

KATARZYNA BRZOZOWSKA–RUP (Kielce)  
ANTONI LEON DAWIDOWICZ (Kraków)

## Metoda filtru cząsteczkowego

**Streszczenie.** Celem pracy jest zaprezentowanie idei metod filtracji opartych na sekwencyjnej metodzie Monte Carlo, w literaturze nazywanych również metodami filtru cząsteczkowego oraz odniesienia do odpowiedniej literatury. Istotę omawianych algorytmów przedstawiamy rozważając problem estymacji silnie nieliniowych i niegaussowskich modeli przestrzeni stanów. W praktyce bowiem w takich przypadkach dobrze znany i najczęściej wykorzystywany algorytm rozszerzonego filtru Kalmana (ang. *Extended Kalman Filter*, EKF) może wykazywać istotną nieefektywność. Idea filtru cząsteczkowego polega na estymacji rozkładu prawdopodobieństwa rozkładem empirycznym wyznaczonym na podstawie cząsteczek, tzw. „chmury punktów”. Zaimplementowanie algorytmu filtru cząsteczkowego wymaga zasadniczo przeprowadzenia trzech procedur: (1) losowania cząsteczek z odpowiednio dobranej sekwencyjnej funkcji ważności, (2) określenia istotności cząsteczek, (3) powtórnego losowania, tzw. *resampling*. Metody te są coraz bardziej popularne w dziedzinie ekonomii, statystyki, medycynie i teorii sygnałów.

**Słowa kluczowe:** model przestrzeni stanów, ukryte procesy Markowa, filtracja optymalna, sekwencyjne metody Monte Carlo, sekwencyjna funkcja ważności, „re-próbkowanie” (ang. *resampling*).

**1. Wstęp.** W praktyce wiele zagadnień wnioskowania statystycznego, analizy szeregów czasowych, ekonometrii oraz przetwarzania sygnałów wygodnie jest rozważać w postaci modeli przestrzeni stanów. Jeżeli posiadamy pewną wiedzę a priori na temat badanego układu wówczas, jego dynamikę możemy opisać dwoma równaniami. Równanie stanów określane jest rozkładem a priori ukrytego procesu Markowa  $\{x_t\}_{t \in N}$ , a równanie obserwacji  $\{y_t\}_{t \in N}$  warunkowym rozkładem obserwacji  $p(y_t | x_t)$ .

$$\begin{aligned}x_t &\sim p(x_t | x_{t-1}), \\y_t &\sim p(y_t | x_t).\end{aligned}$$

Znajomość wspomnianych rozkładów umożliwia wykorzystanie reguły Bayesa, czyli połączenia rozkładów a priori nieznanymi stanów z prawdopodobieństwem wystąpienia obserwacji (zdarzenia zwanego „skutkiem”). Wszyst-

kie istotne informacje na temat zmiennej stanu  $x_{0:t} = \{x_0, x_1, \dots, x_t\}$  przy założeniu, że znane są obserwacje do chwili  $t$   $y_{1:t} = \{y_1, \dots, y_t\}$ , zawarte są w łącznym rozkładzie a posteriori  $p(x_{0:t} | y_{1:t})$ . Idea metody filtracji polega na estymacji wspomnianego rozkładu  $p(x_{0:t} | y_{1:t})$ , jego charakterystyk oraz rozkładu brzegowego  $p(x_t | y_{1:t})$ . Okazuje się jednak, że o ile zasady wnioskowania bayesowskiego są proste, to w praktyce pojawiają się problemy natury numerycznej. Jedyne w kilku wyjątkowych przypadkach, tj. liniowych modelach gaussowskich oraz ukrytych łańcuchach Markowa określonych na skończonej przestrzeni stanów, badany rozkład a posteriori ma postać analityczną. Wyznaczanie rozkładów a posteriori i a priori związane jest bowiem z obliczaniem wysoko-wymiarowych całek. W przypadku wielowymiarowych przestrzeni zmiennych ukrytych, będących przedmiotem wnioskowania, praktycznie jedynymi metodami obliczeniowymi są metody Monte Carlo. W liniowych modelach gaussowskich problem estymacji rozwiązuje się za pomocą filtru Kalmana. Niestety w sytuacji, gdy model jest nieliniowy i/lub niegaussowski, filtr Kalmana, jak również jego modyfikacje (np. Extended Kalman Filter, Unscented Kalman Filter) mogą być jedynie wykorzystywane do wstępnej aproksymacji [3]. W literaturze jedną z najpopularniejszych metod uzyskiwania próby z łącznego rozkładu a posteriori zmiennych ukrytych jest algorytm Gibbsa [13], [17], [18]. Metoda ta poprzez konstrukcję odpowiedniego (o zadanym rozkładzie granicznym<sup>(1)</sup>) łańcucha Markowa, umożliwia losowanie z rozkładu wielowymiarowego. Po dostatecznie dużej liczbie losowań z warunkowych rozkładów a posteriori, otrzymuje się próbę z łącznego rozkładu a posteriori. W praktyce często rozkład a posteriori musi być aktualizowany w momencie pojawienia się nowych danych (zarówno obserwacji jak i zmiennej ukrytej). Okazuje się, że w takich przypadkach algorytm Gibbsa jest wysoce nieefektywny, co związane jest z koniecznością powtarzania iteracji dla nowego zbioru danych. W takich przypadkach, dość naturalną propozycją jest zastosowanie metod sekwencyjnych, w których rozważa się ciąg rozkładów  $\{\pi_t\}_{t \geq 1}$  zamiast statycznego rozkładu  $\pi$ .

W ostatnich latach wyraźnie wzrasta zainteresowanie sekwencyjnymi metodami Monte Carlo SMC. Od momentu zaproponowania przez Gordona klasycznego algorytmu filtracji (*bootstrap algorithm*) [14] w literaturze pojawiło się wiele prac, w których proponuje się usprawnienia pierwotnego algorytmu, niemniej pewne problemy nadal pozostają nierozwiązane. Naszym zamiarem jest stworzenie aktualnego przewodnika (opartego na literaturze zawartej w bibliografii) na temat algorytmu filtru cząsteczkowego (ang. *Particle Filter*, PF) w kontekście nieliniowych, niegaussowskich, markowowskich

---

<sup>(1)</sup> Do symulacji łańcucha Markowa z zadanego rozkładu często wykorzystuje się algorytm Metropolisa i Hastingsa.

modeli przestrzeni stanów. Mamy również nadzieję, że niniejsza praca przyczyni się do tego, iż grono polskich statystyków i ekonomistów zainteresuje się prezentowaną metodą, szeroko stosowaną w literaturze angielskiej.

W praktyce symulacyjne metody Monte Carlo wykorzystywane są między innymi w ekonomii do modelowania zmienności danych finansowych, medycynie, statystyce, ekonometrii oraz inżynierii (analizie procesów sygnałowych, w lokalizacji robotów mobilnych), szczegóły można znaleźć w pozycjach umieszczonych w bibliografii [10], [14], [16].

**2. Markowski model przestrzeni stanów.** Rozważamy przestrzeń probabilistyczną  $(\Omega, B(\Omega), P)$  oraz przestrzeń  $P(R^{n_x})$  wszystkich miar probabilistycznych na  $B(R^{n_x})$  z topologią słabej zbieżności. Na przestrzeni  $(\Omega, B(\Omega), P)$  definiujemy dwa rzeczywiste procesy stochastyczne: proces  $\{X_t\}_{t \in N}$ , zwany procesem zmiennych ukrytych (sygnałowym) oraz proces  $\{Y_t\}_{t \in N}$ , zwany procesem obserwacji, dla których przestrzenie stanów mają odpowiednio wymiary  $n_x$  i  $n_y$ . Realizację omawianych procesów od chwili  $i$  do  $j$  będziemy odpowiednio oznaczać  $x_{i:j} = (x_i, \dots, x_j)$ ,  $y_{i:j} = (y_i, \dots, y_j)$ . W dalszej części pracy będziemy głównie rozważać funkcje gęstości prawdopodobieństwa  $p(x)$ , natomiast dla zaznaczenia miary prawdopodobieństwa przyjęliśmy oznaczenie  $p(dx)$ . Jak wspomnieliśmy we wstępie, będziemy rozważać nieliniowe modele przestrzeni stanów będące modelami ukrytych łańcuchów Markowa (*Hidden Markov Model*, HMM), w których dynamikę procesów  $\{X_t, Y_t\}_t$  opisują równania

$$(2.1) \quad x_t \sim k_\phi(x_t | x_{t-1}),$$

$$(2.2) \quad y_t \sim g_\theta(y_t | x_t),$$

gdzie  $[\phi, \theta] \in \Theta$  oznacza wektor parametrów modelu.

W prezentowanym modelu przyjmujemy następujące założenia:

- 1) Proces zmiennych ukrytych  $\{X_t\}_{t \in N}$  jest procesem Markowa, o znanym rozkładzie początkowym  $X_0 \sim \pi_0(dx_0)$  oraz prawdopodobieństwie przejścia

$$(2.3) \quad p(X_t \in dx_t | F_{t-1}) = p(X_t \in dx_t | X_{t-1} = x_{t-1}) := K_\phi(x_{t-1}, dx_t),$$

$$(2.4) \quad p(X_t \in A | X_{t-1} = x_{t-1}) = \int_A k_\phi(x_t | x_{t-1}) dx_t, \quad A \in B(\Omega),$$

gdzie przez  $F_t$  oznaczmy filtrację  $F_t := \sigma(X_k, Y_k; k \in \{0, \dots, t\})$ ,  $K$  jest jądrem przejścia Markowa, tzn. funkcją która spełnia następujące warunki  $K : R^{n_x} \times B(R^{n_x}) \rightarrow R$ :

- i)  $\forall A \in B(R^{n_x}) K(\cdot | A)$  jest funkcją mierzalną w sensie Lebesgue'a,
- ii)  $\forall x \in R^{n_x} K(x | \cdot)$  jest miarą probabilistyczną.
- iii) funkcja  $k(x | \cdot)$  jest gęstością rozkładu  $K(x, \cdot)$

- 2) Zmienne obserwowalne  $\{Y_t\}_{t \in N}$  mają znany rozkład brzegowy oraz są warunkowo niezależne względem sigma algebry  $F_t^- := \sigma(X_k, Y_k; k \in \{0, \dots, t-1\}, X_t)$ ,

$$(2.5) \quad p(Y_t \in B \mid X_t = x_t) = \int_B g_\theta(dy_t \mid x_t), \quad B \in \mathcal{B}(\Omega),$$

$$(2.6) \quad p(Y_t \in dy_t \mid F_t^-) = p(Y_t \in dy_t \mid X_t).$$

Funkcja  $g_\theta(y_t \mid x_t)$  jest funkcją gęstości względem miary Lebesgue'a

$$(2.7) \quad g(dy_t \mid x_t) = g(y_t \mid x_t)dy_t.$$

Powyższe własności, warunkowa niezależności obserwacji, (2.6) oraz fakt, że pełna informacja o procesie zawarta jest w stanie układu w chwili  $t$  (przeszłe obserwacje nie wnoszą żadnej wiedzy o przyszłych zmianach układu), (2.3) stanowią podstawę prezentowanych w pracy metod filtracji.

Praca ogranicza się do zaprezentowania metod filtracji bez poruszania problemu estymacji parametrów modelu, dlatego w dalszej części pracy będziemy pomijać fakt, iż funkcje gęstości  $k, g$  zależą od parametrów  $\phi, \theta$ .

Problem filtracji polega na wyznaczeniu funkcji gęstości rozkładu brzegowego  $p(x_t \mid y_{1:t})$ . Faktycznie w praktyce szacuje się łączny rozkład warunkowy  $p(x_{0:t} \mid y_{1:t})$ , gdzie łączny rozkład a posteriori zmiennej ukrytej i obserwowanej jest postaci

$$(2.8) \quad p(x_{0:T} \mid y_{1:T}) = \pi_0(x_0) \prod_{t=1}^T k(x_t \mid x_{t-1})g(y_t \mid x_t)$$

Wyznaczenie rozkładów  $p(x_{0:t} \mid y_{1:t})$  i  $p(x_t \mid y_{1:t})$  wymaga wyznaczania wielowymiarowych całek, co analitycznie możliwe jest jedynie dla niewielu modeli. Jak wspomnieliśmy we wstępie w takich przypadkach wskazane jest korzystanie z metod Monte Carlo. Prezentowana metoda polega na wyznaczeniu ciągu rozkładów prawdopodobieństwa  $\{\pi_t(x_t)\}_t$ , które są aktualizowane w czasie. Zagadnienie, którym będziemy się zajmować przedstawia krótko poniższa definicja

DEFINICJA. Ciąg zmieniających się w czasie rozkładów prawdopodobieństwa  $\{\pi_t(x_t)\}_t$  nazywany jest probabilistycznym układem dynamicznym. Ewolucja zmiennej stanu w czasie może być opisana dwoma scenariuszami

- Rozważamy  $X_{0:t}$  pod warunkiem obserwacji  $Y_{1:t} = y_{1:t}$ , wówczas w kolejnych momentach czasowych zwiększa się wymiar próbki  $X_{0:t} = (X_{0:t-1}, X_t)$ , a badane rozkłady prawdopodobieństwa są łącznymi rozkładami warunkowymi  $\pi_t(x_{0:t}) = p(x_{0:t} \mid y_{1:t})$ , określonymi na przestrzeni  $R^{t \cdot n_x}$ .
- Interesuje nas  $X_t$  pod warunkiem obserwacji  $Y_{1:t} = y_{1:t}$ , wtedy wymiar próbki nie zależy od czasu, a badane rozkłady prawdopodobieństwa są

rozkładami brzegowymi  $\pi_t(x_t) = p(x_t | y_{1:t})$ , określonymi na przestrzeni  $R^{n_x}$ .

Ze względu na fakt, iż będziemy rozważać dwa przypadki (warunkowy rozkład łączny i rozkład brzegowy), w dalszej części pracy będziemy przyjmować oznaczenia  $\pi_{0:t|t}(x_{0:t}) = p(x_{0:t} | y_{1:t})$ ,  $\pi_{t|t}(x_t) = p(x_t | y_{1:t})$ .

