N}}$ with for $t \geq 1$

$$X_t = -\sigma^2\Theta_t/2 + \sigma\sqrt{\Theta_t}\varepsilon_t, \quad (24)$$

for $\sigma > 0$ and $(\varepsilon_t)_{t \geq 1}$ being i.i.d. standard Gaussian distributed and being independent of process $(\eta_t)_{t \geq 1}$.

The posterior density of $\Theta_{1:t}$ for given observations $X_{1:t}$ is given by

$$\begin{aligned} \pi(\theta_{1:t} | X_{1:t}, \theta_0) &\propto f(X_{1:t} | \theta_{1:t}) \pi(\theta_{1:t} | \theta_0) \\ &\propto \prod_{s=1}^t \frac{1}{\sqrt{\theta_s \theta_{s-1}}} \left(\frac{\theta_s - b\theta_{s-1}}{\sqrt{\theta_{s-1}}} \right)^{\gamma-1} \\ &\quad \times \exp \left\{ -\frac{(X_s + \sigma^2\theta_s/2)^2}{2\sigma^2\theta_s} - c \left(\frac{\theta_s - b\theta_{s-1}}{\sqrt{\theta_{s-1}}} \right) \right\} 1_{\{\theta_s \geq b\theta_{s-1}\}}. \end{aligned} \quad (25)$$

The transition system of the stochastic volatility process $(\Theta_t)_{t \in \mathbb{N}_0}$ given in (23) is exactly the same as (6). Therefore, Fig. 6 (*lhs*) provides typical trajectories for the parameters (21). In contrast to the previous models we now also have a non-linearity in the measurement system (24). We would like to indicate two different extreme cases: (i) for $\sigma \gg 1$ very large we obtain

$$X_t = -\sigma^2\Theta_t/2 + \sigma\sqrt{\Theta_t}\varepsilon_t \approx -\sigma^2\Theta_t/2. \quad (26)$$

For this reason we expect to detect the transition system rather accurately in this case (in fact the filter becomes almost superfluous); (ii) for $\sigma \ll 1$ very small we obtain

$$X_t = -\sigma^2\Theta_t/2 + \sigma\sqrt{\Theta_t}\varepsilon_t \approx \sigma\sqrt{\Theta_t}\varepsilon_t. \quad (27)$$

In this case we expect the de-trending term $-\sigma^2\Theta_t/2$ to be almost useless in supporting the filtering algorithm.

For the transition system we use the parameters (21) and the importance density $q_t(\cdot | X_t, \theta_{t-1})$ is chosen as $\pi(\cdot | \theta_{t-1})$. In Fig. 10 we present the SIS and SMC algorithm results for $\sigma = 0.25$. These results are compared to the SMC algorithm results provided by model (27) (since $\sigma = 0.25$ is comparably small). The first observation

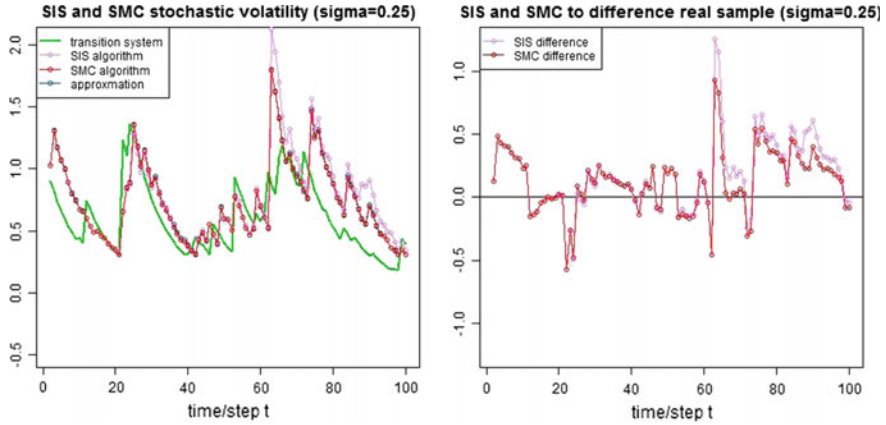


Fig. 10 Comparison between the true sample $\Theta_{1:t}$, the SIS estimate and the SMC estimate in the stochastic volatility model (23)–(24) for $\sigma = 0.25$: (lhs) estimates and (rhs) resulting differences to the true sample $\Theta_{1:t}$; the approximation on the (lhs) uses (27) with SMC and it is almost identically equal to the original SMC estimate (and therefore not visible in the plot)

