On the particular construction of the SMC sampling method for Bayesian filtering

Katarzyna Brzozowska-Rup[1]

Abstract

We consider the theoretical problem of time series which arises when the distribution of the observed variable is the de facto conditional distribution. The Kalman filter provides an effective solution to the linear Gaussian filtering problem. However, when state/measurement functions are highly non-linear, and posterior probability distribution of the state is non-Gaussian, the optimal linear filter and its modifications do not provide satisfactory results. We propose the Sequential Monte Carlo method, known generically as particle filter, which combines importance sampling and resampling schemes. In particular, we present a construction of an auxiliary particle filter algorithm using the Pearson curves technique for approximation of importance weights of simulated particles. The effectiveness of the method is discussed and illustrated by numerical results based on the simulated stochastic volatility process SV.

Keywords: *Bayesian filtering, sequential Monte Carlo methods, nonlinear and non-Gaussian state space models, stochastic volatility process SV, Pearson's curves technique.*

JEL Classification*: C53, C100, C150*
AMS Classification*: 62M20, 62L12, 60J05*

1. Introduction

The article concerns the nonlinear filtering problems which appear in many diverse fields including economics, statistical signal and numerous processing engineering problems. In the following research, we consider a discrete state space representation models DSSM (sometimes termed as 'hidden Markov models', HMM) and show how the Sequential Monte Carlo (SMC) methods can be used to approximate the filtering and predictive distribution functions. A DSSM consists of a stochastic propagation equation, which links the current state vector to the prior state vector, and a stochastic observation equation, which links the observation data to the current state vector.

Let us consider a probability space (Ω, Σ, P) on which we define the following model. For any parameter $\theta \in \Theta$, the hidden (latent) state process $\{X_t : t \geq 1\}$ is a stationary and ergodic Markov process, characterized by its Markov transition probability distribution $p(x'|x)$

[1] Kielce University of Technology, Faculty of Mathematics, al. Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, brzozows@poczta.fm

$$X_t \left| (X_{1:t-1} = x_{1:t-1}, Y_{1:t-1} = y_{1:t-1}, \theta) \sim p(x_t \left| x_{t-1}, \theta \right.), \right. \tag{1}$$

and a known, invariant initial probability density of state $X_1 \sim p\ (x_1 \left| \theta \right.)$. As indicated by its name, $\{X_t\}$ is observed not directly but through another process $\{Y_t : t = 1,...,T\}$. The observations are assumed to be conditionally independent when given $\{X_t\}$, and their common marginal probability distribution is expressed as follows

$$Y_t \left| (X_{1:t-1} = x_{1:t-1}, X_t = x_t, X_{t+1:T} = x_{t+1:T}, Y_{1:T} = y_{1:T}, \theta) \sim p(y_t \left| x_t, \theta \right.). \right. \tag{2}$$

Depending on context, p will denote a probability distribution or a probability density function. In addition, for any process $\{Z_t\}$ the realizations from time $t = i$ to $t = j$ will be denoted as $z_{i:j} = (z_i, z_{i+1},..., z_j)$. We assume that the static parameter θ is known, then sequential inference on the latent process $\{X_t\}$ is typically based on the sequence of posterior distributions $p(x_{1:t} \left| y_{1:t}, \theta \right.)$, which each summarizes all the information collected about $X_{1:t}$ up to time t. In a Bayesian context, sequential estimation of these distributions can be easily achieved using the following updating formula, for $t \geq 2$

$$p(x_{1:t} \left| y_{1:t}, \theta \right.) = p(x_{1:t-1} \left| y_{1:t-1}, \theta \right.) \frac{p(y_t \left| x_t, \theta \right.) p(x_t \left| x_{t-1}, \theta \right.)}{p(y_t \left| y_{1:t-1}, \theta \right.)}. \tag{3}$$

In relevant literature, the optimal filtering problem is defined by the recursion satisfied by the marginal distribution $p(x_t \left| y_{1:t} \right.)$, then we have an equation known as the updating step

$$p(x_t \left| y_{1:t}, \theta \right.) = \frac{p(y_t \left| x_t, \theta \right.) p(x_t \left| y_{1:t-1}, \theta \right.)}{p(y_t \left| y_{1:t-1}, \theta \right.)},. \tag{4}$$

and the prediction step

$$p(x_t \left| y_{1:t-1}, \theta \right.) = \int p(x_t \left| x_{t-1}, \theta \right.) p(x_{t-1} \left| y_{1:t-1}, \theta \right.) dx_{t-1}. \tag{5}$$