**3. Filtracja bayesowska.** Klasyczny problem filtracji sprowadza się do rekurencyjnego wyznaczania rozkładu brzegowego

$$(3.1) \quad \pi_{t|t}(dx_t) = p(X_t \in dx_t | y_{1:t}).$$

Powyższy rozkład zawiera wszystkie dostępne w chwili  $t$  informacje na temat zmiennej ukrytej. Niestety analityczna postać funkcji  $\pi_{t|t}$  istnieje jedynie dla kilku szczególnych modeli dynamicznych. Ze względu na brak analitycznej postaci gęstości rozważanego rozkładu proponuje się zastosowanie metod numerycznych. W pracy estymator filtru optymalnego będziemy definiować za pomocą rozkładu empirycznego

$$(3.2) \quad \pi_{t|t}^N(dx_t) = \frac{1}{N} \sum_{i=1}^N \delta(x_t - x_{t|t}^{(i)}),$$

gdzie cząsteczki  $\{x_{t|t}^{(i)}\}_{i=1}^N$  generowane są z odpowiednio dobranej funkcji  $q(\cdot)$  nazywanej funkcją ważności. Sposób wyboru funkcji  $q(\cdot)$  i jej istotność w poprawnym działaniu algorytmu zostanie przedstawiona w czwartym i piątym rozdziale pracy.

Idea prezentowanego algorytmu wynika z konieczności aktualizacji rozkładu a posteriori w momencie pojawienia się nowych danych. Reguła Bayesa pozwala oszacować brzegowy rozkład a posteriori  $\pi_{t|t}(dx_t)$  uwzględniając zarówno prawdopodobieństwa a priori jak i informacje spoza próby. Przy założeniu, że zmienna ukryta ma własność procesu Markowa<sup>(2)</sup>, funkcja gęstości filtracji może być wyrażona rekurencyjnie w postaci

$$(3.3) \quad \begin{aligned} \pi_{t|t}(x_t) &= p(x_t | y_{1:t}) = p(x_t | y_t, y_{1:t-1}) \\ \text{(korekta/uaktualnienie)} &= \frac{p(y_t | x_t)p(x_t | y_{1:t-1})}{p(y_t | y_{1:t-1})} \\ &= \frac{g(y_t | x_t)\pi_{t|t-1}(x_t)}{\int_{R^{n_x}} g(y_t | x_t)\pi_{t|t-1}(x_t)dx_t} \\ &\propto g(y_t | x_t)\pi_{t|t-1}(x_t), \end{aligned}$$

<sup>(2)</sup> Z równania Chapmana-Kołmogorowa dla procesów markowowskich zachodzi następująca równość  $p(x_t | y_{1:t-1}) = \int_R p(x_t, x_{t-1} | y_{1:t-1}) dx_{t-1} = \int_R k(x_t | x_{t-1}) p(x_{t-1} | y_{1:t-1}) dx_{t-1}$ .

gdzie funkcja

$$(3.4) \text{ (prognoza)} \quad \begin{aligned} \pi_{t|t-1}(x_t) &= \int_{R^{n_x}} p(x_t, x_{t-1} | y_{1:t-1}) dx_{t-1} \\ &= \int_{R^{n_x}} k(x_t | x_{t-1}) \pi_{t-1|t-1}(x_{t-1}) dx_{t-1} \end{aligned}$$

określa gęstość prognozy. Zakładamy, że  $p(y_t | y_{1:t-1}) > 0$ .

Analogiczne rozumowanie można przeprowadzić dla łącznych rozkładów warunkowych

$$(3.5) \text{ (korekta)} \quad \begin{aligned} \pi_{0:t|t}(x_{0:t}) &= p(x_{0:t} | y_{1:t}) = \frac{p(x_{0:t}, y_{1:t})}{p(y_{1:t})} \\ &= \frac{\pi_{0:t|t-1}(x_{0:t}) g(y_t | x_t)}{p(y_t | y_{1:t-1})}, \end{aligned}$$

$$(3.6) \text{ (prognoza)} \quad \begin{aligned} \pi_{0:t|t-1}(x_{0:t}) &= p(x_{0:t} | y_{1:t-1}) \\ &= \pi_{0:t-1|t-1}(x_{0:t-1}) k(x_t | x_{t-1}) \end{aligned}$$

Zauważmy jednak, że powyższe równania rekurencyjne (3.3–3.6) będące podstawą optymalizacji Bayesowskiej mają tylko charakter koncepcyjny, ponieważ występujące w nich funkcje gęstości nie mają postaci analitycznej.

Warto również wspomnieć, że w wielu przypadkach estymację rozkładu wygodnie jest sprowadzić (za pomocą wartości oczekiwanej) do problemu estymacji funkcjonału, który jednoznacznie wyznacza rozkład zmiennej losowej

$$(3.7) \quad \pi_{t|t}(\varphi) := E(\varphi(X_t) | y_{1:t}) = \int \varphi(x_t) p(x_t | y_{1:t}) dx_t, \quad \varphi \in C_b(R^{n_x}).$$

Powyższe przejście w szczególności wykorzystywane jest w dowodzeniu zbieżności omawianej metody (rozdział 9). Prezentując sposób dowodzenia zbieżności wygodnie jest wprowadzić dodatkowe oznaczenia. W tym celu definiujemy

- $(\mu, \varphi) \in P(R^{n_x}) \times C_b(R^{n_x}) \rightarrow \mu(\varphi) = \int \varphi(x) \mu(dx)$

Niech  $K(x, dz)$  będzie operatorem całkowym  $K : P(R^{n_x}) \rightarrow P(R^{n_x})$

- $(K\varphi)(x) := \int K(x, dz) \varphi(z)$
- $\mu K(dz) := \int \mu(dx) K(x, dz), \Rightarrow (\mu K)(f) = \mu(K(f))$

Na podstawie powyższych notacji, przy założeniu, że  $(\pi_{t|t-1}, g_t) > 0$  omawianą zależność rekurencyjną (3.3-3.4) można przedstawić w postaci układu

$$(3.8) \text{ (prognoza)} \quad (\pi_{t|t-1}, \varphi) = (\pi_{t-1|t-1}, K\varphi),$$

$$(3.9) \text{ (korekta)} \quad (\pi_{t|t}, \varphi) = (\pi_{t|t-1}, g\varphi)(\pi_{t|t-1}, g)^{-1}.$$

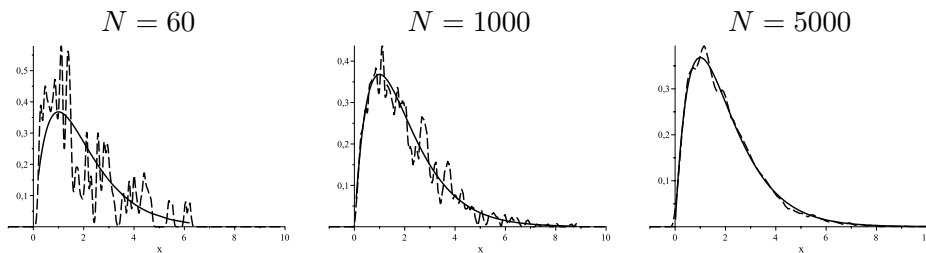
**4. Metoda Monte Carlo i funkcja ważności** (ang. *Importance Sampling*). Rozważany w tym rozdziale problem dotyczy aproksymacji funkcji gęstości. Wybrana przez nas metoda jest symulacyjną metodą Monte Carlo, której idea oparta jest na teorii dystrybucji. Wyobraźmy sobie, że zmienna losowa  $X$  ma rozkład o gęstości  $p$ . W przypadku, gdy dla rozważanego rozkładu, istnieje generator liczb pseudo-losowych, wówczas estymatorem miary prawdopodobieństwa  $p$  jest rozkład empiryczny postaci

$$(4.1) \quad \hat{p}_N(dx) = \frac{1}{N} \sum_{i=1}^N \delta(x - x^{(i)}),$$

gdzie  $\delta$  oznacza deltę Diraca<sup>(3)</sup>, natomiast ciąg  $\{x^{(i)}\}_{i=1}^N$  jest ciągiem niezależnych próbek wygenerowanych z rozkładu  $p$  (ang. *perfect Monte Carlo sampling*).

Poniższe wykresy przedstawiają wynik zastosowania przedstawionej metody w estymacji funkcji gęstości rozkładu gamma w zależności od liczby generowanych próbek.

Wykres 1.



Dysponując estymatorem funkcji gęstości możemy zastosować go do oszacowania wybranych charakterystyk rozkładów. Przypuśćmy, że interesuje nas oszacowanie momentu opisanego za pomocą funkcji  $\varphi$

$$(4.2) \quad p(\varphi(x)) = E_p[\varphi(x)] = \int \varphi(x)p(x)dx,$$

gdzie  $\varphi$  jest funkcją całkowalną względem  $p(x)$ .

Wtedy na podstawie estymatora (4.1) otrzymujemy oszacowanie postaci

$$(4.3) \quad \hat{p}_N(\varphi) = \frac{1}{N} \sum_{i=1}^N \varphi(x^{(i)}).$$

---

<sup>(3)</sup> Przyjmujemy konwencję, że zmienna losowa zawsze ma gęstość, którą definiujemy, jako pochodną dystrybuanty. W przypadku rozkładu, który nie jest absolutnie ciągle pochodną rozumiemy w sensie dystrybucyjnym.

Z mocnego prawa wielkich liczb estymator (4.3) jest prawie wszędzie zbieżny, dla  $N \rightarrow \infty$ . Przy dodatkowym założeniu  $p(\varphi^2) < \infty$  spełnione są założenia centralnego twierdzenia granicznego, tzn.

$$(4.4) \quad \sqrt{N}(\hat{p}_N(\varphi) - E_p[\varphi]) \xrightarrow[N \rightarrow \infty]{D} N(0, \frac{1}{N^2} \sum_{i=1}^N [\varphi(x^{(i)}) - \hat{p}_N(\varphi)]^2),$$

gdzie  $\xrightarrow{D}$  oznacza zbieżność w sensie dystrybuanty.

W przypadku gdy badany rozkład jest złożony, wysoko wymiarowy i nie-standardowy, w dodatku znany z dokładnością do stałej normującej, bezpośrednie generowanie próbek jest zadaniem skomplikowanym lub wręcz niemożliwym. W literaturze popularne są trzy metody generowania próby na podstawie empirycznego rozkładu, tj. funkcja ważności, metoda eliminacji lub markowskie metody Monte Carlo (MCMC). W naszej analizie proponujemy zastosować metodę funkcji ważności.

Idea metody polega na zastąpieniu analizowanego rozkładu  $p$  rozkładem  $q$  (o możliwie zbliżonych własnościach), dla którego istnieje generator liczb pseudolosowych. Za pomocą wprowadzonego rozkładu  $q$  generuje się próbkę niezależnych, ważonych zmiennych losowych  $\{x^{(i)}, w(x^{(i)})\}_{i=1}^N$ . Próbkę uznajemy za odpowiednio ważoną jeżeli prawdziwa jest tożsamość

$$(4.5) \quad E_q[\varphi(x)w(x)] = cE_p(\varphi(x)),$$

gdzie  $c$  jest stałą normującą (wspólną dla każdego  $N$ ).

Jeżeli powyższe założenie jest spełnione, wówczas metoda funkcji ważności umożliwia zamianę wartości oczekiwanej zmiennej  $\varphi(X)$  względem miary  $p$  na wartość oczekiwaną zmiennej  $\varphi(X) \circ w(X)$  względem gęstości  $q$

$$(4.6) \quad \begin{aligned} E_p(\varphi(x)) &= \int \varphi(x)p(x)dx = \int \varphi(x)\frac{p(x)}{q(x)}q(x)dx \\ &= \int \varphi(x)w(x)q(x)dx \\ &= E_q[\varphi(x)w(x)] \end{aligned}$$

Wagi  $w(x)$  określają prawdopodobieństwo, że próbka  $\{x^{(i)}\}_{i=1}^N$  pochodzi z rozkładu  $p$ . Wyznacza się je ze wzoru  $w(x^{(i)}) = \frac{p(x^{(i)})}{q(x^{(i)})}$ ,  $i = 1, \dots, N$  i nazywa wagami ważności. Gęstość  $q$ , która ma zasadnicze znaczenie w tej procedurze, nazywana jest funkcją ważności.

Dysponując próbką  $N$  niezależnych zmiennych losowych o gęstości  $q$  definiujemy estymator miary prawdopodobieństwa

$$(4.7) \quad \hat{p}_N(dx) = \hat{p}(dx) = \sum_{i=1}^N \tilde{w}(x^{(i)})\delta(x - x^{(i)}),$$



oraz estymator wartości oczekiwanej

$$(4.8) \quad \hat{p}_N(\varphi) = \sum_{i=1}^N \tilde{w}(x^{(i)}) \varphi(x^{(i)}),$$

gdzie  $\sum_{i=1}^N \tilde{w}(x^{(i)}) = 1$ ,  $\forall i = 1, \dots, N$   $\tilde{w}(x^{(i)}) \geq 0$ ,  $\{x^{(i)}\}_{i=1}^N$ , oznacza próbkę prostą z rozkładu o gęstości  $q$ .

Formalne matematyczne uzasadnienie powyższych rozważań związane jest z pochodną Radona-Nikodyma, można je znaleźć między innymi w pracy Geweke'a [13]. Efektywność metody funkcji ważności ściśle zależy od wariancji wag ważności. Okazuje się, że estymator (4.8) jest zbieżny z prawdopodobieństwem 1, jeżeli oszacowania wag są nieobciążone. Jeżeli wariancja wag nie jest skończona, wówczas wariancja estymatora (4.8) również jest nieskończona, to jednak nie przekreśla teoretycznej zbieżności omawianych estymatorów.

**5. Sekwencyjna funkcja ważności** (ang. *Sequential Importance Sampling*, SIS) W praktyce bardzo często bezpośrednie losowanie z rozkładu brzegowego  $p(x_t | y_{1:t})$  nie jest zadaniem prostym, dlatego próbki losuje się z łącznego rozkładu  $p(x_{0:t} | y_{1:t})$ , a następnie eliminuje się wartości  $x_{0:t-1}$ .

