Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/100303

A New Approach to the Construction of the APF Algorithm by Applying the Pearson Curves Technique

Katarzyna Brzozowska-Rup[1] *and Antoni Leon Dawidowicz*[2,*]

[1] Faculty of Management and Computer Modelling, Kielce University of Technology, Poland
[2] Faculty of Mathematics and Computer Science, Jagiellonian University, Kraków, Poland

Received: 4 Oct. 2015, Revised: 4 Jan. 2016, Accepted: 5 Jan. 2016
Published online: 1 May 2016

Abstract: We consider the theoretical question concerning time series which arises when the distribution of the observed variable is in fact a conditional distribution. The Kalman filter provides an effective solution to the linear Gaussian filtering problem. However, when state or measurement, or both, 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. The Sequential Monte Carlo method (SMC) have become one of the familiar tools that allowed the Bayesian paradigm to be applied to approximation of sophisticated models. In this paper we propose a novel construction of an auxiliary particle filter (APF) algorithm using the Pearson curves technique (PC) 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: Sequential Monte Carlo methods; state-space models; stochastic volatility process SV; Pearson's curves technique

1 Introduction

In practice, many problems in stochastic dynamical systems require an estimation of the state of a system changing over time using a sequence of noisy measurements made on the systems. The article concerns the nonlinear filtering problems which appear in many diverse fields including economics, statistical signal and engineering.

Optimal filtering consists in a recursive estimation of the sequence of posterior densities $\{p(x_t|y_{1:t})\}_{t>0}$, which summarizes all the information about the system states $X_{1:t}$ under the assumption that the observations $Y_{1:t}$ are available. Where, for any process $\{Z_t\}_{t\geq 1}$ the realizations from time $t = i$ to $t = j$ are denoted as: $z_{i:j} = (z_i, \ldots, z_j)$. Complex models often lead to integrals that cannot be solved analytically, therefore it is advisable to use Monte Carlo approximation methods. Recently, for the estimation of multiple distributions, the Markov Chain Monte Carlo (MCMC) methods have been preferred. However, MCMC methods require a complete "browsing" of the observations set, which is why they do not make the right strategy for sequential estimation. MCMC- based algorithms are very time consuming when performing online estimations. In this paper, we propose a different approach which is based upon the Sequential Monte Carlo (SMC) methods - the technique known as the particle filtering (PF). This method is becoming increasingly popular in economics and finance as an alternative to MCMC methods. The particle filter is a set of Monte Carlo schemes that enable Kalman-type recursions when normality or linearity, or both, are abandoned. The PF methods are sampling algorithms which combine importance sampling and resampling schemes. Although it has been two decades since PF first appeared in Gordon, Salmond and Smith [11], and immense literature can be found on their theory, it still represents an area of active research. The standard reference for SMC methods is Doucet, De Freitas and Gordon [10], and Cappé, Godsill, Moulines [3], Arulampalam, Maskell, Gordon, Clapp [1], Doucet, Johansen [11], Särkkä [23] for recent reviews. We propose a new construction of an auxiliary particle filter (APF) algorithm using the Pearson curves technique for the approximation of importance weights of simulated particles. The algorithm is verified against the particle filter in application to the stochastic volatility process SV. Numerical results show that the proposed approximation

* Corresponding author e-mail: Antoni.Leon.Dawidowicz@im.uj.edu.pl

proves far more accurate than the basic particle filter (known as the bootstrap filter). All simulations and calculations were carried out on a 1.6 GHz Pentium 4CPU using the author's software written in Visual Studio.

2 A Bayesian approach to state-space modelling

Although the methodology can be applied to more complex systems, in the following research we consider a discrete state space model (DSSM) with a first-order Markov state process and conditionally independent observations (sometimes termed as hidden Markov models, HMM). 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), and define the following model: X_t - a hidden (latent) state process as a stationary and ergodic Markov process, characterized by invariant initial probability density of state

$$X_1 \sim p_0(x_1), \tag{1}$$

and Markov transition probability distribution $p(x'|x)$, for $t > 1$:

$$X_t|(X_{1:t-1} = x_{1:t-1}, Y_{1:t-1} = y_{1:t-1}) \sim p(x_t|x_{t-1}). \tag{2}$$

As indicated by its name, X_t is observed not directly but through another process $\{Y_t\}_{t\in\mathbb{N}}$. The observations are assumed to be conditionally independent when given X_t, and their common marginal probability distribution is expressed as follows:

$$Y_t|(X_{1:t} = x_{1:t}, Y_{1:t-1} = y_{1:t-1}, \theta) \sim p(y_t|x_t). \tag{3}$$