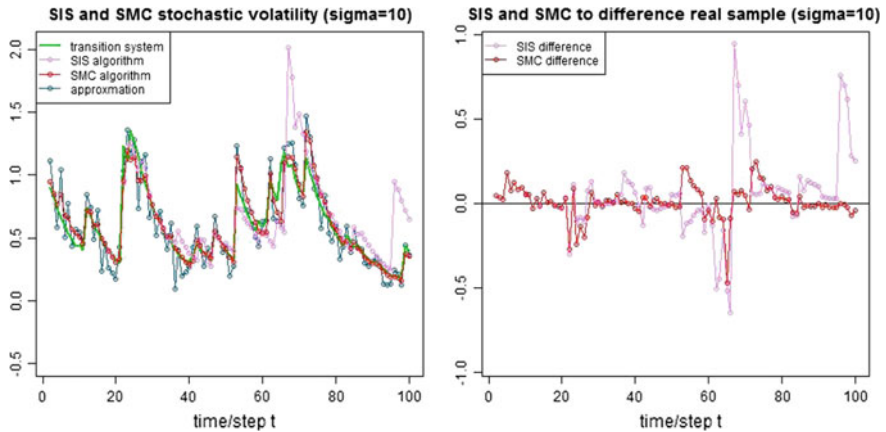


Fig. 11 Comparison between the true sample $\Theta_{1:t}$, the SIS estimate and the SMC estimate in the stochastic volatility model (23)–(24) for $\sigma = 10$: (lhs) estimates and (rhs) resulting differences to the true sample $\Theta_{1:t}$; the approximation on the (lhs) uses (26)

is that we cannot distinguish the SMC results from models (24) and (27), thus, our de-trending term is too small to be helpful to improve inference of the transition system. Secondly, we observe based on scaling (27) that the transition innovation η_t and the measurement innovation ε_t live on a competing scale which makes it difficult to infer Θ_t from $X_{1:t}$. In fact, as can be seen from Fig. 10 (lhs), this leads to a visible delay in the filtered estimation of Θ_t . This is quite a typical phenomenon in filtering and it comes from the fact that “we cannot look into the future for smoothing”.

In our second example we choose a large $\sigma = 10$ and compare the solution of model (24) to the (deterministic) one of model (26), see Fig. 11 (lhs). We see that

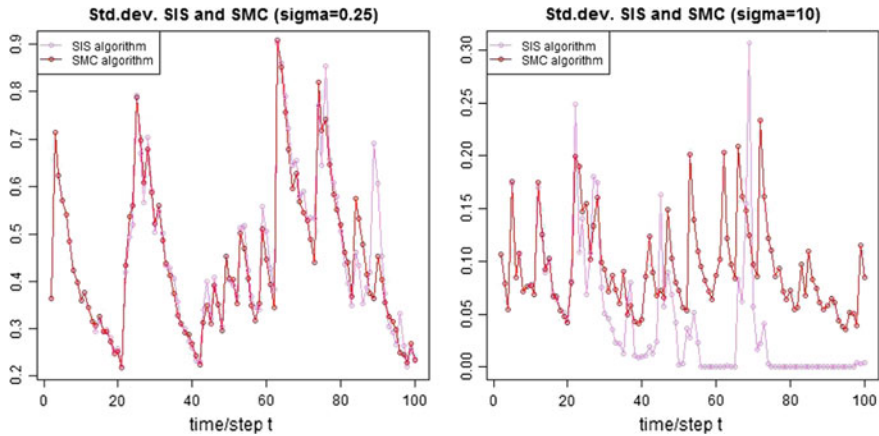


Fig. 12 Estimated posterior standard deviation $\text{Var}(\Theta_t | X_{1:t})^{1/2}$ from the SIS and SMC algorithms for stochastic volatility model (23)–(24) for (lhs) $\sigma = 0.25$ and (rhs) $\sigma = 10$

approximation (26) clearly fluctuates around the true Θ_t , but fluctuation is still a bit too large to directly extract Θ_t from the observations X_t . This means that σ is not sufficiently large and we should use SMC filtering. The SMC filter provides very good results, in fact much better results than in the previous example $\sigma = 0.25$, because the predictive power of the de-trending term is already quite large in this situation. This can also be seen by comparing Figs. 10 (rhs) and 11 (rhs). Finally, in Fig. 12 we present the posterior standard deviations which (in a smoothed version) allow us to construct confidence bounds for the prediction. For $\sigma = 0.25$ they are in the range of 0.4, for $\sigma = 10$ they are of size 0.1 which is clearly smaller (due to the higher predictive power of the de-trending term).