Idea sekwencyjnej estymacji rozkładu  $p(x_{0:t} | y_{1:t})$  polega na rekurencyjnym wyznaczaniu oszacowania  $\hat{p}_N(x_{0:t} | y_{1:t})$  bez zmiany historii stanów  $\{x_{0:t}^{(i)}\}_{i=1}^N$ . W wyniku metody uzyskujemy ciąg rozkładów  $\{\pi_{0:t|t}\}_{t \geq 1}$  zależnych od chwili  $t$ . Poniżej przedstawimy krótko schemat algorytmu. Zakładamy, że w chwili  $t-1$  cząsteczki  $\{x_{0:t-1}^{(i)}\}_{i=1}^N$  zostały wygenerowane z funkcji ważności  $q_{t-1}(x_{0:t-1})$ , ich wagi ważności wyznaczamy ze wzoru

$$(5.1) \quad w_{t-1}^{(i)} \propto \frac{p(x_{0:t-1}^{(i)} | y_{1:t-1})}{q_{t-1}(x_{0:t-1}^{(i)})}.$$

Następnie na ich podstawie w chwili  $t$  generujemy cząsteczki  $x_t^{(i)}$  z rozkładu  $q_t(\cdot | x_{0:t-1}^{(i)})$ .

Podstawą metody sekwencyjnej jest funkcja ważności, która daje się wyrazić rekurencyjną zależnością (Cappe, [4])

$$(5.2) \quad \begin{aligned} q_t(x_{0:t}) &= q_0(x_0) \prod_{j=1}^t q_j(x_j | x_{0:j-1}) \\ &= q_{t-1}(x_{0:t-1}) q_t(x_t | x_{0:t-1}). \end{aligned}$$

W markowowskich modelach zmiennych ukrytych funkcja ważności (5.2) przyjmuje postać

$$(5.3) \quad q_t(x_{0:t} | y_{1:t}) = q_{t-1}(x_{0:t-1} | y_{1:t-1}) q_t(x_t | x_{t-1}, y_t).$$

Generowanie próbki losowej z proponowanej sekwencyjnej funkcji ważności odbywa się zgodnie z przedstawionym poniżej algorytmem sekwencyjnej funkcji ważności

#### ALGORYTM SEKWENCYJNEJ FUNKCJI WAŻNOŚCI

Dla  $t = 0$ ,  $x_0^{(i)} \sim q_0(x_0)$ .

Dla  $t = 1, \dots, T$

- Dla  $j = 1 \dots N$  generujemy  $x_t^{(j)} \sim q_t(x_t | x_{0:t-1}^{(j)})$  i definiujemy  $x_{0:t}^{(j)} = \{x_{0:t-1}^{(j)}, x_t^{(j)}\}$
- Wyznaczamy wagi wygenerowanych cząstek

$$w_t^{(j)} \propto w_{t-1}^{(j)} u_t^{(j)}$$

- Normalizujemy wagi ważności cząsteczek  $\tilde{w}_t^{(j)} = \frac{w_t^{(j)}}{\sum_{i=1}^N w_t^{(i)}}$ ,  
 $\sum_j \tilde{w}_t^{(j)} = 1$ .

Pojawiające się w algorytmie wagi cząsteczek wyznaczane są z zależności

$$\begin{aligned}
 (5.4) \quad w_t^{(i)} &\propto \frac{p(x_{0:t}^{(i)} | y_{1:t})}{q_t(x_t^{(i)} | x_{0:t-1}^{(i)})q_{t-1}(x_{0:t-1}^{(i)})} \\
 &\propto \frac{p(x_{0:t}^{(i)} | y_{1:t})}{q_t(x_t^{(i)} | x_{0:t-1}^{(i)})p(x_{0:t-1}^{(i)} | y_{1:t-1})} \cdot \frac{p(x_{0:t-1}^{(i)} | y_{1:t-1})}{q_{t-1}(x_{0:t-1}^{(i)})} \\
 &\propto w_{t-1}^{(i)} \frac{p(x_{0:t}^{(i)} | y_{1:t})}{q_t(x_t^{(i)} | x_{0:t-1}^{(i)})p(x_{0:t-1}^{(i)} | y_{1:t-1})} \\
 &\propto w_{t-1}^{(i)} u_t^{(i)},
 \end{aligned}$$

gdzie  $u_t$ , w ogólnym przypadku, definiujemy wzorem

$$(5.5) \quad u_t^{(i)} = u_t(x_{0:t}^{(i)}) = \frac{p(x_{0:t}^{(i)} | y_{1:t})}{q_t(x_t^{(i)} | x_{0:t-1}^{(i)})p(x_{0:t-1}^{(i)} | y_{1:t-1})}.$$

W literaturze czynnik  $u_t$  nazywany jest współczynnikiem przyrostu wagi (ang. *incremental weight*). Efektywność metody silnie zależy od trafności wyboru funkcji ważności, dlatego w dalszej części pracy bardziej szczegółowo zajmiemy się tym problemem. Zauważmy, że jeżeli za kryterium optymalizacji wyboru funkcji ważności przyjmujemy minimalizację warunkowej wariancji wag w chwili  $t$ , wówczas optymalną funkcją ważności jest funkcja

[(<sup>4</sup>)][(<sup>5</sup>)]

$$(5.6) \quad q_t(x_t | x_{0:t-1}) = p(x_t | x_{0:t-1}).$$

W dalszej części pracy będziemy pomijać indeks określający moment, dla którego określana jest funkcja  $q$  (użyliśmy go aby w przejrzysty sposób zaprezentować ideę metody).

Używając optymalnej funkcji ważności, współczynnik przyrostu wagi wyraża się wzorem

$$(5.7) \quad u_t = \frac{p(x_{0:t}^{(i)} | y_{1:t}^{(i)})}{q(x_t^{(i)} | x_{0:t-1}^{(i)})p(x_{0:t-1}^{(i)} | y_{1:t-1}^{(i)})} = \frac{p(x_{0:t-1}^{(i)} | y_{1:t}^{(i)})}{p(x_{0:t-1}^{(i)} | y_{1:t-1}^{(i)})} \\ = \frac{\pi_{0:t-1|t}(x_{0:t-1}^{(i)})}{\pi_{0:t-1|t-1}(x_{0:t-1}^{(i)})}.$$

Podobnie, jak w klasycznej metodzie funkcji ważności, estymator łącznego rozkładu warunkowego zmiennych ukrytych wyraża się wzorem

$$(5.8) \quad \hat{\pi}_{0:T|T}^N(dx_{0:T}) = \hat{p}_N(dx_{0:T} | y_{1:T}) = \sum_{i=1}^N \tilde{w}_T^{(i)} \delta(x_{0:T} - x_{0:T}^{(i)}),$$

zaś estymator wartości oczekiwanej funkcji  $\varphi$

$$(5.9) \quad \hat{p}(\varphi) = \hat{\varphi}_N(x_{0:T}) = E_{\hat{p}_N(x_{0:T}|y_{1:T})}(\varphi(x_{0:T})) = \sum_{i=1}^N \varphi(x_{0:T}^{(i)}) \tilde{w}_T^{(i)},$$

gdzie  $x_{0:T}^{(i)}$  oznacza próbkę z rozkładu  $q(x_{0:T}^{(i)})$ .

Otrzymany estymator jest estymatorem obciążonym ale zgodnym (obciążenie jest rzędu  $O(N)$ ). W pracach [9], [13] Doucet i Geweke pokazali, że jeżeli  $\{x_{0:T}^{(i)}\}_{i=1}^N$  jest zbiorem niezależnych zmiennych losowych o tym samym rozkładzie, zachodzi inkluzja nośników  $\text{supp}p(x_{0:T} | y_{1:T}) \subseteq \text{supp}q(x_{0:T} | y_{1:T})$  oraz warunkowa wartość oczekiwana  $E_{\hat{p}_N(x_{0:T}|y_{1:T})}(\varphi(x_{0:T}))$  istnieje i jest skończona, to powyższy estymator jest zbieżny z prawdopodobieństwem jeden do badanej wartości oczekiwanej, przy  $N \rightarrow \infty$  (mocne prawo wielkich liczb)

$$(5.10) \quad \hat{\varphi}_N(x_{0:T}) \rightarrow \frac{E_{q(\cdot|y_{1:T})}[\varphi(x_{0:T})w_T]}{E_{q(\cdot|y_{1:T})}[w_T]}.$$

---

(<sup>4</sup>) Z postaci warunkowej wariancji wag ważności  $\text{Var}_q(w(X)) = \int \left[ \frac{p(x)}{q(x)} - 1 \right]^2 q(x) dx$  wynika, że optymalną funkcją ważności jest sama funkcja  $p(x)$ , tj.  $q = p$ .

(<sup>5</sup>) W przypadku, gdy rozkład  $p$  znany jest z dokładnością do czynnika normującego  $Z = \int p(x) dx$ , którego nie możemy wyznaczyć analitycznie, zamiast funkcji  $p$  rozważamy funkcję  $\gamma$ , taką że  $p(x) = \frac{\gamma(x)}{Z}$ .

Przy założeniach, że wartości oczekiwane

$$E_{p(x_{0:T}|y_{1:T})} \left[ \frac{p(x_{0:T} | y_{1:T})}{q(x_{0:T} | y_{1:T})} \right] \quad \text{i} \quad E_{p(x_{0:T}|y_{1:T})} \left[ \varphi^2(x_{0:T}) \frac{p(x_{0:T} | y_{1:T})}{q(x_{0:T} | y_{1:T})} \right]$$

istnieją i są skończone estymator (5.10) posiada graniczny rozkład normalny (zachodzi centralne twierdzenie graniczne), tzn.

$$(5.11) \quad \sqrt{N}(\hat{\varphi} - E_{p(x_{0:T}|y_{1:T})}[\varphi]) \sim N(0, \Sigma_\varphi),$$

gdzie  $\Sigma_f = \text{Var}_q[\tilde{w}(x)(\varphi(x) - E_p[\varphi(x)])]$ .

Wybór funkcji ważności jest istotnym elementem procedury, dlatego w dalszej części przedstawimy najczęściej stosowane sposoby określenia funkcji ważności.

### 5.1. Metody wyboru funkcji ważności rozkładu.

1) ROZKŁAD A PRIORI (ang. *bootstrap filter*) jest to najpopularniejsza i najszerzej stosowana postać funkcji ważności [14], w której za funkcję ważności przyjmuje się prawdopodobieństwo przejścia

$$(5.12) \quad q(x_t | x_{0:t-1}, y_{1:t}) = p(x_t | x_{t-1}) = k(x_t | x_{t-1}).$$

Wówczas warunkową funkcję gęstości można przedstawić w postaci

$$(5.13) \quad p(x_t | y_{0:t}) = p(x_t | y_t, y_{0:t-1}) = \frac{p(y_t | x_t)p(x_t | y_{0:t-1})}{p(y_t | y_{0:t-1})} \\ \propto \underbrace{g(y_t | x_t)}_{\text{waga}} \underbrace{p(x_t | y_{0:t-1})}_{\text{generator}}$$

$$(5.14) \quad p(x_t | y_{0:t-1}) = \int p(x_t | x_{t-1})p(x_{t-1} | y_{0:t-1})dx_{t-1} \\ \approx \int p(x_t | x_{t-1}) \frac{1}{N} \sum_{i=1}^N \delta(x_{t-1} - x_{t-1|t-1}^{(i)}) dx_{t-1} \\ = \frac{1}{N} \sum_{i=1}^N p(x_t | x_{t-1|t-1}^{(i)}) \\ = \frac{1}{N} \sum_{i=1}^N k(x_t | x_{t-1|t-1}^{(i)}).$$

W tym przypadku przyrost wagi ważności wyraża się proporcją

$$(5.15) \quad u_t^{(i)} \propto g(y_t | x_t^{(i)}).$$

Główną wadą tak skonstruowanej funkcji ważności jest niezależność od obserwacji, co może powodować brak odporności na wartości odstające [4]. Z drugiej jednak strony w większości przypadków taki wybór jest najkorzystniejszy ze względu na łatwość uzyskiwania próby, a dalej estymacji

badanego rozkładu (co znacznie zmniejsza stopień komplikacji algorytmu oraz czas jego działania).

2) JĄDRO OPTYMALNE (ang. *Optimal distribution*). Metody filtracji optymalnej zostały zaproponowane przez Liu i Chen [16] (choć wcześniej już były stosowane do szczególnych przypadków w pracy [22]). W tym podejściu funkcja ważności zależy od wartości  $x_{t-1}$  oraz  $y_t$

$$(5.16) \quad q(x_t | x_{0:t-1}, y_{1:t}) = p(x_t | x_{t-1}, y_t)$$

$$(5.17) \quad q(x_t | x_{t-1}, y_t) = \frac{k(x_t | x_{t-1})g(y_t | x_t)}{\int k(x_t | x_{t-1})g(y_t | x_t)dx_t}.$$

Uwzględniając założenia (2.3)–(2.5), na podstawie ogólnego wzoru (5.4) otrzymujemy następujący wzór na wagi ważności

$$(5.18) \quad w_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(y_t | x_t^{(i)})p(x_t^{(i)} | x_{t-1}^{(i)})}{q(x_t^{(i)} | x_{t-1}^{(i)}, y_t)}.$$

Taki wybór funkcji ważności jest optymalny ze względu na możliwość minimalizacji warunkowej wariancji wag ważności w chwili  $t$ . Zauważmy bowiem, że wariancja wyraża się wzorem

$$(5.19) \quad \begin{aligned} & \text{var}_{q(x_t | x_{t-1}^{(i)}, y_t)} [w_t^{(i)}] \\ &= \left( \frac{w_{t-1}^{(i)}}{p(y_t | y_{1:t-1})} \right)^2 \left[ \int \frac{(k(x_t | x_{t-1}^{(i)})g(y_t | x_t))^2}{q(x_t | x_{t-1}^{(i)}, y_t)} dx_t - p^2(y_t | x_{t-1}^{(i)}) \right] \\ &\stackrel{(*)}{=} \left( \frac{w_{t-1}^{(i)}}{p(y_t | y_{1:t-1})} \right)^2 \\ & \quad \times \left[ \int q(x_t | x_{t-1}^{(i)}, y_t) \left( \int k(\tilde{x}_t | x_{t-1}^{(i)})g(y_t | \tilde{x}_t)d\tilde{x}_t \right)^2 dx_t - p^2(y_t | x_{t-1}^{(i)}) \right] \\ &= \left( \frac{w_{t-1}^{(i)}}{p(y_t | y_{1:t-1})} \right)^2 \\ & \quad \times \left[ \int q(x_t | x_{t-1}^{(i)}, y_t) dx_t \left( \int k(\tilde{x}_t | x_{t-1}^{(i)})g(y_t | \tilde{x}_t)d\tilde{x}_t \right)^2 - p^2(y_t | x_{t-1}^{(i)}) \right] \\ &\stackrel{(**)}{=} \left( \frac{w_{t-1}^{(i)}}{p(y_t | y_{1:t-1})} \right)^2 \left[ \left( \int k(\tilde{x}_t | x_{t-1}^{(i)})g(y_t | \tilde{x}_t)d\tilde{x}_t \right)^2 - p^2(y_t | x_{t-1}^{(i)}) \right] = 0. \end{aligned}$$

W równości (\*) wykorzystujemy zależności (5.17), na podstawie której wyznaczamy wzór na postać iloczynu  $k(x_t | x_{t-1}^{(i)})g(y_t | x_t)$ . Przejście (\*\*) uzyskujemy zakładając, że  $q(x_t | x_{t-1}^{(i)}, y_t) = p(x_t | x_{t-1}^{(i)}, y_t)$ , wówczas całka

z funkcji gęstości wynosi 1. Przy takim wyborze funkcji ważności wagi spełniają zależność rekurencyjną postaci

$$(5.20) \quad w_t^{(i)} \propto w_{t-1}^{(i)} p(y_t | x_{t-1}^{(i)}).$$

W praktyce stosując jądro optymalne można napotkać dwa poważne problemy. Pierwszy związany jest z koniecznością losowania z rozkładu  $p(x_t | x_{t-1}^{(i)}, y_t)$ , drugi z obliczeniem całki  $p(y_t | x_{t-1}^{(i)}) = \int k(x_t | x_{t-1}^{(i)}) g(y_t | x_t) dx_t$  (może nie mieć postaci analitycznej). Postać analityczna wag istnieje tylko dla wybranych modeli. W literaturze problemu pojawia się coraz więcej technik efektywnego wyboru i estymacji funkcji ważności  $q(x_t | x_{t-1}, y_t)$  [5], [6], [9].