Depending on context, p will denote a probability distribution or a probability density function.
The sequential inference on the latent process X_t is typically based on the sequence of posterior distributions $p(x_{1:t}|y_{1:t})$, where 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:

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

In relevant literature, the optimal filtering problem is defined by a recursion satisfied by the marginal distribution $p(x_t|y_{1:t})$

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})}, \tag{5}$$

which is known as the updating step. The formula (5) employs the dynamics of the model defined earlier (2), (3) and a one-step ahead forecast distribution:

$$p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1}. \tag{6}$$

The likelihood function for a single observation is given by the formula:

$$p(y_t|y_{1:t-1}) = \int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t. \tag{7}$$

In this research, it is assumed that density functions: $p(y_t|x_t)$, $p(x_t|x_{t-1})$ are known and $p(y_t|y_{1:t-1}) > 0$. Despite its apparent simplicity, the recursion formula in equation (5) rarely admits a closed-form expression (notably the linear-Gaussian case, which leads to the Kalman filter), hence it is necessary to employ approximations.

3 The idea of particle filter and auxiliary particle filter

It is worth noting that 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). Therefore, the two most important elements of the method which determine its effectiveness are; the way of selecting an importance sampling distribution (IS) (also referred to as the proposal distribution, or the importance function, IF) and the property of resampling procedure.
Recall that the idea of SIS is to propose an importance function $q(\cdot)$ which: should resemble $p(\cdot|y_{1:t})$ as much as possible, is easy to sample from and fulfills the implication $p(\cdot|y_{1:t}) > 0 \implies q(\cdot) > 0$.
In literature we can distinguish between two versions of IS: the first, where the importance function is taken to be the transition density (approach known as the kernel density), i.e.:

$$q(x_t|x_{1:t-1}, y_{1:t}) = p(x_t|x_{t-1}), \tag{8}$$

and the second, known as the optimal IS, where in order to make the method more effective the definition incorporates both the state and observation processes:

$$q(x_t|x_{1:t-1}, y_{1:t}) = p(x_t|x_{t-1}, y_t). \tag{9}$$

Detailed derivatives of the importance distribution can be found in the references section. It is worth noting that the importance sampling density determines the form of the importance weights which are used for required density approximation:

$$w_t^{(i)} = w(x_{1:t}^{(i)}) \propto \frac{p(x_{1:t}|y_{1:t})}{q(x_{1:t}|y_{1:t})}$$
$$\propto \begin{cases} w_{t-1}(x_{1:t-1}^{(i)})p(y_t|x_t^{(i)}) \text{ dla } q(x_t|x_{1:t-1}, y_t) \overset{1}{=} p(x_t|x_{t-1}) \\ w_{t-1}(x_{1:t-1}^{(i)})p(y_t|x_{t-1}^{(i)}) \text{ dla } q(x_t|x_{t-1}, y_t) \overset{2}{=} p(x_t|x_{t-1}, y_t) \end{cases} \tag{10}$$

In the paper we assume that the weights $w_t^{(i)}$, $t = 1,\ldots,T$, $i = 1,\ldots,N$ are determined up to a multiplicative constant and $\sum_{i=1}^{N} w_t^{(i)} = 1$.
By optimal IS we understand 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.

$$\mathrm{var}_{q(x_t|x_{1:t-1}^{(i)},y_{1:t})}[w_t^{(i)}] = 0. \tag{11}$$

It is worth pointing out that on the one hand, the optimal importance function limits the degeneracy algorithm, 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 evaluate $p(y_t|x_{t-1}^{(i)})$ (which, in general, do not occur in an analytical expression). In the case where the formula (9) cannot be expressed analytically, a suboptimal approach is proposed, see Doucet et al. [10] (we will discuss it in detail later).
In literature, it is a well-known fact that importance sampling method (and sequential importance sampling) suffers from severe drawbacks. One of its most important shortcomings is a sample impoverishment. As the variance of the unnormalized importance weights tends to increase exponentially with time, it results in a potential degeneracy of the approximation. In practice this means that after a few time steps, a very small subset of particles takes all the probability mass, which implies that a discrete measure approximation of the filtering density becomes invalid and useless. The degeneracy is routinely measured using the effective sample size criterion $\hat{N}_{ESS} = \left(\sum_{i=1}^{N}(w_t^{(i)})^2\right)^{-1}$. The $\hat{N}_{ESS}$ below a pre-specified threshold (typically $N_T = 0.5N$) implies that the degeneracy is too high. To prevent the degeneracy problem, it is recommended to re-draw the generated samples (known as a resampling step). The idea of resampling is very intuitive and relies on replicating the particles with higher weights and discarding those with negligible weights. The most common resampling schemes are discussed for example in Bolić, Djurić, Hong [2].
In this paper we make use of the residual systematic resampling (RSR), which can be implemented by using the following steps:

Algorithm 1: RSR

1. Sample $U^{(0)} \sim \mathrm{Unif}\left((0,\frac{1}{N}]\right)$
2. For i from 1 to N calculate
   $N_t^{(i)} = \left\lfloor (w_t^{(i)} - U^{(i-1)})N \right\rfloor + 1$,
   $U^{(i)} = U^{(i-1)} + N_t^{(i)}N^{-1} - w_t^{(i)}$
   The resampled particles are
   $x_{t|t}^{(i)} = x_{t|t-1}^{(N_t^{(i)})}$, $w_t^{(i)} = N^{-1}$
   end for

It is worth noting that the algorithm 1 is substantially more efficient than the most popular multinomial. Its processing time is reduced, whereby it is very attractive for hardware implementation, for details see: Bolić et al. [2], Hol, Schön, Gustafsson [13]. Now, we can present the simplest case of particle filtering algorithm, also known as the bootstrap filter (algorirhm 2), see Gordon et al. [12], Doucet et al. [10,11]:

Algorithm 2: Bootstrap filter, BPF, [10]

Put $t = 1$
For i from 1 to N
– Sample $x_{1|1}^{(i)} \sim q_1(\cdot)$,
– Compute the particle weights
$w_1^{(i)} = \frac{p_1(x_{1|1}^{(i)})p(y_1|x_{1|1}^{(i)})}{q_1(x_{1|1}^{(i)})}$
end for
For t from 2 to T
For i from 1 to N
– Sample $x_{t|t-1}^{(i)} \sim q(x_t^{(i)}|x_{t-1|t-1}^{(i)})$;
set $\tilde{x}_{1:t|t-1}^{(i)} = \left\{x_{1:t-1|t-1}^{(i)}, x_{t|t-1}^{(i)}\right\}$;
– Compute the importance weights
$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(y_t|x_{t|t-1}^{(i)})p(x_{t|t-1}^{(i)}|x_{t-1|t-1}^{(i)})}{q(x_{t|t-1}|x_{t-1|t-1}^{(i)},y_t)}$
end for
Iff $\hat{N}_{ESS} < N_T$, then
Resample $\left\{x_{1:t|t-1}^{(i)}, w_t^{(i)}\right\} \to \left\{x_{1:t|t}^{(i)}, \frac{1}{N}\right\}$
else $x_{1:t|t}^{(i)} = \tilde{x}_{1:t|t-1}^{(i)}$
set $t = t+1$ end for

Through PF in particular, we obtain 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

$$\widehat{p}(x_{1:t}|y_{1:t},\theta) = \sum_{i=1}^{N} w_t^{(i)} \delta(x_{1:t} - x_{1:t|t-1}^{(i)}). \tag{12}$$

More formally (after the resampling step), the weighted distribution (12) is replaced by the unweighted measure

$$\widetilde{p}(x_{1:t}|y_{1:t},\theta) = \frac{1}{N}\sum_{i=1}^{N} \delta(x_{1:t} - x_{1:t|t}^{(i)}), \tag{13}$$

where $\delta(\cdot)$ is the Dirac delta function, and $w_t^{(i)}$ denotes the normalized importance weight attached to particle $x_t^{(i)}$.
The random measure (12) can be used to approximate any integral. As an example we can consider an expected value of an integrable real function φ:

$$\hat{p}_N(\varphi) = \sum_{i=1}^{N} w_t^{(i)} \varphi(x_{1:t|t-1}^{(i)}). \tag{14}$$

According to the strong law of large numbers (under mild assumptions)

$$\hat{p}_N(\varphi) \xrightarrow[N\to\infty]{a.s} p(\varphi), \tag{15}$$

where a.s. denotes an almost sure convergence. Details of this basic proposition and its proof can be found, among others, in Doucet et al. [10].
Literature features various methods of PF optimization which mainly consist in applying selected suboptimal function approximation methods, or some modifications of the resampling procedure, for example: Johansen, Doucet Prokhorov, Yu [15], Douc, Moulines, Olsson [7], Doucet et al. [11], Del Moral, Doucet, Jasra [4].
The modified construction of APF we propose consists in the use of Pearson curves (PC) for an approximation of simulated particles weights. The idea of APF is strongly associated with the fact that when we use the optimal importance function (9), the weight (10) (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 considering y_t. Instead, it is proposed to employ the knowledge about the next observation before resampling to ensure that particles are compatible with that observation. APF was first described by Pitt and Shephard [19], and almost simultaneously by Carpenter, Clifford and Fearnhead [8]. It can be shown that the APF proposed in Carpenter et al. [8], may be interpreted as SIR with a target function $p(x_{1:t}|y_{1:t+1})$. Having selected the appropriate importance function $q(x_t|x_{t-1},y_t)$ and the resampling strategy, we can now summarize the considered APF algorithm as the pseudo code described below in Algorithm 3.
It is worth noting that by employing the following algorithm, we can obtain the approximation:

$$\tilde{p}(x_{1:t}|y_{1:t+1}) \propto p(x_{1:t}|y_{1:t})\tilde{p}(y_{t+1}|x_t), \tag{16}$$

where $\tilde{p}(y_{t+1}|x_t)$ is the approximation of function $p(y_{t+1}|x_t)$. However, $p(x_{1:t}|y_{1:t})$ is not approximated directly, hence it is necessary to modify the associated importance weights, see Doucet et al. [11] for details.

Algorithm 3: Auxiliary particle filter, APF

1. Put $t=1$
   For i from 1 to N
   – Sample $x_{1|0}^{(i)} \sim q_1(\cdot)$,
   – Compute the particle weights $w_1^{(i)}$ as in BPF end for
2. For t from 2 to T
   For i from 1 to N
   – set $\tilde{w}_{t-1}^{(i)} \propto w_{t-1}^{(i)} \tilde{p}(y_t|x_{t-1|t-2}^{(i)})$
   Resample $\{x_{1:t-1|t-2}^{(i)}, \tilde{w}_{t-1}^{(i)}\} \to \left\{x_{1:t-1|t-1}^{(i)}, \frac{1}{N}\right\}$;
   end for

   – For i from 1 to N
   sample $x_{t|t-1}^{(i)} \sim q(x_t^{(i)}|x_{t-1|t-1}^{(i)}, y_t)$
   set $w_t^{(i)} \propto \frac{p(y_t|x_{t|t-1}^{(i)})p(x_{t|t-1}^{(i)}|x_{t-1|t-1}^{(i)})\tilde{p}(y_{t+1}|x_{t|t-1}^{(i)})}{q(x_{t|t-1}|x_{t-1|t-1}^{(i)}, y_t)\tilde{p}(y_t|x_{t-1|t-1}^{(i)})}$
   end for
   end for

4 Simulations and results

The performance of the method under discussion is demonstrated through a simulated standard stochastic volatility model SV with uncorrelated measurement. The SV model is the alternative to the Autoregressive Conditional Heteroscedasticy (ARCH) and GARCH type processes which assumes two error processes; SV allows the variance of the returns to be an unobserved random process. This implies that the SV models can be more flexible than the ARCH - type models in fitting the data. It allows modelling two of the main features of the financial time series i.e. time varying volatility and clustering phenomena in volatility. The first SV model appeared in 1973 (attributed to Clark); in its most simple continuous form was proposed by Taylor (1986). During the last two decades extensive research has been carried out on this type of models. Many extensions to the SV models have been proposed in the literature. For a comprehensive discussion on the models see, for example Casarin[9], Liesenfeld, Richard [17], Shephard [21], Shephard and Andersen [22], and Tsyplakov [25].
We define the model assuming that y_t is the observed return, and x_t the unobserved log-volatility and in order to proceed sequentially, identify latent variables and observations by conditional distributions

$$p(x_{t+1}|x_t,\theta) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp\left(-\frac{(x_{t+1}-\alpha-\phi x_t)^2}{2\sigma_\eta^2}\right), \tag{17}$$

$$p(y_t|x_t,\theta)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{y_t^2}{\exp(x_t)}+x_t\right)\right), \quad (18)$$

where ϕ is the persistent parameter that allows for volatility clustering ($|\phi|<1$), α is interpreted as the drift (shift) parameter, σ_η^2 is the volatility of the volatility factor, and $\theta=[\alpha,\phi,\sigma_\eta^2]$ is the parameter vector.

To implement APF, we propose using a linearized method (LM) based on the first and second Taylor's series approximation.

In order to simplify the necessary calculations, we introduce the function as the logarithm of the target functions:

$$l(x_t)=\ln p(x_t|x_{t-1},y_t). \quad (19)$$

Assuming that $l(x_t)$ is twice differentiable with respect to x_t, it can be approximated as follows

$$l(x_t)\approx l(x)+[l'(x)]^T(x_t-x)+\frac{1}{2}(x_t-x)^T l''(x)(x_t-x), \quad (20)$$

where $l'(x)=\frac{\partial l(x_t)}{\partial x_t}\Big|_{x_t=x}$, $l''(x)=\frac{\partial^2 l(x_t)}{\partial x_t \partial x_t^T}\Big|_{x_t=x}$, and the point x is chosen arbitrarily. If the function $l(x_t)$ is concave, formula (16) is then equivalent to the following

$$C-\frac{1}{2}(x_t-x-m(x))^T\Sigma^{-1}(x)\,(x_t-x-m(x)), \quad (21)$$

where $\Sigma(x)=-l''(x)^{-1}$, $m(x)=\Sigma(x)l'(x)$, and C is a constant independent of x.