5.4 Backward Smoothing and Resample-Moves

In Fig. 10 (lhs) we have seen that the filter estimates $\theta_{t|t}$ always come with a delay in reaction. Backward smoothing means that we use later information to re-assess the value of Θ_t . We have already met this idea in Sect. 3.2 and after the description of the SIS algorithm. Basically this means that we infer Θ_t by considering

$$\theta_{t|T} = \mathbb{E} [\Theta_t | X_{1:T}] \quad \text{for later time points } T \geq t.$$

Using the simulated samples we calculate empirically at time $T \geq t$, see also (12),

$$\widehat{\mathbb{E}}^{(I)} [\Theta_t | X_{1:T}] = \sum_{i=1}^I \tilde{\Theta}_t^{(i)} \frac{w_T(\tilde{\Theta}_{1:T}^{(i)})}{\sum_{j=1}^I w_T(\tilde{\Theta}_{1:T}^{(j)})}.$$

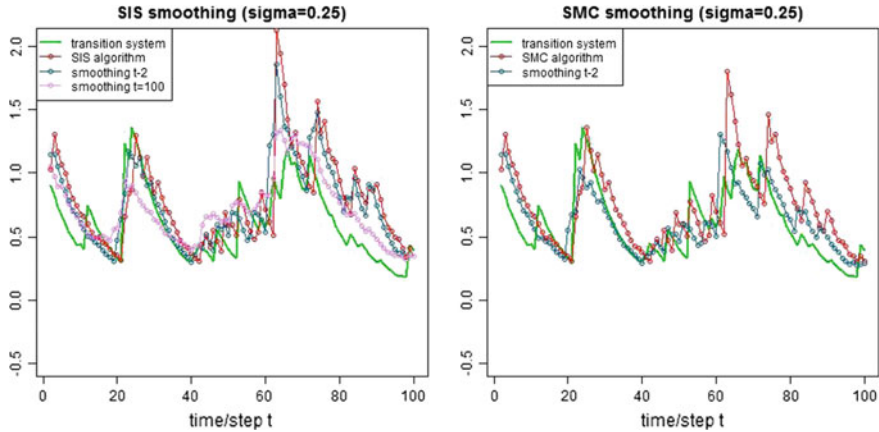


Fig. 13 Comparison between the true sample $\Theta_{1:T}$, the SIS estimate and the SMC estimate in the stochastic volatility model (23)–(24) for $\sigma = 0.25$: (lhs) SIS estimates of $\theta_{t|t}$ and of $\theta_{t|t+2}$ and $\theta_{t|100}$ (backward smoothed) and (rhs) SMC estimates of $\theta_{t|t}$ and of $\theta_{t|t+2}$ (backward smoothed)

This seems straightforward, however, this empirical method needs some care! It can be done in a direct manner for the SIS algorithm. We provide the results in Fig. 13 (lhs) for the stochastic volatility model with $\sigma = 0.25$ for $T = t + 2$ and $T = 100$. We see a clear left shift of the estimates $\theta_{t|T}$, that is, for more information $X_{1:T}$ we can better distinguish the competing innovations ε_t and η_t . If T is too large the model looks like it is over-smoothing, which means that an appropriate time lag $T - t$ needs to be determined for smoothing.

For the SMC algorithm backward smoothing is much more delicate due to the resampling step. Observe that the resampling step (17) at time s tends to select only the particles $\tilde{\Theta}_{1:s}^{(i)}$ that have a sufficiently large importance weight $W_s(\tilde{\Theta}_{1:s}^{(i)})$. From these selected particles I new trajectories are simulated into the future after time s . But, this selection also implies a thinning of the past trajectories (because part of the particles are dropped in the resampling step and, thus, also their history). Applying the resampling step several times therefore leads to very poor properties at the beginning of the trajectories because of the successive selection of the fittest particles. For this reason in SMC sampling it is preferable to do backward smoothing for time lags that are smaller than the time lags between adaptive resampling steps. An example for $T = t + 2$ is presented in Fig. 13 (rhs). We see that also here backward smoothing leads to better inference compared to the true value Θ_t (which can be seen by the left shift of $\theta_{t|t+2}$ versus $\theta_{t|t}$).