**Przykład 1.** [9] Rozważmy nieliniowy gaussowski model przestrzeni stanów

$$x_t = f(x_{t-1}) + v_t \quad v_t \sim N(0_{n_v \times 1}, \Sigma_v), \quad (5.21)$$

$$y_t = Cx_t + w_t \quad w_t \sim N(0_{n_w \times 1}, \Sigma_w), \quad (5.22)$$

gdzie  $f : R^{n_x} \rightarrow R^{n_x}$  jest nieliniową funkcją,  $C \in R^{n_y \times n_x}$  jest macierzą obserwacji. O składnikach losowych  $v_t$  i  $w_t$  zakładamy, że są wzajemnie niezależne,  $\Sigma_v > 0$  i  $\Sigma_w > 0$ .

Wykorzystując własności wielowymiarowego rozkładu gaussowskiego można pokazać, że warunkowy rozkład zmiennej ukrytej  $x_t$  względem zmiennej  $x_{t-1}$  i obserwacji  $y_t$  jest rozkładem normalnym

$$(5.23) \quad x_t | x_{t-1}, y_t \sim N(m_t, \Sigma),$$

o parametrach

$$(5.24) \quad \Sigma^{-1} = \Sigma_v^{-1} + C^T \Sigma_w^{-1} C$$

$$(5.25) \quad m_t = \Sigma(\Sigma_v^{-1} f(x_{t-1}) + C^T \Sigma_w^{-1} y_t),$$

oraz rozkładzie warunkowym

$$(5.26) \quad p(y_t | x_{t-1}) \propto \exp\left(-\frac{1}{2}(y_t - Cf(x_{t-1}))^T (\Sigma_v + C\Sigma_w C^T)^{-1} (y_t - Cf(x_{t-1}))\right).$$

W prezentowanym przykładzie wykorzystuje się zależność, jaką spełniają warunkowe funkcje gęstości

$$(5.27) \quad p(x_t | x_{t-1}, y_t) = \frac{p(x_t, y_t | x_{t-1})}{p(y_t | x_{t-1})}.$$

Na początku zauważmy, że równanie obserwacji możemy zapisać w postaci

$$(5.28) \quad y_t = Cf(x_{t-1}) + Cv_t + w_t.$$

W takim razie rozkład  $p(y_t | x_{t-1})$  jest rozkładem normalnym, w celu określenia gęstości wystarczy znać warunkową wartość oczekiwaną oraz wariancję  $E(y_t | x_{t-1}) = Cf(x_{t-1})$ ,  $\text{var}(y_t | x_{t-1}) = C\Sigma_v C^T + \Sigma_w$ .

Aby wyznaczyć rozkład  $p(x_t | x_{t-1}, y_t)$ , wystarczy znaleźć funkcję gęstości łącznego rozkładu  $p(x_t, y_t | x_{t-1})$ . Rozważany model (5.21, 5.22) możemy zapisać w postaci macierzowej

$$(5.29) \quad \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} f(x_{t-1}) \\ Cf(x_{t-1}) \end{bmatrix} + \begin{bmatrix} I & 0 \\ C & I \end{bmatrix} \begin{bmatrix} \nu_t \\ w_t \end{bmatrix},$$

wówczas korzystając z teorii wielowymiarowego rozkładu gaussowskiego warunkowy rozkład zmiennych  $(X_t, Y_t)$  względem zmiennej  $X_{t-1}$  jest rozkładem normalnym

$$(5.30) \quad \begin{bmatrix} x_t \\ y_t \end{bmatrix} \sim N \left( \begin{bmatrix} f(x_{t-1}) \\ Cf(x_{t-1}) \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right),$$

gdzie  $\Sigma_{11} = \Sigma_\nu$ ,  $\Sigma_{12} = \Sigma_\nu C^T - \Sigma_{21}^T$ ,  $\Sigma_{22} = C\Sigma_\nu C^T + \Sigma_w$ . Dokonując podstawienia oraz żmudnych przekształceń wzoru (5.27) otrzymujemy rozkład normalny zadany parametrami (5.24), (5.25).

3) ROZKŁADY SUBOPTYMALNE (ang. *suboptimal distributions*). W związku z trudnościami stosowania jądra optymalnego pojawiły się metody, w których proponuje się różne sposoby estymacji jądra. Rozkłady otrzymane wówczas nazywamy suboptymalnymi, ponieważ mówiąc „po literacku” są one optymalne w klasie takich, które się da policzyć. Jedną z takich metod (Doucet [9]) opiera się na idei Rozszerzonego Filtru Kalmana, w którym wykorzystuje się lokalną linearyzację modelu przestrzeni stanów. Niestety w literaturze ciągle brakuje ogólnej teorii dotyczącej przeprowadzania takiej procedury. W pracy [9] Arulampalam i Maskell twierdzą, że wprowadzanie do modelu dodatkowego kroku związanego z estymacją funkcji ważności może istotnie podnieść koszty działania algorytmu bez gwarancji istotnych korzyści. Z tego powodu najstosowniejszy ich zdaniem jest wybór jądra a priori.

**3.1. Funkcja ważności wyznaczana metodą lokalnej linearyzacji.** Jeden ze sposobów wyznaczenia funkcji ważności  $q(x_t | x_{t-1}, y_t)$  polega na rozważaniu funkcji ważności w postaci rozkładu parametrycznego  $q(x_t | \theta(x_{t-1}, y_t))$ . Przy czym zakładamy, że  $\theta$  jest skończone wymiarowym parametrem  $\theta \in \Theta \subset R^{n_\theta}$ , którego wartość determinowana jest przez  $x_{t-1}$  i  $y_t$ ,  $\theta : R^{n_x} \times R^{n_y} \rightarrow \Theta$  jest odwzorowaniem deterministycznym. Prezentowaną metodę przedstawimy na dwóch wybranych podejściach zaczerpniętych z pracy Douceta [9].

### A. LOKALNA LINEARYZACJA MARKOWOWSKIEGO MODELU PRZESTRZENI STANÓW.

Ideę metody przedstawimy rozważając następujący model

$$(5.31) \quad x_t = f(x_{t-1}) + v_t \quad v_t \sim N(0_{n_v \times 1}, \Sigma_v),$$

$$(5.32) \quad y_t = g(x_t) + w_t \quad w_t \sim N(0_{n_w \times 1}, \Sigma_w),$$

gdzie  $f : R^{n_x} \rightarrow R^{n_x}$ ,  $g : R^{n_x} \rightarrow R^{n_y}$  są funkcjami różniczkowalnymi, o składnikach losowych  $v_t$  i  $w_t$  zakładamy, że są wzajemnie niezależne,  $\Sigma_v > 0$  i  $\Sigma_w > 0$ . Szacując równanie obserwacji, za pomocą wzoru Peano w punkcie  $f(x_{t-1})$  otrzymujemy równanie liniowe

$$(5.33) \quad y_t = g(x_t) + w_t \\ \approx g(f(x_{t-1})) + \frac{\partial g(x_t)}{\partial x_t} \Big|_{x_t=f(x_{t-1})} \cdot (x_t - f(x_{t-1})) + w_t.$$

Nowy model złożony z równań (5.31) i (5.33) nie jest już modelem markowskim, ale jest liniowym modelem gaussowskim, dzięki temu szukana funkcja ważności jest postaci

$$(5.34) \quad q(x_t | x_{t-1}, y_t) \sim N(m_t, \Sigma_t),$$

ze średnią  $m_t$  i macierzą kowariancji  $\Sigma_t$  wyznaczaną dla każdej trajektorii na podstawie wzorów

$$(5.35) \quad \Sigma_t^{-1} = \Sigma_v^{-1} + \left[ \frac{\partial g(x_t)}{\partial x_t} \Big|_{x_t=f(x_{t-1})} \right]^T \Sigma_w^{-1} \frac{\partial g(x_t)}{\partial x_t} \Big|_{x_t=f(x_{t-1})},$$

$$(5.36) \quad m_t = \Sigma_t \left( \Sigma_v^{-1} f(x_{t-1}) + \left[ \frac{\partial g(x_t)}{\partial x_t} \Big|_{x_t=f(x_{t-1})} \right]^T \Sigma_w^{-1} (y_t - g(f(x_{t-1}))) \right. \\ \left. + \frac{\partial g(x_t)}{\partial x_t} \Big|_{x_t=f(x_{t-1})} f(x_{t-1}) \right).$$

Wagi ważności wyznacza się (z dokładnością do stałej) na podstawie wzoru (5.26).

**Przykład 2.** Model nieliniowy [14]. W literaturze jako klasyczny przykład modelu nieliniowego rozważa się model postaci

$$(5.37) \quad x_t = f(x_{t-1}) + v_t = 0,5x_{t-1} + 25 \frac{x_{t-1}}{1 + x_{t-1}^2} + 8 \cos(1, 2t) + v_t$$

$$(5.38) \quad y_t = g(x_t) + w_t = \frac{x_t^2}{20} + w_t,$$

gdzie  $x_0 \sim N(0, 5)$ , zmienne losowe  $v_t$  i  $w_t$  są wzajemnie niezależnymi zmiennymi losowymi generowanymi z rozkładów normalnych,  $v_t \sim N(0, \sigma_v^2)$ ,  $w_t \sim N(0, \sigma_w^2)$ . W tym przypadku nie da się wyznaczyć w sposób analityczny rozkładu  $p(y_t | x_{t-1})$  oraz nie można generować bezpośrednio z rozkładu



$p(x_t | x_{t-1}, y_t)$ . Z tego względu w celu wyznaczenia funkcji ważności proponuje się zastosowanie metody lokalnej linearyzacji równania obserwacji.

$$(5.39) \quad \begin{aligned} y_t &= g(x_t) + w_t \approx g(f(x_{t-1})) \\ &\quad + \frac{\partial g(x_t)}{\partial x_t} \Big|_{x_t=f(x_{t-1})} \cdot (x_t - f(x_{t-1})) + w_t \\ &= -\frac{f^2(x_{t-1})}{20} + \frac{f(x_{t-1})}{10}x_t + w_t \end{aligned}$$

Szukaną postacią rozkładu  $p(x_t | x_{t-1}, y_t)$  jest rozkład normalny  $N(m_t, \sigma_t^2)$  z parametrami zadanymi wzorami

$$(5.40) \quad \sigma_t^{-2} = \sigma_v^{-2} + \sigma_w^{-1} \left[ \frac{\partial g(x_t)}{\partial x_t} \Big|_{x_t=f(x_{t-1})} \right]^2 = \sigma_v^{-2} + \sigma_w^{-1} \frac{f^2(x_{t-1})}{100},$$

$$(5.41) \quad m_t = \sigma_t^2 \left( \sigma_v^{-2} f(x_{t-1}) + \sigma_w^{-2} \frac{f(x_{t-1})}{10} (y_t + \frac{f^2(x_{t-1})}{20}) \right).$$

Funkcja wag wyraża się również rozkładem normalnym takim jak niżej

$$(5.42) \quad p(y_t | x_{t-1}) = N((f^2(x_{t-1}) + \sigma_v^2)/20, \sigma_w^2).$$

## B. LOKALNA LINEARYZACJA OPTYMALNEJ FUNKCJI WAŻNOŚCI.

Zakładamy, że  $\ell(x_t) \stackrel{df}{=} \ln p(x_t | x_{t-1}, y_t)$  jest dwukrotnie różniczkowalna względem  $x_t$  na przestrzeni  $R^{n_x}$ , dzięki czemu możemy aproksymować funkcję  $\ell(x_t)$  za pomocą wzoru Taylora rzędu dwa w punkcie  $x$

$$(5.43) \quad \ell(x_t) \approx \ell(x) + [\ell'(x)]^T (x_t - x) + \frac{1}{2} (x_t - x)^T \ell''(x) (x_t - x),$$

gdzie

$$(5.44) \quad \ell'(x) = \frac{\partial \ell(x_t)}{\partial x_t} \Big|_{x_t=x},$$

$$(5.45) \quad \ell''(x) = \frac{\partial^2 \ell(x_t)}{\partial x_t \partial x_t^T} \Big|_{x_t=x}.$$

Punkt  $x$ , w którym dokonujemy aproksymacji, jest wybierany arbitralnie (ale określany jest przez odwzorowanie  $x_{t-1}$  i  $y_t$ ). Jeżeli dodatkowo założymy, że funkcja  $\ell''(x_t)$  jest ujemnie określona (warunek zachodzi, gdy funkcja  $\ell(x_t)$  jest wklęsła), wówczas określając

$$(5.46) \quad \Sigma(x) = -\ell''(x_t)^{-1}$$

$$(5.47) \quad m(x) = \Sigma(x) \ell'(x_t),$$

otrzymujemy

$$(5.48) \quad [\ell'(x)]^T (x_t - x) + \frac{1}{2} (x_t - x)^T \ell''(x) (x_t - x)$$

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

Z powyższych rozważań wynika, że za funkcja ważności jest gaussowską funkcją rozkładu

$$(5.49) \quad q(x_t | x_{t-1}, y_t) = N(m(x) + x, \Sigma(x)).$$

W przypadku, gdy funkcja  $p(x_t | x_{t-1}, y_t)$  jest unimodalna, najkorzystniej jest przyjąć, że wartość  $x$  jest medianą rozkładu  $p(x_t | x_{t-1}, y_t)$ , wtedy  $m(x) = 0_{n_x \times 1}$ .