The above considerations as well as the definition of the function illustrate that the importance function is a Gaussian distribution function of known parameters:

$$q(x_t|x_{t-1},y_t)=N\left(m(x)+x,\Sigma(x)\right). \quad (22)$$

For more details, please see Doucet et al.[10].

Another method is the Laplace approximation (LA) technique, which consists in an expansion of the logarithm of $p(x_t|x_{t-1},y_t)$ around a point $x_{t,\max}=\arg\max_{x_t} l(x_t)$. The derived approximation has the following form

$$l(x_t)\approx\frac{\partial^2 l(x_t)}{2\partial x_t^2}\Big|_{x_t=x_{t,\max}}(x_t-x_{t,\max})^2, \quad (23)$$

which resembles the log-kernel of Gaussian density. Consequently, we approximate $q(x_t|x_{t-1},y_t)$ by an unnormalized Gaussian

$$q(x_t|x_{t-1},y_t)=C\exp\left(-\frac{(x-x_{t,\max})^2}{2H_t}\right), \quad (24)$$

where $H_t=\left(\frac{\partial^2 l(x_t)}{2\partial x_t^2}\Big|_{x_t=x_{t,\max}}\right)^{-1}$, and C is normalizing constant.

In this paper, our main goal is to discuss the possibility of improving the PF method by applying a Pearson curves technique (PC) for approximation of functions $p(y_t|x_{t-1})$. The details of the method are placed in the Appendix.

It can be easily checked that the function under consideration is symmetric, and its kurtosis β_2 depends on σ_η^2, ($\beta_2=\exp(\sigma_\eta^2)$). It takes the form of either the Pearson type II curve (if $\beta_2>3$) or type VII (for $\beta_2<3$), or the Gaussian curve (if $\beta_2=3$), see Johanson, Nixon and Amos [14]. 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 weights should be upper bounded). Therefore, we assume that $\tilde{p}(y_t|x_{t-1})$ is a Pearson type VII distribution with a shape parameter m and a scale parameter a, defined by the density

$$f_{a,m}^{VII}(y)=\frac{2\Gamma(m)}{a\sqrt{\pi}\Gamma(m-0.5)}\left(1+\frac{y^2}{a^2}\right)^{-m}I_{[0,\infty)}(y), \quad (25)$$

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 assume both positive and negative values, it is necessary to extend the function in (4.9) to the negative axis. Assuming that they appear equally often, we consider a combination of PC VII defined on the whole of the real line

$$f_{a_-,a_+,m_-,m_+}^{VII}(y)=\frac{1}{2}\left(f_{a,m}^{VII}(y)I_{(-\infty,0)}(y)+f_{a,m}^{VII}(y)I_{[0,\infty)}(y)\right), \quad (26)$$

where the parameters (a_-,m_-), (a_+,m_+) are determined separately for negative and additive sets of $\{y_t\}_t$.

Figure 1 indicates how the auxiliary particle filter combines with the PC technique approximates the true value of the state variable of the considered SV model.

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

$$RMSE[t]=\left(t^{-1}\sum_{k=1}^{t}(x_t-\hat{x}_{t|t})\right)^{\frac{1}{2}}, \quad (27)$$

which measures the distance between the true x_t and the filtered series $\hat{x}_t$, where

$$\hat{x}_t=E[x_t|y_{1:t}]=\sum_{i=1}^{N}w_t^{(i)}x_t^{(i)}. \quad (28)$$

The presented theory is justified by several computer simulations. In these simulations we assume that parameters α, ϕ are constant ($\alpha=0.3$, $\phi=0.8$) while manipulating the level of the σ (most filters are sensitive to the size of the disturbance which is exposed to hidden variables). For this we denote models M1, M2, M3 for $\sigma=0.5$, $\sigma=1$, $\sigma=2$, respectively. Additionally, we

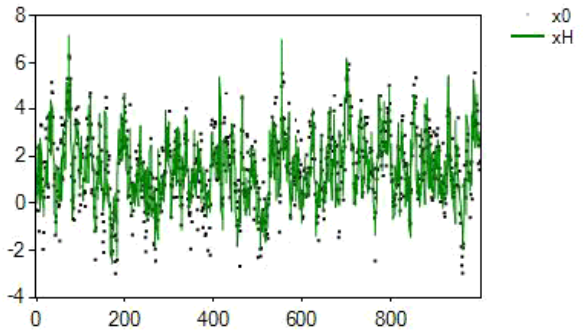

Fig. 1: The results of filtering estimates obtained for the stochastic volatility model by the APF_PC algorithm for the considered SV model (4.1)-(4.2) with parameters $\phi = 0.8$, $\sigma^2 = 1$, $\alpha = 0.3$, $T = 1000$, number of particles $N = 1000$ where x0 is the latent state of simulated process and xH denotes the state of filtering process.