However simple it may be, the recursion formula in equation (4) rarely admits a closed form expression (notably the linear-Gaussian case, which leads to the Kalman filter). In general, it is necessary to employ approximations. In this paper, we consider the application of the Sequential Monte Carlo (SMC) methods, particularly particle filter (PF). Since their

introduction by Gordon, Salmond and Smith in 1993, particle filtering methods have become a very popular class of algorithms to solve these estimation problems numerically in an online manner. Our attention is focused on an algorithm known as the auxiliary particle filter (APF).

2. Methodology - the idea of particle filter and auxiliary particle filter

The basic variants of PF are the extended version of the Sequential Importance Sampling (SIS) algorithm with an added resampling step (known as the Sequential Importance Resampling, SIR). Resampling is necessary to keep as many samples as possible with non-zero weights. The idea behind it is to reduce the degeneracy problem and increase speeds of PF, for details see [2], [5], [8]. Through PF in particular, one obtains the Monte Carlo approximation of the filtering distribution, which is an empirical distribution formed from a set of random samples (known as particles) with associated weights

$$\hat{p}_N(x_{1:t}|y_{1:t}) = \sum_{i=1}^{N} w_t^{(i)} \delta(x_{1:t} - x_{1:t}^{(i)}), \tag{6}$$

where $\delta(\cdot)$ is the Dirac delta function, and $w_t^{(i)}$ denotes the normalized importance weight attached to particle $x_t^{(i)}$, details can be found in [1], [3], [5], [6]. The importance weight is calculated from the formula:

$$w_t^{(i)} = w_t^{(i)}(x_{1:t}^{(i)}) \propto \frac{p(x_{1:t}^{(i)}|y_{1:t})}{q(x_{1:t}^{(i)}|y_{1:t})}$$
$$\propto \begin{cases} w_{t-1}^{(i)} p(y_t|x_t^{(i)}) & dla \quad q(x_t|x_{1:t-1}^{(i)}, y_{1:t}) \overset{1}{=} p(x_t|x_{t-1}^{(i)}) \\ w_{t-1}^{(i)} p(y_t|x_{t-1}^{(i)}) & dla \quad q(x_t|x_{1:t-1}^{(i)}, y_{1:t}) \overset{2}{=} p(x_t|x_{t-1}^{(i)}, y_t) \end{cases}. \tag{7}$$

where $q(\cdot)$ is the importance sampling function.

The choice of an importance density is one of the critical elements of the method that affects its efficiency. Note that depending on the form of the importance density function $q(\cdot)$, in literature we can distinguish two versions: the first denoted in formula (7), where "1" is known as a prior kernel, and the second "2" is the optimal importance density [2]. It is worth noting that on the one hand, the optimal importance function limits the degeneracy algorithm,

[2] By optimal we understand such a function which minimizes the variance of the importance weights conditional upon the simulated trajectory $x_{1:t-1}^{(i)}$ and observations $y_{1:t}$, i.e. $\mathrm{var}_{q(x_t|x_{1:t-1}^{(i)}, y_{1:t})}[w_t^{(i)}] = 0$.

inasmuch as it takes into account the information about the current observation. But on the other, it suffers from two major drawbacks: it requires the ability to sample from $p(x_t|x_{t-1}^{(i)}, y_t)$ and to evaluate up to $p(y_t|x_{t-1}^{(i)})$, which in general does not occur in analytical form.

Literature features various methods of PF optimization which mainly consist in applying selected suboptimal function approximation methods, see for example [3], [6], or some modifications of the resampling procedure [2], [8], [9] and the references therein.