Przykład 3. Klasycznym przypadkiem, w którym warto zastosować powyższe rozumowanie, jest liniowy model gaussowski z obserwacjami generowanymi z wykładniczej rodziny rozkładów

$$(5.50) \quad x_t = Ax_{t-1} + v_t,$$

$$(5.51) \quad p(y_t | x_t) = \exp(y_t^T Cx_t - b(Cx_t) + c(y_t)),$$

gdzie  $C$  jest macierzą  $n_y \times n_x$ ,  $b: R^{n_y} \rightarrow R$ ,  $c: R^{n_y} \rightarrow R$ ,  $\Sigma_v > 0$ . Wtedy

$$(5.52) \quad \begin{aligned} \ell(x_t) &= \ln p(x_t | x_{t-1}, y_t) \\ &= C + y_t^T Cx_t - b(Cx_t) - \frac{1}{2}(x_t - Ax_{t-1})^T \Sigma_v^{-1}(x_t - Ax_{t-1}), \end{aligned}$$

$$(5.53) \quad \ell''(x) = \frac{\partial^2 \ell(x_t)}{\partial x_t \partial x_t^T} \Big|_{x_t=x} = \frac{\partial^2 b(Cx_t)}{\partial x_t^2} \Big|_{x_t=x} - \Sigma_v^{-1}.$$

Przykład 4. Kolejny przykład dotyczy zastosowania metody filtru cząsteczkowego do procesu stochastycznej zmienności SV. Ciekawą cechą, która odróżnia model SV od modeli typu ARCH, GARCH, jest to, iż warunkowa wariancja jest odrębnym procesem stochastycznym. W praktyce oznacza to, iż zmiany warunkowej wariancji nie są utożsamiane z przyczynami zmian obserwacji procesu. Umożliwia to lepsze modelowanie rzeczywistej zmienności niż w przypadku modeli GARCH. Dodatkowymi zaletami omawianego modelu jest możliwość modelowania zjawiska „grubych ogonów”, oraz leptokurtozy. Powyższe własności procesu stanowią o jego konkurencyjności oraz powodują, że coraz częściej model SV stosowany jest do analizy zmienności danych finansowych [18].

Najczęściej proces stochastycznej zmienności  $\{x_t\}_{t \in N}$  definiuje się układem równań

$$(5.54) \quad x_t = \alpha x_{t-1} + \sigma \eta_t, \quad x_0 \sim N\left(0, \frac{\sigma_0^2}{1 - \alpha_0^2}\right),$$

$$(5.55) \quad y_t = \phi \exp\left(\frac{x_t}{2}\right) \varepsilon_t.$$

Zmienna  $x_t$  jest logarytmem wariancji warunkowej w chwili  $t$ , składniki losowe  $\eta_t$  i  $\varepsilon_t$  są wzajemnie niezależne, o standardowym rozkładzie normal-

nym. W praktyce zmienna  $y_t$  może na przykład oznaczać dzienną stopę zwrotu danego instrumentu finansowego. Przyjmując, że  $F_t$  jest zbiorem wszystkich informacji dostępnych o procesie obserwacji do chwili  $t$  włącznie (filtracja), a  $\theta$  jest wektorem parametrów modelu  $\theta = [\gamma, \phi, \sigma^2]$ , wówczas warunkowe gęstości omawianego modelu przyjmują postać  $f_\theta(x_t | x_{t-1}, \theta, F_{t-1}) = N(\alpha x_{t-1}, \sigma^2)$ ;  $g_\theta(y_t | x_t, F_{t-1}) = N(0, \phi^2 \exp(x_t))$ . Z dokładnością do stałej potrafimy wyznaczyć łączny rozkład zmiennej ukrytej względem obserwacji.

Istotnie, gęstość warunkowa wyrażona zależnością

$$\begin{aligned} p(x_k | x_{k-1}, y_k) &= \frac{p(x_k, x_{k-1}, y_k)}{p(x_{k-1}, y_k)} \\ &= \frac{p(y_k | x_k)p(x_k | x_{k-1})p(x_{k-1})}{\int p(y_k | x_k)p(x_k | x_{k-1})dx_k p(x_{k-1})} = \frac{p(y_k | x_k)p(x_k | x_{k-1})}{\int p(y_k | x_k)p(x_k | x_{k-1})dx_k} \\ &\propto p(y_k | x_k)p(x_k | x_{k-1}) \end{aligned}$$

przyjmuje postać

$$(5.56) \quad \exp\{-[\sigma^{-2}(x_k - \alpha x_{k-1})^2 + \phi^{-2} \exp(-x_k)y_k^2 + x_k]/2\}.$$

Zatem

$$(5.57) \quad p(x_{1:t} | y_{1:t}) = \prod_{k=2}^t \exp\{-[\sigma^{-2}(x_k - \alpha x_{k-1})^2 + \phi^{-2} \exp(-x_k)y_k^2 + x_k]/2\}.$$

Niestety ze względu na zależność zmiennych ukrytych, ich wysoki wymiar oraz obecność składnika  $\exp(-x_k)y_k^2$  nie można w sposób bezpośredni generować próbek z rozkładu (5.57). Dodatkowo sytuację komplikuje fakt, że nie istnieje analityczna postać funkcji określającej wagi cząsteczek  $p(y_t | x_{t-1})$ . W celu wyznaczenia postaci funkcji ważności proponujemy rozwinięcie składnika wykładniczego  $\exp(-x_k)$  w punkcie, który jest warunkową wartością oczekiwaną  $\alpha x_{k-1}$  zmiennej  $x_k$

$$(5.58) \quad \exp(-x_k) \approx \exp(-\alpha x_k) \left\{ 1 - (x_k - \alpha x_{k-1}) + \frac{1}{2}(x_k - \alpha x_{k-1})^2 \right\}.$$

Wykorzystując oszacowanie (5.58) we wzorze (5.56) w wykładniku funkcji  $\exp(\cdot)$  otrzymujemy następującą formę kwadratową

$$\begin{aligned} (5.59) \quad & -\frac{1}{2} (\sigma^{-2}(x_k - \alpha x_{k-1})^2 + \phi^{-2} y_k^2 \exp(-x_k) + x_k) \\ &= -\frac{1}{2} (\sigma^{-2}(x_k - \alpha x_{k-1})^2 + \phi^{-2} y_k^2 \exp(-\alpha x_{k-1}) \\ & \quad - \phi^{-2} y_k^2 \exp(-\alpha x_{k-1})(x_k - \alpha x_{k-1}) \\ & \quad + \frac{1}{2} \phi^{-2} y_k^2 \exp(-\alpha x_{k-1})(x_k - \alpha x_{k-1})^2 + x_k) \end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{2} \left( (\sigma^{-2} + \frac{1}{2}\phi^{-2}y_k^2 \exp(-\alpha x_{k-1})) (x_k - \alpha x_{k-1})^2 \right. \\
&\quad \left. + (-\phi^{-2}y_k^2 \exp(-\alpha x_{k-1}) + 1)(x_k - \alpha x_{k-1}) \right) + C(x_{k-1}, y_k) \\
&= -\frac{1}{2} \left( (\sigma^{-2} + \frac{1}{2}\phi^{-2}y_k^2 \exp(-\alpha x_{k-1})) \right. \\
&\quad \left. \times \left( (x_k - \alpha x_{k-1}) + \frac{1 - \frac{1}{2}\phi^{-2}y_k^2 \exp(-\alpha x_{k-1})}{2(\sigma^{-2} + \frac{1}{2}\phi^{-2}y_k^2 \exp(-\alpha x_{k-1}))} \right)^2 \right) \\
&\quad + C'(x_{k-1}, y_k)
\end{aligned}$$

Powyższa postać kanoniczna wskazuje, że szukany rozkład (funkcja ważności) będzie rozkładem normalnym z parametrami, wartością oczekiwaną

$$(5.60) \quad m_k = \alpha x_{k-1} - \frac{\{1 - \phi^2 y_k^2 \exp(-\alpha x_{k-1})\}/2}{\sigma^{-2} + \phi^2 y_k^2 \exp(-\alpha x_{k-1})/2},$$

oraz wariancją

$$(5.61) \quad \tau_k^2 = (\sigma^{-2} + \phi^{-2}y_k^2 \exp(-\alpha x_{k-1})/2)^{-1}.$$

Stosujemy algorytm SIR generując na początku  $X_0 \sim N(0, \sigma)$ , a kolejne wartości stanów  $X_k$  z wyznaczonej funkcji ważności, natomiast wagi generowanych cząsteczek wyznaczmy ze wzoru

$$(5.62) \quad \frac{\exp\{-[\sigma^{-2}(x_k - \alpha x_{k-1})^2 + \phi^{-2} \exp(-x_k)y_k^2 + x_k]/2\}}{\exp\{-[\tau_k^{-2}(x_k - m_k)^2/2\}\tau_k^{-1}}.$$

**6. Resampling.** W praktyce, po kilku iteracjach algorytmu SIS okazuje się, że wszystkie oprócz jednej wagi ważności przyjmują wartości bliskie zeru – pojawia się tzw. zjawisko degeneracji próby. W związku z tym do algorytmu wprowadza się dodatkowe losowanie mające na celu „odnowienie” trajektorii, z angielskiego nazywane resamplingiem. Podstawowa procedura resamplingu polega na losowaniu  $N$  wartości z dostępnej populacji cząsteczek (jest to losowanie z szacowanego rozkładu  $\tilde{\pi}_t^N$ ). Losowanie odbywa się zgodnie ze znormalizowanymi wagami wyznaczanymi podczas algorytmu. W literaturze znanych jest co najmniej kilka metod przeprowadzania resamplingu. A mianowicie resampling wielomianowy (ang. *multinomial resampling*), rezidualny (ang. *residual resampling*), systematyczny (ang. *systematic resampling*). Niestety pełne własności teoretyczne wymienionych metod ciągle stanowią otwarty problem. W pracy proponujemy korzystanie z algorytmu systematycznego oraz wielomianowego.

Prezentowane poniżej algorytmy są przedstawione dla ustalonego momentu  $t$ .

## RESAMPLING SYSTEMATYCZNY

1. Losujemy  $N$  uporządkowanych wartości  $u_k = \frac{k-1}{N} + u$ , gdzie  $u \sim U\left(\left(0, \frac{1}{N}\right]\right)$ . [6]
2. Powtórnie losowane cząsteczki  $x_{t|t}^{(j)}$ ,  $j = 1, \dots, N_i$  są otrzymywane jako kopie cząsteczki  $x_{t|t-1}^{(i)}$ , gdzie  $N_i$  jest liczbą wartości  $u_k \in \left(\sum_{s=1}^{i-1} \tilde{w}_t^{(s)}, \sum_{s=1}^i \tilde{w}_t^{(s)}\right]$ .

Zaletą powyższego algorytmu jest łatwość w jego konstrukcji, niestety cząsteczki generowane na jego podstawie przestają być niezależne, co znacznie utrudnia badanie zbieżności metody. W literaturze metoda często występuje pod nazwą próbkowania minimalizującego wariancję (*minimum variance sampling*), ponieważ wagi replikowanych cząsteczek wybierane są tak, aby minimalizować wartość wariancji  $Var[N_i] = E[N_i - E[N_i]]^2$ .

## RESAMPLING WIELOMIANOWY

1. Losujemy  $N$  niezależnych zmiennych losowych  $\{u_k\}_{k=1}^N$  o rozkładzie jednostajnym na przedziale  $(0, 1]$ .
2. Definiujemy indeks  $I(i) = F^{-1}(u_i)$  dla  $i = 1, \dots, N$ . Funkcja  $F^{-1}(\cdot)$  jest odwrotnością dystrybuanty wyznaczonej na podstawie znormalizowanych wag cząsteczek  $\{\tilde{w}_t^{(i)}\}_{i=1}^N$ , rozumianą w taki sposób, że  $F^{-1}(u) = i$ , jeśli  $u \in \left(\sum_{s=1}^{i-1} \tilde{w}_t^{(s)}, \sum_{s=1}^i \tilde{w}_t^{(s)}\right]$ .
3. Powtórnie losowane cząsteczki  $x_{t|t}^{(i)}$  otrzymujemy podstawiając  $x_{t|t}^{(i)} = x_{t|t-1}^{(I(i))}$ .

Resampling wielomianowy umożliwia wygenerowanie  $N$  nowych, niezależnych cząsteczek ze znanego zbioru. Każda nowa cząsteczka pojawia się  $N_i$  razy ( $N_i$  może być zerem), gdzie  $\sum_{i=1}^N N_i = N$ ,  $E(N_i) = N\tilde{w}_t^{(i)}$ ,  $Var(N_i) = N\tilde{w}_t^{(i)}(1 - \tilde{w}_t^{(i)})$ .