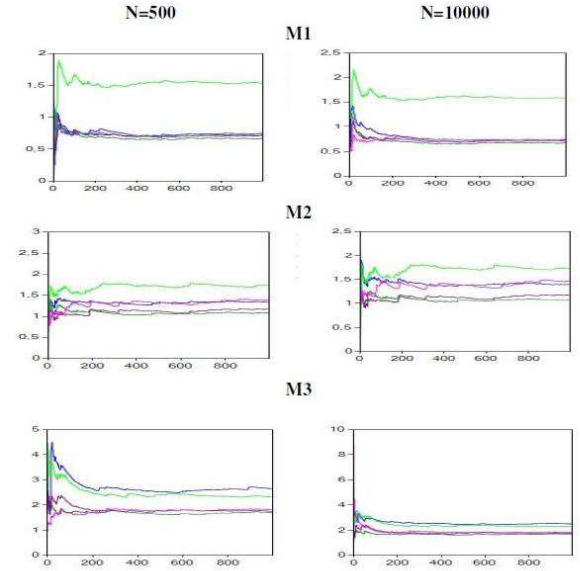

Fig. 2: Results of RMSE for SIR and the APF techniques from top row to bottom M1, M2, M3 respectively, which are computed using simulated time series with length $T = 1000$ and two different numbers of particles $N = 500$ (left colum), $N = 10000$ (right column). In the plot we have used the following colours to denote: pink -KPF, green -APF_LA, violet -APF_LM, navy blue -APF_P, dark green -APF_PC.).

compare the three algorithms: the proposed APF (APF_PC), the well-known SIR filter (denoted as KPF), and the APF described in Pitt et al. [19] (APF_P). Depending on IS, the algorithms are denoted as APF_LM (for the linearized method), APF_LA (when IS is calculated by the Laplace approximation) and APF_PC (when IS is a prior kernel). Simulation results for the proposed technique are presented in Figure 2. The conducted simulations unequivocally illustrate the fact that the PF method with the IS function determined by linearization is burdened with the biggest estimation error. The result is particularly worthy of notice as LM is one of the most frequently proposed methods of determining the importance sampling density in literature. Therefore, in our further analysis, we will deal with a comparison of KPF, APF_P, APF_PC.

Finally Figure 2 shows that the proposed technique, regardless of the model, outperforms the conventional particle filter. This fact is emphesized by Figure 3 where we can see that APF_PC performs better than other methods. Surprisingly, an increased number of generated particles does not noticeably minimize the RMSE error, which is a valuable observation in practical terms. It is interesting to point out that the evaluated RMSE is close to the volatility of underlying state process.

Due to the time-consuming resulting from SMC computational complexity and generally known fact that the Monte Carlo estimation methods require a large number of MC simulations to ensure the desired efficiency of the estimation it is extremely valuable

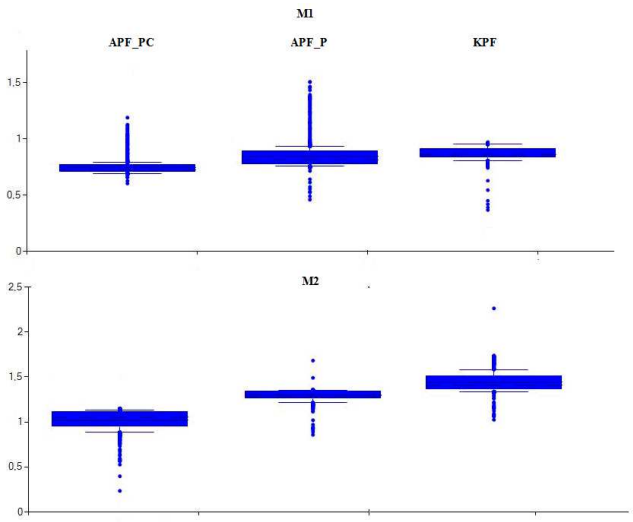

Fig. 3: Box and whiskers plot of RMSE for the APF_PC, APF_P and KPF techniques based on 100 independent realizations. From left to right and from top row to bottom M1, M2, M3 respectively, which are computed using simulated time series with length $T = 1000$ and $N = 500$ numbers of particles.

advantage of the method is that it allows to obtain significantly more accurate estimates of the state variable for a relatively small number of simulations N.