There are ways to deal with the deficiency of the SMC algorithm that it cannot be used for arbitrary backward smoothing because of potential degeneracy of trajectories for big time lags. Ways to fix these problems are, for instance, a resample-move or a block sample step. Such methods mainly aim at spreading the degenerate part of the trajectories by a Markov chain Monte Carlo (MCMC) step using the Metropolis–

Hastings algorithm, the Gibbs sampler or related techniques. We briefly explain the resample-move, for more details we refer to Doucet and Johansen [6].

Consider the posterior density $\pi(\theta_{1:t}|X_{1:t}, \theta_0)$ as the invariant (stationary limit) distribution of a Markov process $(\Theta_{1:t}^{(s)})_{s \in \mathbb{N}}$ having transition kernel $K(\theta_{1:t}^{(s+1)}|\theta_{1:t}^{(s)})$. As a consequence we obtain identity

$$\int \pi(\theta_{1:t}|X_{1:t}, \theta_0) K(\theta'_{1:t}|\theta_{1:t}) d\theta_{1:t} = \pi(\theta'_{1:t}|X_{1:t}, \theta_0).$$

This immediately implies that for given $\Theta_{1:t} \sim \pi(\cdot|X_{1:t}, \theta_0)$ we can resample from the transition kernel $\Theta'_{1:t} \sim K(\cdot|\Theta_{1:t})$ and the resulting sample is still distributed according to $\pi(\cdot|X_{1:t}, \theta_0)$. As a consequence if we obtain in the adaptive resampling step of the SMC algorithm $N_t^{(i)} > 1$ particles that have the same past history $\tilde{\Theta}_{1:t}^{(i)}$, see (17), we can spread these particles by applying an independent resample-move to each particle using transition kernel $K(\cdot|\tilde{\Theta}_{1:t}^{(i)})$. In addition, MCMC sampling theory provides explicit constructions of transition kernels K for given invariant distributions $\pi(\cdot|X_{1:t}, \theta_0)$ as soon as the latter are known up to the normalizing constants (which is the case in our situation, see for instance (8)). In practice, only a fixed time lag is resampled by this MCMC step, firstly because then one does not need to deal with different lengths of resample-moves as t increases, and secondly because filtering is also only applied to limited time lags (to preserve stationarity in real world time series).

6 Conclusions and Outlook

Gaussian linear state space models can be solved with Kalman filter techniques. Non-Gaussian or/and non-linear state space models can only be solved numerically. A powerful simulation method is sequential Monte Carlo (SMC) sampling. The resulting sampler is a version of importance sampling that benefits from the underlying Markovian structure of state space models. We have presented the SMC sampler and we have illustrated it in terms of several examples.

This outline and the examples presented were always based on known densities (up to the normalizing constants). Unknown model parameters add an additional complexity to the problem. Solving the latter problem may take advantage of Markov chain Monte Carlo (MCMC) methods, in particular, the particle marginal Metropolis-Hastings (PMMH) algorithm may be useful.

References

1. Bühlmann H, Gisler A (2005) A course in credibility theory and its applications. Springer, Berlin
2. Creal D (2012) A survey of sequential Monte Carlo methods for economics and finance. *Econ Rev* 31(3):245–296
3. Del Moral P (1996) Non linear filtering: interacting particle solution. *Markov Process Relat Fields* 2(4):555–580
4. Del Moral P, Doucet A, Jasra A (2006) Sequential Monte Carlo samplers. *J R Stat Soc Ser B* 68(3):411–436
5. Del Moral P, Peters GW, Vergé C (2012) An introduction to stochastic particle integration methods: with applications to risk and insurance. In: Dick J, Kuo FY, Peters GW, Sloan IH (eds) *Monte Carlo and Quasi-Monte Carlo Methods*. Springer Proceedings in Mathematics & Statistics. Springer, Berlin
6. Doucet A, Johansen AM (2011) A tutorial on particle filtering and smoothing: fifteen years later. In: Crisan D, Rozovsky B (eds) *Handbook of nonlinear filtering*. Oxford University Press, Oxford
7. Gordon NJ, Salmond DJ, Smith AFM (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc F Radar Sig Process* 140(2):107–113
8. Heston SL (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev Financ Stud* 6(2):327–343
9. Johansen AM, Evers L (2007) *Monte Carlo Methods*. Lecture Notes. University of Bristol, Bristol
10. Kalman RE (1960) A new approach to linear filtering and prediction problems. *J Basic Eng* 82:35–45

Robustness in Econometrics

Kreinovich, V.; Sriboonchitta, S.; Huynh, V.-N. (Eds.)

2017, X, 705 p. 129 illus., 120 illus. in color., Hardcover

ISBN: 978-3-319-50741-5