Niestety stosowanie procedury resamplingu oprócz korzyści – wyboru najbardziej prawdopodobnych trajektorii w efekcie wielokrotnego stosowania – powoduje degenerację różnorodności próby (ang. *sample impoverishment*) oraz utratę niezależności cząsteczek. Zauważmy bowiem, że trajektoria  $\{x_{0:t}^{(i)}, i = 1, \dots, N\}$ , która ma odpowiednio wysokie wagi ważności  $\tilde{w}_t^{(i)}$ , wybierana jest wiele razy, co w konsekwencji powoduje, że trajektorie  $x_{0:k}^{(i_1)}$

---

<sup>(6)</sup> Wówczas zmienna losowa  $u_k$  ma rozkład jednostajny na przedziale  $\left(\frac{k-1}{N}, \frac{k}{N}\right]$ .

i  $x_{0:k}^{(i_2)}$ , dla  $i_1 \neq i_2$ ,  $i_1, i_2 \in 1, \dots, N$  są identyczne [14]. Jednym ze sposobów zapobiegania degeneracji trajektorii jest ograniczenie częstotliwości przeprowadzania resamplingu. W literaturze (między innymi [3], [4]) miarą degeneracji algorytmu jest współczynnik definiowany za pomocą wariancji wag ważności

$$(6.1) \quad N_{ESS} = \frac{N}{1 + \text{var}_q(\tilde{w}(x_{0:t}))} = \frac{N}{E_q[\tilde{w}(x_{0:t})^2]} \leq N.$$

Współczynnik ten nazywany jest efektywną wielkością próbki (ang. *effective sample size*). W praktyce, niestety nie można wyznaczyć go analitycznie, dlatego posługujemy się oszacowaniem powyższej wielkości wyznaczanej ze wzoru

$$(6.2) \quad \hat{N}_{ESS} = \left( \sum_{i=1}^N (w_t^{(i)})^2 \right)^{-1}.$$

Na podstawie wzoru (6.1) możemy wyznaczyć przedział zmienności dla współczynnika efektywnej wielkości próby tj.  $1 \leq N_{ESS} \leq N$ . W praktyce proponuje się przeprowadzanie resamplingu w przypadku, gdy  $\hat{N}_{ESS} < N_T$ , gdzie  $N_T$  jest wartością krytyczną (zazwyczaj ustalaną na poziomie równym połowie liczby iteracji algorytmu  $N_T = \frac{N}{2}$ ).

**7. Klasyczny algorytm filtru cząsteczkowego** (ang. *Sampling importance resampling filter*, SIR). Klasyczny algorytm filtru cząsteczkowego SIR jest najbardziej znaną i najczęściej stosowaną postacią symulacyjnej metody Monte Carlo wykorzystywaną do numerycznej aproksymacji rozkładu  $\{p(x_t | y_{1:t})\}_{t \geq 1}$ . Algorytm SIR jest rozszerzoną postacią procedury SIS. Modyfikacja polega na wprowadzeniu ponownego próbkowania, (ang. *resampling*), który ma zapobiegać problemowi degeneracji próbek, czyli propagowaniu nieistotnych cząsteczek do kolejnych iteracji algorytmu. W wersji klasycznej wyznaczenie brzegowego rozkładu  $p(x_t | y_{1:t})$  wymaga przeprowadzenia dwóch następujących po sobie etapów predykcji i tzw. korekty (uaktualnienia). W prezentowanym poniżej algorytmie funkcja ważności oraz współczynnik  $u_t^{(j)}$  (przyrost wagi) występują w ogólnej postaci, a resampling jest dokonywany w każdej iteracji (*bootstrap* PF, [14]).

#### SIR

- 1) Dla  $t = 0$  z zadanego rozkładu początkowego generujemy  $N$  stanów początkowych  $x_{0|-1}^{(j)} \sim \pi_0(x_0)$ ,  $j = 1, \dots, N$ ,  
Dla  $t = 1, \dots, T$  oraz  $j = 1, \dots, N$  przeprowadzamy dwukrokową indukcję ze względu na  $t$   
(LOSOWANIE)

- 2) Dla  $j = 1 \dots N$  losujemy  $x_{t|t-1}^{(j)} \sim q(x_{t|t-1} | x_{t-1|t-1}^{(j)})$   
 3) Wyznaczamy wagi wygenerowanych cząsteczek

$$w_t^{(j)} \propto w_{t-1}^{(j)} u_t^{(j)}$$

Po normalizacji otrzymujemy wagi cząsteczek  $\tilde{w}_t^{(j)} = \frac{w_t^{(j)}}{\sum_{i=1}^N w_t^{(i)}}$ ,

$\sum_j^N \tilde{w}_t^{(j)} = 1$ . Otrzymujemy estymatory

$$\pi_{t|t-1}^N(dx_t) = \frac{1}{N} \sum_{i=1}^N \delta_{x_{t|t-1}^{(i)}}(dx_t),$$

$$\tilde{\pi}_{t|t}^N(dx_t) = \sum_{i=1}^N \tilde{w}_t^{(i)} \delta_{x_{t|t-1}^{(i)}}(dx_t)$$

(RESAMPLING)

- 4) Zgodnie z wyznaczonym prawdopodobieństwem losujemy niezależnie (ze zwracaniem) cząsteczki  $\{x_{t|t}^{(j)}\}_{j=1}^N$

$$P(x_{t|t}^{(j)} = x_{t|t-1}^{(i)}) = \tilde{w}_t^{(i)}, \quad j = 1, \dots, N.$$

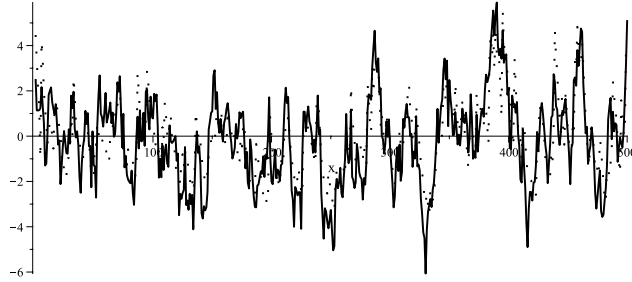
Na podstawie próbki  $\{x_{t|t}^{(j)}, \frac{1}{N}\}_{j=1}^N$  otrzymujemy empiryczną miarę prawdopodobieństwa

$$\pi_{t|t}^N(dx_t) = \frac{1}{N} \sum_{i=1}^N \delta_{x_{t|t}^{(i)}}(dx_t)$$

- 5) Podstawiamy  $t = t + 1$  i powtarzamy algorytm od punktu 2.

Poniżej prezentujemy wynik zastosowania algorytmu SIR do estymacji stanów zmiennej ukrytej modelu SV z parametrami  $\alpha = 0,85$ ,  $\beta = 0,65$ ,  $\sigma = 1$ .

Wykres 2. Realizacja zmiennej ukrytej. Linia ciągła – realizacja na podstawie modelu SV, linia przerywana – wynik estymacji na podstawie algorytmu SIR z  $N = 500$ , dla szeregu o długości  $T = 500$ .



**7.1. Ograniczenia.** Zaprezentowany algorytm umożliwia uzyskanie zgodnych estymatorów, w wielu przypadkach jest łatwy w implementacji, efektywność algorytmu nie zależy od wymiaru badanego układu, ale w praktyce musimy być świadomi jego niedoskonałości. Okazuje się bowiem, że nawet w przypadkach, gdy potrafimy właściwie oszacować funkcję ważności algorytm, SIR może okazać się nieefektywny. Taki problem pojawia się wówczas, gdy ciągi dystrybuant  $\pi_{n-1|n}(x_{n-1})$  i  $\pi_{n-1|n-1}(x_{n-1})$  są rozbieżne. Wówczas wariancja przyrostu wag jest duża i konieczne jest częste przeprowadzanie resamplingu. To z kolei może spowodować, że estymator

$$\tilde{\pi}_{t|t}^N(dx_t) = \sum_{i=1}^N \tilde{w}_t^{(j)} \delta_{x_{t|t-1}^{(i)}}(dx_t)$$

będzie niewiarygodny, wprowadzenie dodatkowego losowania wiąże się bowiem z dodatkową wariancją<sup>(7)</sup>. Chociaż naszym celem jest zachęcenie Czytelnika do korzystania z metod cząsteczkowych, musimy jednak zwrócić Jego uwagę na kilka niedoskonałości algorytmu SIR. A mianowicie algorytm wykazuje wysoka wrażliwość na trafność wyboru funkcji ważności, w chwili  $t$  następuje tylko losowanie zmiennej  $X_t^{(i)}$ , a pozostałe wartości (ścieżki)  $X_{1:t-1}^{(i)}$  pozostają niezmiennie w świetle nowych obserwacji. W szczególności problem polega na tym, że dla  $k \ll n$  dystrybuanta brzegowa  $\pi_{k:n|n}(x_{k:n})$  będzie szacowana na podstawie tylko kilku cząsteczek (a może się zdarzyć, że tylko jednej), ponieważ w algorytmie między chwilą  $k$  a  $n$  wielokrotnie wykonano resampling. Można więc oczekiwać, że wiarygodną estymację uzyskamy tylko wtedy, gdy  $n - k$  będzie mała.

W pracy przedstawimy dwie wybrane metody będące modyfikacją klasycznego algorytmu filtru cząsteczkowego:

- *auxiliary particle filter*,
- *block sampling strategy*.

## 8. Zaawansowane metody SMC. W poszukiwaniu optymalnej po-

<sup>(7)</sup> W ten sposób pojawia się często spotykany w statystyce problem, który można sformułować: wypośredkować pomiędzy stratą informacji i kumulacją błędów.



stacji algorytmu w rozdziale 3 przedstawiliśmy możliwości wyboru sekwencyjnej funkcji ważności. Kolejna z proponowanych modyfikacji dotyczy sposobu przeprowadzenia resamplingu.

**8.1. Auxiliary particle filter (APF).** Metoda APF zaproponowana przez Pitta i Shepharda [19] jest próbą zwiększenia efektywności działania metody funkcji ważności. W omawianym podejściu funkcję ważności definiuje się za pomocą jądra optymalnego  $q(x_t | x_{0:t-1}, y_{1:t}) = p(x_t | x_{t-1}, y_t)$ . W celu zmniejszenia wariancji wag ważności zmienia się kolejność kroków algorytmu SIR. Najpierw przeprowadzany jest resampling, a potem odbywa się generowanie cząsteczek (poprawność rozumowania wynika z faktu, iż wynik losowania z funkcji  $q$  nie zależy od  $x_t$ ). Okazuje się, że taka zmiana daje podwójną korzyść. Z jednej strony pozwala na zwiększenie różnorodności cząsteczek, co umożliwi lepsze szacowanie rozkładu. Z drugiej losowość związana z resamplingiem nie wpływa na wagi ważności, dzięki czemu można ograniczyć stratę informacji (metoda APF umożliwia redukcję wariancji wag ważności). Resampling cząsteczek w chwili  $t$  przeprowadza się z prawdopodobieństwem bliskim  $p(y_t | x_{t-1})$ , tak więc dokonuje się próby przewidzenia najbardziej prawdopodobnych cząsteczek.

W praktyce zaproponowana modyfikacja oznacza generowanie cząsteczek z rozszerzonego rozkładu brzegowego o funkcji gęstości  $p(x_t, l | y_{1:t})$ . Rozkład ten można wyrazić z dokładnością do stałej następującym równaniem

$$(8.1) \quad \begin{aligned} p(x_t, l | y_{1:t}) &\propto g(y_t | x_t) p(x_t, l | y_{1:t-1}) \\ &= g(y_t | x_t) p(x_t | l, y_{1:t-1}) p(l | y_{1:t-1}^{(l)}). \end{aligned}$$

Jeżeli bezpośrednio generowanie z rozkładu  $p(x_t, l | y_{1:t})$  jest skomplikowane, wówczas jest on aproksymowany rozkładem wyznaczonym przez funkcję ważności postaci

$$(8.2) \quad q(x_t, l | y_{1:t}) \propto g(y_t | \mu_t^{(l)}) k(x_t | x_{t-1}^{(l)}) w_{t-1}^{(l)}.$$

Parametr  $l \in \{1, \dots, N\}$  nazywany jest zmienną pomocniczą (ang. *auxiliary variable*), wartość  $\mu_t^{(l)}$  jest pewną charakterystyką zmiennej  $x_t$  wyznaczaną pod warunkiem  $x_{t-1}^{(l)}$ . Najczęściej  $\mu_t^{(l)}$  wyznaczana jest jako wartość oczekiwana, mediana lub dowolna wartość wylosowana z rozkładu  $K(x_t | x_{t-1}^{(l)})$ . Dla ogólnego zapisu będziemy przyjmować  $\mu_t^{(l)} = h(x_{t-1}^{(l)})$ .

Przedstawiając funkcję (8.1) w postaci iloczynu

$$(8.3) \quad q(x_t, l | y_{1:t}) = q(l | y_{1:t}) q(x_t | l, y_{1:t}),$$

oraz przyjmując, że

$$(8.4) \quad q(x_t | l, y_{1:t}) = k(x_t | x_{t-1}^{(l)}),$$

otrzymujemy następującą zależność

$$(8.5) \quad q(l | y_{1:t}) \propto g(y_t | \mu_t^{(l)}) w_{t-1}^{(l)}$$

Powyższe rozumowanie pokazuje, że generowanie z rozkładu  $q(x_t, l | y_{1:t})$  jest równoważne dwustopniowemu losowaniu, a mianowicie najpierw generujemy wartość indeksu  $l^{(j)} \sim q(l | y_{1:t})$ , a na jego podstawie losujemy z funkcji przejścia  $k(x_t | x_{t-1}^{(l)})$ .

#### Algorytm auxiliary particle filter

- 1) Dla  $j = 1, \dots, N$  generujemy  $x_0^{(j)} \sim \pi_0(x_0)$ , wyznaczamy wagi ważności  $\tilde{w}_0^{(j)} \propto g(y_0 | x_0^{(j)})$   
Dla  $t = 1, \dots, T$ .
- 2) Dla  $t = 1, \dots, N$  wyznaczamy  $\mu_t^{(j)} = h(x_{t-1}^{(j)})$ ,
- 3) Losujemy  $N$  indeksów cząsteczek  $l^j \in \{1, \dots, N\}$  zgodnie z rozkładem dyskretnym o wagach  $v_t^{(j)} = w_{t-1}^{(j)} g(y_t | \mu_t^{(j)})$ ,  $j = 1, \dots, N$ ,  
(RESAMPLING)
- 4) Dla  $j = 1, \dots, N$  generujemy próbkę  $\{\tilde{x}_t^{(j)}\}$ , taką że  $P(x_t^{(j)} = \tilde{x}_t^{(j)}) = u_t^{(j)}$ , gdzie  $u_t^{(j)} = \frac{w_t^{(j)}}{v_t^{(j)}}$ ,  
(LOSOWANIE)
- 4) Dla  $j = 1, \dots, N$  losujemy  $\tilde{x}_t^{(j)} \sim q(\tilde{x}_t^{(j)} | x_{t-1}^{(j)}, y_t)$ . Wyznaczamy wagi ważności

$$\tilde{w}_t^{(j)} = u_{t-1}^{(j)} \frac{g(y_t | \tilde{x}_t^{(j)}) k(\tilde{x}_t^{(j)} | x_{t-1}^{(j)})}{q(\tilde{x}_t^{(j)} | x_{t-1}^{(j)}, y_t)}$$

Po normalizacji otrzymujemy wagi cząsteczek  $w_t^{(j)} = \frac{\tilde{w}_t^{(j)}}{\sum_{i=1}^N \tilde{w}_t^{(i)}}$ ,

$j = 1, \dots, N$ .