5 Conclusions

The theory of the PF framework emphasizes the important role played by the importance density, but it is clear that the quality of state's estimates improves together with the importance weights approximation efficiency. Thus in order to estimate these importance weights, we have proposed the Pearson curves technique. The performance of the method is tested for three different models of stochastic volatility process, which is fairly common in financial time series analysis. The experimental comparison of the three techniques under consideration demonstrates superior performance of the APF_PC algorithm, regardless of the model. Additionally, our modification, as outlined above, makes APF straightforward and quick to implement. The results we have arrived at confirm the familiar fact that the effectiveness of the PF estimation is heavily dependent on the appropriate choice of the importance function sampling as well as the selection of importance particles (correctness of resampling).
An altogether separate issue in the field of PF is the question of estimation of structural model parameters. It is worth noting that relevant literature distinguishes between two approaches: in the first one parameters are treated as hidden variables (extension of state-space is carried out); see: Liu and West [18] and Polson, Stroud and Müller [20], whereas the second approach makes use of the EM method and particle filtering jointly (EM-PF). Some examples of estimation of structural parameters of a model using EM-PF can be found in Kantas, Doucet, Singh, Maciejowski [16] and Cappé [5]. In the future, it would be interesting to compare and study the efficiency of estimation of our modification of PF combined with online Expectation Maximisation algorithm and the HMM-particle learning method for model parameters.

Appendix: Person curves technique for APF

In this section we demonstrate a Pearson curves technique approximation which relies on the fact that the first four moments are available although the density itself is analytically intractable or unknown. Karl Pearson introduced the system of classification densities by the shape parameters: skewness $\beta_1 = \frac{\mu_3^2}{\mu_2^2}$ and kurtosis, $\beta_2 = \frac{\mu_4}{\mu_2^2}$ (they are tabulated), where $\mu_k = E[(y_t - E[y_t|x_{t-1}])^k|x_{t-1}]$ is a k-th central moment, $k = 2,3,4$.
To provide approximations of $p(y_t|x_{t-1})$) distributions for SV model we rewrite the model as a discrete state space model, see Taylor (1986)

$$x_t = \log \sigma_t^2$$

$$x_1 \sim N\left(\frac{\alpha}{1-\phi}, \frac{\sigma_\eta^2}{1-\phi^2}\right),$$

$$x_{t+1} = \alpha + \phi x_t + \sigma_\eta \eta_t,$$

$$y_t = \exp\left(\frac{x_t}{2}\right)\varepsilon_t,$$

$$\begin{bmatrix}\varepsilon_t \\ \eta_t\end{bmatrix} \sim N\left(0, \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}\right), t = 1,\ldots,T.$$

We can compute the explicit form of the mentioned moments:

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

$$\mu_2 = Var[y_t|x_{t-1}] = \exp(\alpha + \phi x_{t-1} + 0.5\sigma_\eta^2),$$

$$\mu_3 = 0,$$

$$\mu_4 = E[y_t^4|x_{t-1}] = \exp(2\alpha + 2\phi x_{t-1})\exp(2\sigma_\eta^2).$$

Accordingly, we can conclude that $p(y_t|x_{t-1})$ is a symmetric function, described by kurtosis $\beta_2 = \exp(\sigma_\eta^2)$ which depends on σ_η^2 (the variance of a stochastic component of state equation).

Funding

This research was in part supported by a grants from the Polish National Science Centre UMO-2013/11/D/HS4/04014
and 04.0.09.00/2.02.03.01.0007 NCN. ZKEZ.14.002.

Notes

The results presented formed part of PhD thesis of first-noted author written under the aegis of second-noted author.

Part of the presented results were discussed at The 7th Scientific Conference on Modelling and Forecasting of Socio-Economic Phenomena May 7 - 10, 2013, Zakopane, Poland.

References

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

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

[3] Cappé, O, Godsill, S and Moulines, E. An overview of existing methods and recent advances in sequential Monte Carlo. IEEE PProceedings in Signal Processing. 2007; 95(5): 899–924.

[4] Del Moral, PA, Doucet A and Jasra A. On adaptive resampling strategies for seuential Monte Carlo methods. Bernoulli. 2012;18(1): 252–278.

[5] Cappé, O. Online Expectation Maximisation, Chichester (UK): John Wiley & Sons, Ltd; 2011. p. 31–53.