In this paper, we present a new idea for the construction of APF, which for a suitable weighing of simulated particles involves the use of Pearson curves (PC) for an approximation of $p(y_t|x_{t-1}^{(i)})$. The idea of APF algorithm is strongly associated with the fact that when we use the optimal importance function, the weight at time t does not depend on the state x_t. Therefore, it seems wasteful to resample particles at the end of iteration $t-1$ prior to looking at y_t, so it is proposed to employ knowledge about the next observation before resampling to ensure that particles are compatible with that observation. APF was first described by Pitt and Shephard [10], Carpenter, Clifford and Fearnhead [4]. Despite the above dissimilarity (7), it can be shown that the APF proposed in [4] may be interpreted as SIR with a different choice of importance weights. Having selected the appropriate importance function $q(x_t|x_{t-1}, y_t)$ and resampling strategy the considered APF algorithm can be summarised as the following pseudo code:

APF algorithm

1. At $t=1$ for $i=1,...,N$ do sample $x_1^{(i)} \sim q(\cdot|y_1)$ set $w_1^{(i)} \propto \dfrac{p(y_1|x_1^{(i)})\,p(x_1^{(i)})}{q(x_1^{(i)})}$ end for
2. For $t \geq 2$, for $i=1,...,N$ do

   set $\tilde{w}_{t-1}^{(i)} \propto w_{t-1}^{(i)} \tilde{p}(y_t|x_{t-1}^{(i)})$, $RESAMPLE\{x_{t-1}^{(i)}, \tilde{w}_{t-1}^{(i)}\} \rightarrow \{\tilde{x}_{t-1}^{(i)}, \dfrac{1}{N}\}$, set $x_{1:t-1}^{(i)} = \tilde{x}_{1:t-1}^{(i)}$ end for

   for $i=1,...,N$ do

   sample $x_t^{(i)} \sim q(x_t|x_{t-1}^{(i)}, y_t)$, set $w_t^{(i)} \propto \dfrac{p(y_t|x_t^{(i)})\,p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{t-1}^{(i)}, y_t)\,\tilde{p}(y_t|x_{t-1}^{(i)})}$, end for.

It is worth noting that by employing the above algorithm, we can obtain the approximation $\hat{p}(x_{1:t}|y_{1:t+1}) \propto p(x_{1:t}|y_{1:t})\hat{p}(y_{t+1}|x_t)$. However, $p(x_{1:t}|y_{1:t})$ is not approximated directly and the importance sampling is proposed

$$\hat{p}(x_{1:t+1}|y_{1:t+1}) \propto \hat{p}(x_{1:t}|y_{1:t})\tilde{p}(y_{t+1}|x_t)q(x_{t+1}|x_t, y_{t+1}),. \tag{8}$$

where $\tilde{p}(y_{t+1}|x_t)$ is the approximation of function $p(y_{t+1}|x_t)$.

3. Simulations and results

The performance of the method under discussion is demonstrated through a simulated standard stochastic volatility model SV with uncorrelated measurement. We assume that y_t is the observed return, x_t the unobserved log-volatility. In terms of proceeding sequentially, we identify latent variables and observations by conditional distributions

$$p(x_{t+1}|x_t, \theta) = \frac{1}{\sqrt{2\pi}\,\sigma_\eta}\exp\left(-\frac{(x_{t+1}-\alpha-\phi x_t)^2}{2\sigma_\eta^2}\right), \tag{9}$$

$$p(y_t|x_t) = \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{y_t^2}{\exp(x_t)}+x_t\right)\right], \tag{10}$$

where ϕ is the persistence of volatility process that allows for volatility clustering, α is interpreted as the modal volatility, σ^2 is the volatility of the volatility factor, and $\theta = [\alpha, \phi, \sigma^2]$ is the parameter vector.

In this paper, our main goal is to discuss the possibility of improving the PF method by applying PC technique for approximation of functions $p(y_t|x_{t-1})$. Pearson Curve can be obtained as functions of the shape parameters: skewness $\beta_1 = \frac{\mu_3^2}{\mu_2^2}$ and kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2}$, (which are tabulated), where μ_k is a kth central moment $\mu_k = E((y_t - E(y_t|x_{t-1}))^k|x_{t-1})$, $k = 2,3,4$. We can compute the explicit form of the mentioned moments:

$$E[y_t|x_{t-1}] = E\left[exp\left(\frac{\alpha+\phi x_{t-1}+\sigma\eta_t}{2}\varepsilon_t\right)\middle| x_{t-1}\right] = 0, \tag{11}$$

$$\mu_2 = Var\left[y_t \middle| x_{t-1}\right] = \exp(\alpha + \phi x_{t-1} + 0.5\sigma^2), \tag{12}$$

$$\mu_3 = 0, \tag{13}$$

$$\mu_4 = E\left[y_t^4 \middle| x_{t-1}\right] = \exp(2\alpha + 2\phi x_{t-1})\exp(2\sigma^2). \tag{14}$$