Oczekuje się, że wyznaczone wagi  $w_t$  charakteryzują się mniejszą zmiennością niż wagi otrzymywane w klasycznym algorytmie PF.

Na podstawie powyższych rozważań otrzymujemy następujący estymator rozkładu brzegowego

$$(8.6) \quad p(x_t | y_{1:t}) \propto \sum_{i=1}^N g(y_t | \mu_t^{(l^i)}) K(x_t | x_{t-1}^{(l^i)}) w_{t-1}^{(l^i)}$$

Dla zobrazowania działania powyższej procedury przytoczymy przykład prezentowany przez Pitta i Shepharda w pracy [19].

**Przykład 7.** Ponownie rozważamy model stochastycznej zmienności (5.52)–(5.53), ale tym razem wykorzystamy fakt, że funkcja  $\log g(y_t | x_t)$  jest

wkłęśła. Na podstawie zależności (8.2)–(8.5) funkcję  $p(x_t, l \mid y_{1:t})$  możemy przedstawić jak niżej

$$\begin{aligned} p(x_t, l \mid y_{1:t}) &\propto g(y_t \mid x_t, \mu_t^l) k(x_t \mid x_{t-1}^l) \\ &= g(y_t \mid \mu_t^l) p(x_t \mid x_{t-1}^l, y_t, \mu_t^l). \end{aligned} \quad (8.7)$$

Rozszerzoną postać funkcji  $\log g(y_t \mid x_t)$  rozwijamy w punkcie  $\mu_t^l$ ,  $\mu_t^l = \alpha x_{t-1}^l$ , otrzymując

$$\begin{aligned} (8.8) \quad \log p(y_t \mid x_t, \mu_t^l) &= \log g(y_t \mid \mu_t^l) + (x_t - \mu_t^l)^T \frac{\partial \log g(y_t \mid \mu_t^l)}{\partial x_t} \\ &\quad + \frac{1}{2} (x_t - \mu_t^l)^T \frac{\partial^2 \log g(y_t \mid \mu_t^l)}{\partial x_t \partial x_t^T} (x_t - \mu_t^l)^T \\ &= C - \frac{1}{2} x_t - \frac{y_t^2}{2\phi^2} \exp(-\mu_t^l) [1 - (x_t - \mu_t^l)], \end{aligned}$$

gdzie  $C$  jest stałą.

Postać (8.7) oznacza, że generowanie z rozkładu  $p(x_t, l \mid y_{1:t})$  polega na wygenerowaniu indeksu z rozkładu  $g(y_t \mid \mu_t^l)$ , a następnie generowaniu cząsteczek z rozkładu  $p(x_t \mid x_{t-1}^l, y_t, \mu_t^l)$ . Dla rozważanego modelu stochastycznej zmienności

$$\begin{aligned} (8.9) \quad p(x_t \mid x_{t-1}^l, y_t, \mu_t^l) &= N \left( \mu_t^l + \frac{\sigma_\eta^2}{2} \left[ \frac{y_t^2}{\phi^2} \exp(-\mu_t^l) - 1 \right], \sigma_\eta^2 \right) \\ &= N(\mu_t^{*l}, \sigma_\eta^2), \end{aligned}$$

$$(8.10) \quad g(y_t \mid \mu_t^l) = \exp \left( \frac{1}{2\sigma^2} (\mu_t^{*l2} - \mu_t^{l2}) \right) \exp \left( -\frac{y_t^2}{2\phi^2} \exp(-\mu_t^l) (1 + -\mu_t^l) \right).$$

Resampling przeprowadzany jest na podstawie wag

$$\begin{aligned} (8.11) \quad w_t^{(j)} &= u_{t-1}^{(j)} \frac{g(y_t \mid \tilde{x}_t^{(j)}) k(\tilde{x}_t^{(j)} \mid x_{t-1}^{(j)})}{g(y_t \mid \mu_t^{(l^{(j)})}) p(\tilde{x}_t^{(j)} \mid x_{t-1}^{(l^{(j)})}, y_t, \mu_t^{(l)})} \\ &= u_{t-1}^{(j)} \frac{g(y_t \mid \tilde{x}_t^{(j)})}{p(y_t \mid x_t^j, \mu_t^{(l^{(j)})})}. \end{aligned}$$

**8.2. Block sampling SIR.** Kolejną, wybraną przez nas modyfikacją klasycznego algorytmu filtru cząsteczkowego, jest metoda bloków. Idea metody BSIR polega na tym, iż losowana jest nie tylko próbka  $\{X_t^{(i)}\}$ , ale pewien fragment trajektorii o ustalonej długości  $\{X_{t-L+1:t-1}^{(i)}\}$ . Intuicyjnie wydaje się, iż aktualizacja (powtórne losowanie) ustalonej części trajektorii  $x_{t-L+1:t-1}^{(i)}$  na podstawie nowych informacji dostępnych w chwili  $t$  może istotnie podnieść efektywność metody. Poniżej prezentujemy schemat metody zaproponowany w pracy [11].

Załóżmy, że w chwili  $t - 1$  dysponujemy zbiorem ważonych cząsteczek  $\{X_{0:t-1}^{(i)}, W_{t-1}^{(i)}\}_{i=1}^N$  (aproxymacja rozkładu  $\pi_{0:t-1|t-1}$ ) oraz, że funkcja  $q_t(x'_{t-L+1:t} | x_{0:t-1})$  określa prawdopodobieństwo przejścia z bieżącej trajektorii  $x_{1:t-1}$  do stanów  $x'_{t-L+1:t}$ . Wówczas losując blok trajektorii  $X'_{t-L+1:t}$ , z rozkładu  $q_t(\cdot | x_{0:t-1}^{(i)})$  możemy skonstruować nową ścieżkę  $\{X_{0:t-L}^{(i)}, X'_{t-L+1:t}\}$ , przy czym poprzednio wyznaczone stany trajektorii  $\{X_{t-L+1:t-1}^{(i)}\}$  zastępujemy wartościami  $X'_{t-L+1:t-1}$ . Niech  $\gamma_{t-1}$  będzie rozkładem  $\{X_{0:t-1}^{(i)}\}$  w chwili  $t - 1$ , wtedy łączny rozkład  $\{X_{0:t-1}^{(i)}, X'_{t-L+1:t}\}$  możemy przedstawić w postaci

$$(8.12) \quad \gamma_t(x_{0:t-1}, x'_{t-L+1:t}) = \gamma_{t-1}(x_{0:t-1})q_t(x'_{t-L+1:t} | x_{0:t-1}).$$

Pierwszy problem związany z powyższym rozumowaniem polega na wyznaczeniu łącznego rozkładu próbki  $\{X_{0:t-L}^{(i)}, X'_{t-L+1:t}\}$ . Wspomniany rozkład wymaga obliczenia wielowymiarowej całki, a mianowicie

$$(8.13) \quad \gamma_t(x_{0:t-L}, x'_{t-L+1:t}) = \int \gamma_t(x_{0:t-1}, x'_{t-L+1:t}) dx_{t-L+1:t-1}.$$

W praktyce napotykamy kolejną trudność związaną z ustaleniem analitycznej postaci wag ważności. W tej sytuacji Docet i Briers [11] zaproponowali wprowadzenie funkcji ważności definiowanej na rozszerzonej przestrzeni kojarzonej z  $\{X_{0:t-1}^{(i)}, X'_{t-L+1:t}\}$ . Definiowanie funkcji ważności jako funkcji łącznego rozkładu pozwala ominąć problem związany z całkowaniem. W miejsce dotychczas analizowanych rozkładów  $\pi_{t|t}(x_t)$ ,  $\pi_{0:t|t}(x_{0:t})$  w algorytmie BSIR rozważa się rozkład

$$(8.14) \quad \pi_t(x_{0:t-L}, x'_{t-L+1:t})\lambda_t(x_{t-L+1:t-1} | x_{0:t-L}, x'_{t-L+1:t}).$$

Wprowadzona „sztucznie” funkcja  $\lambda_t(x_{t-L+1:t-1} | x_{0:t-L}, x'_{t-L+1:t})$  (ang. *artificial conditional distribution*) umożliwia analizowanie funkcji  $\pi_t(x_{0:t-L}, x'_{t-L+1:t})$  jako brzegowej funkcji rozkładu. Przy powyższych założeniach wagi ważności wyznaczane są z następującej relacji

$$(8.15) \quad \frac{\pi_t(x_{0:t-L}, x'_{t-L+1:t})\lambda_t(x_{t-L+1:t-1} | x_{0:t-L}, x'_{t-L+1:t})}{\gamma_t(x_{0:t-1}, x'_{t-L+1:t})} \\ = \frac{\pi_t(x_{0:t-1})}{\gamma_t(x_{0:t-1})} \cdot \frac{\pi_t(x_{0:t-L}, x'_{t-L+1:t})\lambda_t(x_{t-L+1:t-1} | x_{0:t-L}, x'_{t-L+1:t})}{\pi_t(x_{0:t-1})q_t(x'_{t-L+1:t} | x_{0:t-1})}.$$

#### Block sampling SIR/BSIR

Dla  $t = 0$ ,  $i = 1, \dots, N$

1) Losujemy  $x_0^{(i)} \sim \pi_0(x_0)$ ,

2) Wyznaczamy wagi cząsteczek  $w_0^{(i)} = w(x_0^{(i)}) \propto g(y_0 | x_0^{(i)})$ ,

- 3) Zgodnie z wyznaczonymi wagami (po normalizacji) losujemy niezależnie, ze zwracaniem cząsteczki  $\{\bar{x}_0^{(j)}\}_{j=1}^N$  otrzymując jednakowo ważne próbki  $\{x_0^{(j)}, \frac{1}{N}\}_{j=1}^N$ .

Dla  $1 \leq t < L$ ,  $i = 1, \dots, N$

- 4) Losujemy  $x'_{1:t}{}^{(i)} \sim q_t(\cdot | x_{0:t-1}^{(i)})$ .

- 5) Wyznaczamy aktualne wagi cząsteczek

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{\pi_t(x'_{0:t}{}^{(i)}) \lambda_t(x_{0:t-1}^{(i)} | x'_{0:t}{}^{(i)})}{\pi_{t-1}(x_{0:t-1}^{(i)}) q_t(x'_{0:t}{}^{(i)} | x_{0:t-1}^{(i)})}$$

- 6) Po znormalizowaniu wag  $w_t^{(i)}$ , z zadaniem prawdopodobieństwem ze zbioru  $\{x'_{0:t}{}^{(i)}\}$ , losujemy bloki otrzymując próbkę  $\{x_{0:t}^{(j)}, \frac{1}{N}\}_{j=1}^N$ .

Dla  $t \geq L$ ,  $i = 1, \dots, N$

- 7) Losujemy  $x'_{t-L+1:t}{}^{(i)} \sim q_t(\cdot | x_{0:t-1}^{(i)})$ .

- 8) Wyznaczamy wagi

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{\pi_t(x_{0:t-L}^{(i)}, x'_{t-L+1:t}{}^{(i)}) \lambda_t(x_{t-L+1:t-1}^{(i)} | x_{0:t-L}^{(i)}, x'_{t-L+1:t}{}^{(i)})}{\pi_{t-1}(x_{0:t-1}^{(i)}) q_t(x'_{t-L+1:t}{}^{(i)} | x_{0:t-1}^{(i)})}$$

- 9) Resampling  $\{w_t^{(i)}, x_{0:t-L}^{(i)}, x'_{t-L+1:t}{}^{(i)}\} \rightarrow \{\frac{1}{N}, x_{1:t}^{(i)}\}$ .

Przyjmując jako kryterium optymalizacji minimalizację warunkowej wariancji wag, Doucet i Briers w pracy [11] pokazali, że optymalne funkcje  $q$ ,  $\lambda$  są definiowane odpowiednio

$$(8.16) \quad q_t^{opt}(x'_{t-L+1:t} | x_{0:t-1}) = \pi_t(x'_{t-L+1:t} | x_{0:t-L})$$

$$(8.17) \quad \lambda_t^{opt}(x_{t-L+1:t-1} | x_{0:t-L}, x'_{t-L+1:t}) = \pi_{t-1}(x_{t-L+1:t-1} | x_{0:t-L}).$$

W związku z powyższym zastosowanie algorytmu wymaga wyznaczenia oszacowań rozkładów  $\pi_t(x'_{t-L+1:t} | x_{0:t-L})$ ,  $\pi_{t-1}(x'_{t-L+1:t} | x_{0:t-L})$ . Zauważmy, że przyjmując definicje (8.16), (8.17) wagi ważności wyrażają się wzorem

$$(8.18) \quad w_t^{opt(i)} \propto w_{t-1}^{(i)} \frac{\pi_t(x_{0:t-L}^{(i)})}{\pi_{t-1}(x_{0:t-L}^{(i)})}$$

W przypadku markowowskich modeli przestrzeni stanów optymalna funkcja ważności oraz wagi ważności określane są następująco

$$(8.19) \quad q_t^{opt}(x'_{t-L+1:t} | x_{0:t-1}) = \pi_t(x'_{t-L+1:t} | x_{0:t-L})$$

$$= p(x'_{t-L+1:t} | x_{t-L}, y_{t-L+1:t}),$$

$$(8.20) \quad \pi_t(x_{0:t-L}) = p(x_{0:t-L} | y_{0:t}),$$

$$(8.21) \quad w_t^{opt} \propto w_{t-1} \frac{\pi_t(x_{0:t-L})}{\pi_{t-1}(x_{0:t-L})} \\ = w_{t-1} \frac{p(x_{0:t-L} | y_{1:t})}{p(x_{0:t-L} | y_{1:t-1})} \propto p(y_t | x_{t-L}, y_{t-L+1:t-1}).$$

Przykłady prezentowane w pracy [11] pokazują, że losowanie bloków w metodzie filtru cząsteczkowego pozwala istotnie zredukować częstotliwość przeprowadzania resamplingu, a dzięki temu otrzymać lepszą aproksymację badanego rozkładu. Niestety koszt przeprowadzenia powyższej metody jest większy niż metody klasycznej. Metoda bloków powinna być stosowana w przypadkach, gdy obserwujemy istotne zmniejszenie wskaźnika ESS w stosunku do filtru standardowego.