[6] Clark, PK. A subordinated stochastic process model with finite variances for speculative prices. Econometrica. 1973;41: 135–156.

[7] Douc, R, Moulines, E and Olsson, J. Optimality of the Auxiliary Particle Filter. Probability and Mathematical Statistics. 2009;29: 1–28.

[8] Carpenter, J, Clifford, P,and Fearnhead, P. An Improved Particle Filter for Nonlinear Problems. Radar, Sonar and Navigation, IEE Proceedings. 1999;146(1):2–7.

[9] Casarin, R. Bayesian Monte Carlo Filtering for Stochastic Volatility Models . Cahier on CEREMADE N.0415, UNIVERSITY Paris Dauphine; 2004 (working paper).

[10] Doucet, A, De Freitas, JFG and Gordon, NJ. Sequential Monte Carlo Methods in Practice. New York(NY): Springer-Verlag; 2001.

[11] Doucet, A, Johansen, AM A tutorial on particle filtering and smoothing: fifteen years later. In: Dan Crisan and Boris Rozovskii, editors. The Oxford Handbook of Non-linear Filtering. Oxford: Oxford University Press; 2011.

[12] Gordon, NJ, Salmond, DJ and Smith, AFM. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. on Radar and Signal Processing, IEEE Proceedings F. 1993;140(2): 107–113.

[13] Hol, JD, Schön, TB and Gustafsson, F. On resampling algorithms for particle filtering. Proc. IEEE Nonlinear Statistical Signal Processing Workshop 2006, 79-82.

[14] Johanson, NL, Nixon, E, and Amos, DE. Tables of Percentage Points of Pearson curves for given expressed as Standard Measures. Biometrika. 1963:50:459–471.

[15] Johansen, AM, Doucet, A and Prokhorov, Yu. A note on the auxiliary particle filter. Statistics and Probability Letters. 2008;78(12):1498–1504.

[16] Kantas, N, Doucet, A, Singh, SS. and Maciejowski, J.M. An overview of Sequential Monte Carlo Methods for Parameter Estimation in General State- Space Models. Proceedings of SYSID09. IFAC. 2009 France.

[17] Liesenfeld, R, Richard, J-F. Classical and Bayesian Analysis of Univariate and Multivariate Stochastic Volatility Models.Econometric Reviews, 2006;25(2-3): 335–360.

[18] Liu, JS and West, M. Combined parameter and state estimation in simulation-based filtering in Sequential Monte Carlo in Practice. In: Doucet A, de Freitas JFG and Gordon NJ, editors. New York NY: Springer-Verlag; 2001.

[19] Pitt, MK and Shephard, N. Filtering via simulation: auxiliary particle filters. Journal of the American Statistical Association. 1999;94(446): 590–599.

[20] Polson, NG, Stroud, JR. and Mller, P . Particle Filtering with sequential parameter learning. Journal of the Royal Statistical Society: B. 2008;70(2):413–428.

[21] Shephard, N. Stochastic volatility: readings. Advanced texts in Econometrics. Oxford(UK): Oxford University Press; 2005.

[22] Shephard, N and Andersen T. Stochastic Volatility: origins and overview. In: T.G. Andersen TG, Davis RA, Krei J-P, and Mikosch T, editors. Handbook of financial Time Series. Berlin: Springer; 2009. p. 233–254.

[23] Särkkä, S. Bayesian filtering and smoothing. Cambridge(UK): Cambridge University Press; 2013.

[24] Taylor, SJ. Modelling financial time series. Chichester(UK): John Wiley & Sons; 1986.

[25] Tsyplakov, A. Revealing the Arcane: An Introduction to the Art of Stochastic Volatility Models [cited 2010 Sep 28]. Available form: http://ssrn.com/abstract=1684444

Katarzyna Brzozowska-Rup received the M.Sc degree in mathematics at the Jagiellonian University in Krakow, Poland in 2000, the PhD degree in economics at the University of Lodz in 2012. Since 2006 she has worked at the Kielce University of Technology, where since 2012 she has served as an Associate Professor. Her research interests include sequential Monte Carlo methods, stochastic PDEs, applied statistics and econometrics, with her main focus on the modelling of economic time series and non-linear financial time series.

Antoni Leon Dawidowicz Born on 11 September 1952 in Krakow, is an Associate Professor at the Jagiellonian University in Krakow. In 1976 he graduated in mathematics from the Jagiellonian University, where he has been working since. 1990-1994 he served as a Councillor of Kraków, 2001-2007 president of the Polish Tatra Mountains Society. Member of the Polish Naional Group of International Society of Clinical Biostatistics. Since 2011 he has chaired the Kraków branch of the Polish Mathematical Society.