Accordingly, we can conclude that $p(y_t|x_{t-1})$ is a symmetric function, described by kurtosis $\beta_2 = \exp(\sigma^2)$ which depends on σ^2. It can be easily checked that the function under consideration takes the form of Pearson type II curve (if $\beta_2 > 3$), type VII (for $\beta_2 < 3$) or Gaussian curve (if $\beta_2 = 3$). A thorough theoretical analysis of APF for its effectiveness shows that we should select a function $\tilde{p}(y_t|x_{t-1})$ with thicker tails than $p(y_t|x_{t-1})$ (so that the importance weighs should be upper bounded). Therefore, we assume that $\tilde{p}(y_t|x_{t-1})$ is a Pearson type VII distribution with shape parameter m and scale parameter a, defined by the density

$$f_{a,m}^{\ VII}(y) = \frac{2\Gamma(m)}{a\sqrt{\pi}\Gamma\left(m-\frac{1}{2}\right)}\left(1+\frac{y^2}{a^2}\right)^{-m} \mathrm{I}_{[0,\infty)}(y). \tag{15}$$

where $m = \frac{5\beta_2 - 9}{2\beta_2 - 6}$, $a = \sqrt{\frac{2\mu_2\beta_2}{\beta_2 - 3}}$, $\Gamma(\cdot)$ is the gamma function. Additionally, due to the fact that the observations can take both a positive and negative value, it is necessary to extend the function (15) for the negative axis. Assuming that they appear equally often, we consider a combination of PC VII defined on the whole of the real line

$$f_{a_-,m_-,a_+,m_+}^{\ VII}(y) = \frac{1}{2}\left(f_{a_-,m_-}^{\ VII}(y)\mathrm{I}_{(-\infty,0)}(y) + f_{a_+,m_+}^{\ VII}(y)\mathrm{I}_{[0,\infty)}(y)\right), \tag{16}$$

where the parameters (a_-, m_-), (a_+, m_+) are determined separately for sets: $\{-y_t : y_t < 0\}_t$ and $\{y_t : y_t > 0\}$.

In order to investigate the performance of the proposed algorithm we evaluate its effectiveness by the Root Mean Squared Error (RMSE) defined as:

$$RMSE[t] = \left(t^{-1} \sum_{k=1}^{t} \left(x_k - \hat{x}_k \right)^2 \right)^{1/2}, \tag{17}$$

which measures the distance between the true x_t and filtered series $\hat{x}_t$, where $\hat{x}_t = E[x_t | y_{1:t}] = \sum_{i=1}^{N} w_t^{(i)} x_t^{(i)}$.

In presented research, we assume that parameters α, ϕ are constant, particularly $\alpha = 0,3$, $\phi = 0,8$, while manipulating the level of the σ (most filters are sensitive to the size of the disturbance which is exposed to hidden variable). For this we will respectively denote models M1, M2, M3 for $\sigma = 0,5$, $\sigma = 1, \sigma = 2$. Additionally, we compare the use of proposed APF strategies (APF_PC) with the well known SIR filter, the APF proposed in [9] (APF_P). Simulation results for the proposed technique are presented on the assumption that the sampling distribution q is a Gaussian approximation of p as described in [3].

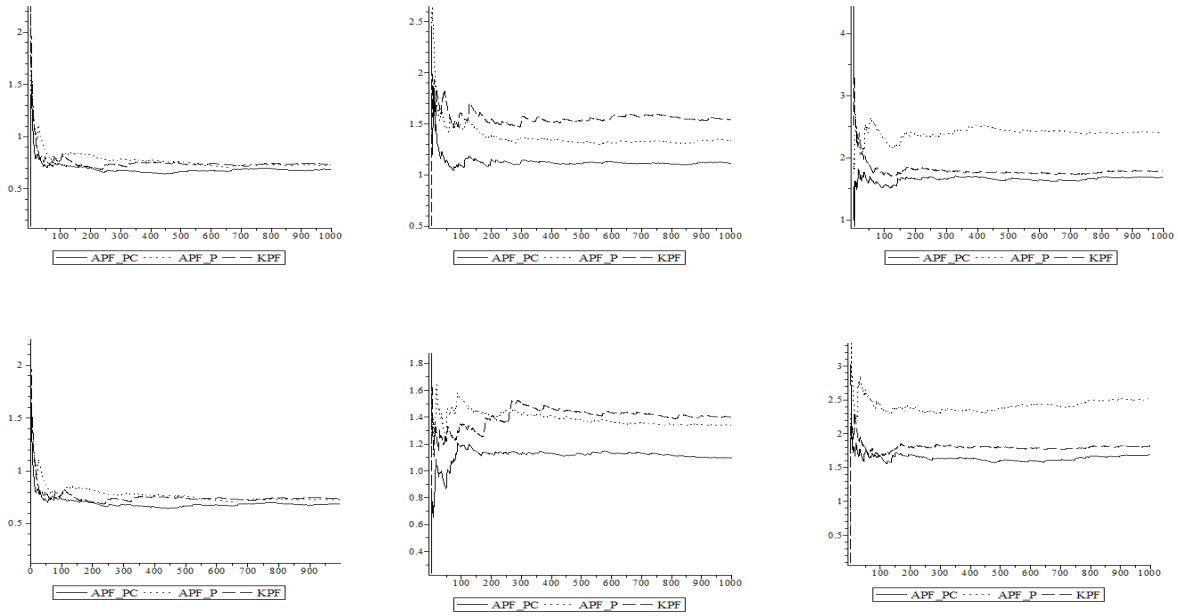

Fig. 1. Plot of RMSE for SIR and APF technique (from left to right M1, M2, M3).

Fig. 1 shows the RMSE values that are computed using simulated time series with length $T = 1000$ and two different numbers of particles $N = 1000$ (the first row in the above frame), $N = 10000$ (second row). The presented results show that the proposed technique, regardless of the model, outperforms the conventional particle filter.

4. Conclusions

Various techniques have been suggested to improve filtering algorithms for nonlinear models. In this paper, we have introduced APF algorithm combined with Pearson curves technique. We have demonstrated that it is possible to develop APF approximation which performs well for nonlinear, non-Gaussian series observations, which are fairly common in financial time series. Additionally, our modification, outlined above, makes APF straightforward and quick to implement. A more extensive analysis of the problem will be presented in the extended version of this article.

Acknowledgements

The results presented are part of my PhD thesis written under the aegis of Professor Antoni Leon Dawidowicz, to whom I would like to express sincere gratitude for scientific supervision and friendly support throughout. Research supported by a grant from the Polish Ministry of Science and Higher Education NN111 632040.

References

[1] Arulampalam, M.S., Maskell, S., Gordon, N., Clapp T., 2002. A tutorial on particle filters for online nonlinear/ non-Gaussian Bayesian Tracking, IEEE Trans. Signal Processing 50(2), 174-188.

[2] Bolić, M., Djurić, P.M., Hong, S., 2004. Resampling algorithms for particle filters: a computational perspective. EURASIP Journal on Applied Signal Processing 15, 2267-2277.

[3] Brzozowska-Rup, K., Dawidowicz, A.L., 2009. Metoda filtru cząsteczkowego. Matematyka Stosowana: matematyka dla społeczeństwa. 10, 69-107.

[4] Carpenter, J., Clifford, P., Fearnhead, P., 1999. An Improved Particle Filter for Nonlinear Problems. IEE Proceedings – Radar, Sonar and Navigation 146, 2-7.

[5] Doucet, A., De Freitas, J.F.G., Gordon, N.J., 2001. Sequential Monte Carlo Methods in Practice. New York: Springer-Verlag.

[6] Doucet, A., Johansen, A.M., 2008. A tutorial on particle filtering and smoothing: fifteen years later. Technical report. Department of Statistics, University of British Columbia.

[7] Gordon, N.J., Salmond, D.J. and Smith, A.F.M., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEEE Proceedings F on Radar and Signal Processing 140(2), 107-113.

[8] Hol, J.D., Schön, T.B., Gustafsson, F., 2006. On resampling algorithms for particle filtering. Proc. IEEE Nonlinear Statistical Signal Processing Workshop, 79-82.

[9] Johansen, A.M., Doucet, A., 2008. A note on the auxiliary particle filter. Statistics and Probability Letters 78(12), 1498-1504.

[10] Pitt, M.K., Shephard, N., 1999. Filtering via simulation: auxiliary particle filters. Journal of the American Statistical Association 94(446), 590-599.