**9. Problem zbieżności (zgodności) omawianego estymatora.** Celem niniejszej pracy jest, jak wiadomo, wyznaczenie estymatorów stanu zmiennej ukrytej, oraz rozkładu warunkowego. Postać estymatora punktowego zazwyczaj zależy od przyjętego kryterium optymalności (funkcji straty). W naszych rozważaniach estymator zmiennej ukrytej będziemy definiować za pomocą warunkowej wartości oczekiwanej

$$(9.1) \quad \hat{x}_t^N = \hat{x}_t = E_{\pi_{t|t}}(X_t | y_{1:t}),$$

ponieważ taki wybór jest optymalny ze względu na minimalizację błędu średniokwadratowego<sup>[(8)]</sup>. Następnie metodą filtru cząsteczkowego wyznaczamy oszacowania rozkładu brzegowego

$$(9.2) \quad \pi_{t|t}^N(dx_t) = \frac{1}{N} \sum_{i=1}^N \delta_{x_t^{(i)}}(dx_t).$$

Wektor  $\hat{x}_t^N$  i funkcja  $\pi_{t|t}^N(dx_t)$  są zmiennymi losowymi (zależą od losowo generowanych cząsteczek), dlatego naturalnym i bardzo istotnym jest pytanie o zachowanie się wyznaczonych estymatorów, gdy  $N \rightarrow \infty$ . Zauważmy, że w rozważanym przypadku będziemy badać zbieżność dla danego  $t$  <sup>[(9)]</sup>. Problem zbieżności będziemy badać w sensie zbieżności średniokwadratowej. W literaturze badane jest również zbieżności z prawdopodobieństwem jeden.

Od tej pory będziemy zakładać, że proces obserwacji jest ustalony  $Y_t = y_t$ ,  $\|f\| = \sup_{x \in R^{n_x}} f(x)$ . Rozważania dotyczące zbieżności średniokwadratowej [9] opierają się na założeniach Z1, Z2, Z3:

<sup>(8)</sup> Warto zwrócić uwagę, że estymację  $E_{\pi_{t|t}}(\varphi(x_t) | y_{1:t})$  powinno się przeprowadzać przed dokonaniem resamplingu, który zwiększa wariancję próbek.

<sup>(9)</sup> Wszystkie stochastyczne wielkości (tj.  $E$ , zbieżność prawie wszędzie) dotyczą losowości związanej z generowaniem cząsteczek.

Z1.  $g(y_t | \cdot)$  jest funkcją ograniczoną,  $\|g\| < \infty$ ,  $\forall x_t \in R^{n_x}$ .

Z2.  $\varphi$  – jednostajnie ciągła i ograniczona  $R^{n_x}$ ,  $\varphi \in B(R^{n_x})$ .

Z3. Jądro  $K$  spełnia warunek Fellera, tzn. dla dowolnej funkcji ciągłej  $\varphi \in C_b(R^{n_x})$  ciągłą jest również funkcja  $x_{t-1} \mapsto \int \varphi(x_t) K(dx_t | x_{t-1})$ .

W naszej pracy zastosowaliśmy wielomianową procedurę losowania (Multinomial sampling procedure), wówczas

$$(9.3) \quad E \left| \sum_{i=1}^N (N_t^{(i)} - N\tilde{w}_t^{(i)}) q^{(i)} \right|^2 \leq CN \max_{i=1 \dots N} |q^{(i)}|^2,$$

gdzie  $N_t^{(i)}$  jest wartością naturalną pochodząca z rozkładu wielomianowego o parametrach  $(N, w_1^{(i)}, \dots, w_t^{(i)})$ ,  $\sum_{i=1}^N N_t^{(i)} = N$ ,  $q = [q^{(1)}, q^{(2)}, \dots, q^{(N)}] \in R^N$ .

Z przyjętego rozkładu wynika, że  $E(N_t^{(i)}) = Nw_t^{(i)}$ ,  $\text{var}(N_t^{(i)}) = Nw_t^{(i)}(1 - w_t^{(i)})$ .

TWIERDZENIE 1. (o średniokwadratowej zbieżności) [9]. *Przy założeniu Z1, Z2, dla  $\forall t > 0 \exists c_{t|t}$  niezależna od  $N$ , taka że*

$$(9.4) \quad E \left[ (\pi_{t|t}^N, \varphi) - (\pi_{t|t}, \varphi) \right]^2 \leq \frac{c_{t|t}}{N} \|\varphi\|^2,$$

gdzie wartość oczekiwana liczona jest względem wszystkich realizacji cząstek losowych.

Dla dowodu twierdzenia wystarczy pokazać, że przy przyjętych założeniach w  $t = 0$  spełniony jest warunek  $\forall \varphi \in B(R^{n_x})$

$$(9.5) \quad E \left[ (\pi_0^N, \varphi) - (\pi_0, \varphi) \right]^2 \leq \frac{\|\varphi\|^2}{N}$$

i indukcyjnie dowieść zależności (9.4), korzystając z trzech następujących pomocniczych lematów, których dowody oparte są na oszacowaniu dla wartości oczekiwanych warunkowych.

LEMAT 1. *Jeżeli dla każdej funkcji  $\varphi \in B(R^{n_x})$  wartość oczekiwana w chwili  $t - 1$  spełnia warunek*

$$E \left[ (\pi_{t-1|t-1}^N, \varphi) - (\pi_{t-1|t-1}, \varphi) \right]^2 \leq \frac{c_{t-1|t-1}}{N} \|\varphi\|^2,$$

wówczas dla każdej funkcji  $\varphi \in B(R^{n_x})$  istnieje stała  $c_{t|t-1}$ , taka że

$$(9.6) \quad E \left[ (\pi_{t|t-1}^N, \varphi) - (\pi_{t|t-1}, \varphi) \right]^2 \leq \frac{c_{t|t-1}}{N} \|\varphi\|^2.$$

LEMAT 2. *Jeżeli dla każdego  $\varphi \in B(R^{n_x})$  i dla każdego  $t > 0$  funkcja  $\pi_{t-1|t-1}^N$  spełnia nierówność (9.4) z parametrem  $c_{t-1|t-1}$  oraz funkcja  $\pi_{t|t-1}^N$*

spełnia nierówność (9.6), to dla każdej funkcji  $\varphi \in \mathcal{B}(R^{n_x})$  istnieje stała  $\tilde{c}_{t|t}$  taka, że

$$(9.7) \quad E \left[ (\tilde{\pi}_{t|t}^N, \varphi) - (\tilde{\pi}_{t|t}, \varphi) \right]^2 \leq \frac{\tilde{c}_{t|t}}{N} \|\varphi\|^2.$$

LEMAT 3. Jeżeli dla każdego  $\varphi \in \mathcal{B}(R^{n_x})$  oraz dla każdego  $t > 0$  spełniona jest nierówność (9.7), wówczas  $\forall \varphi \in \mathcal{B}(R^{n_x})$  taka, że

$$(9.8) \quad E \left[ (\pi_{t|t}^N, \varphi) - (\pi_{t|t}, \varphi) \right]^2 \leq \frac{c_{t|t}}{N} \|\varphi\|^2.$$

Dla wykazania zbieżności algorytmu SIR wystarczy udowodnić prawdziwość powyższych lematów. Ze względu na fakt iż dowody lematów są krótkie, a pozwalają lepiej zrozumieć budowę algorytmu, poniżej prezentujemy sposób ich dowodzenia.

*Dowód twierdzenia 1* dla chwili  $t = 0$ . Generujemy  $i = 1, \dots, N$  cząstek  $x_0^{(i)}$  z rozkładu  $\pi_0(dx)$ , wówczas estymator empiryczny funkcji gęstości rozkładu inicjującego jest postaci  $\pi_0^N(dx) = \frac{1}{N} \sum_{i=1}^N \delta_{x_0^{(i)}}(dx)$ . Na podstawie przyjętych założeń zachodzi równość  $E(\pi_0^N, \varphi) = (\pi_0, \varphi)$ . Teraz dokonując elementarnych przekształceń możemy pokazać prawdziwość twierdzenia dla  $t = 0$

$$\begin{aligned} E \left[ (\pi_0^N, \varphi) - (\pi_0, \varphi) \right]^2 &= E \left[ \frac{1}{N} \sum_{i=1}^N \varphi(x_0^{(i)}) - \frac{1}{N} \sum_{i=1}^N (\pi_0, \varphi) \right]^2 \\ &= \frac{1}{N^2} E \left[ \sum_{i=1}^N (\varphi(x_0^{(i)}) - (\pi_0, \varphi)) \right]^2 \\ &= \frac{1}{N^2} \sum_{i,j=1}^N E(\varphi(x_0^{(i)}) - (\pi_0, \varphi))(\varphi(x_0^{(j)}) - (\pi_0, \varphi)) \\ &= \frac{1}{N^2} \sum_{i=1}^N E(\varphi(x_0^{(i)}) - (\pi_0, \varphi))^2 \\ &= \frac{1}{N} E(\varphi(x_0^{(i)}) - (\pi_0, \varphi))^2 = \frac{1}{N} E(\varphi(x_0^{(i)})^2 - 2\varphi(x_0^{(i)})(\pi_0, \varphi) + (\pi_0, \varphi)^2) \\ &= \frac{1}{N} E(\varphi(x_0^{(i)})^2 - (\pi_0, \varphi)^2) \leq \frac{1}{N} E\varphi(x_0^{(i)})^2 \leq \frac{\|\varphi\|^2}{N}. \end{aligned}$$

*Dowód lematu 1.* Dla udowodnienia zbieżności estymatora prognozy zmiennej ukrytej korzystamy z nierówności trójkąta, wówczas moduł róż-



nicy szacujemy przez sumę modułów

$$\begin{aligned} & \left| (\pi_{t|t-1}^N, \varphi) - (\pi_{t|t-1}, \varphi) \right| = \left| (\pi_{t|t-1}^N, \varphi) - (\pi_{t-1|t-1}, K\varphi) \right| \\ & \leq \left| (\pi_{t|t-1}^N, \varphi) - (\pi_{t-1|t-1}^N, K\varphi) \right| + \left| (\pi_{t-1|t-1}^N, K\varphi) - (\pi_{t-1|t-1}, \varphi) \right|. \end{aligned}$$

Z algorytmu wiemy, że  $\pi_{t|t-1}^N(dx_t) = \frac{1}{N} \sum_{i=1}^N \delta_{x_{t|t-1}^{(i)}}(dx_t)$ , stąd  $(\pi_{t|t-1}^N, \varphi) = \frac{1}{N} \sum_{i=1}^N \varphi(x_{t|t-1}^{(i)})$ .

Przyjmując, że  $\Sigma_{t-1}$  jest  $\sigma$  algebrą generowaną przez  $\{x_{t-1|t-1}^{(j)}\}_{j=1}^N$  możemy pokazać, że

$$E \left[ (\pi_{t|t-1}^N, \varphi) \mid \Sigma_{t-1} \right] = \frac{1}{N} \sum_{i=1}^N E(\varphi(x_{t|t-1}^{(i)}) \mid \Sigma_{t-1}) = (\pi_{t-1|t-1}^N, K\varphi),$$

Z drugiej strony

$$(\pi_{t-1|t-1}^N, K\varphi) = \frac{1}{N} \sum_{i=1}^N K\varphi(x_{t-1|t-1}^{(i)}) = \frac{1}{N} \sum_{i=1}^N \int \varphi(z) K(x_{t-1|t-1}^{(i)}, dz).$$

Z faktu, że  $K$  jest prawdopodobieństwem przejścia otrzymujemy  $\|K\varphi\| \leq \|\varphi\|$ , co pozwala dalej szacować

$$\begin{aligned} & E \left[ \left( (\pi_{t|t-1}^N, \varphi) - E \left[ (\pi_{t|t-1}^N, \varphi) \mid \Sigma_{t-1} \right] \right)^2 \mid \Sigma_{t-1} \right] \\ & = E \left[ \left[ (\pi_{t|t-1}^N, \varphi) - (\pi_{t-1|t-1}^N, K\varphi) \right]^2 \mid \Sigma_{t-1} \right] \\ & = E \left[ \frac{1}{N^2} \left( \sum_{i=1}^N \varphi(x_{t|t-1}^{(i)}) - \int \varphi(z) K(x_{t|t-1}^{(i)}, dz) \right)^2 \mid \Sigma_{t-1} \right] \\ & = \frac{1}{N^2} E \left( \sum_{i,j=1}^N \left[ \varphi(x_{t|t-1}^{(i)}) - \int \varphi(z) K(x_{t|t-1}^{(i)}, dz) \right] \right. \\ & \quad \left. \times \left[ \varphi(x_{t|t-1}^{(j)}) - \int \varphi(z) K(x_{t|t-1}^{(j)}, dz) \right] \mid \Sigma_{t-1} \right) \\ & = \frac{1}{N^2} \left( \sum_{i,j=1}^N E \left[ \varphi(x_{t|t-1}^{(i)}) - \int \varphi(z) K(x_{t|t-1}^{(i)}, dz) \right] \right. \\ & \quad \left. \times \left[ \varphi(x_{t|t-1}^{(j)}) - \int \varphi(z) K(x_{t|t-1}^{(j)}, dz) \right] \mid \Sigma_{t-1} \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{N^2} \sum_{i=1}^N E \left( \left[ \varphi(x_{t|t-1}^{(i)}) - \int \varphi(z) K(x_{t|t-1}^{(i)}, dz) \right]^2 \mid \Sigma_{t-1} \right) \\
&= \frac{1}{N^2} \sum_{i=1}^N E \left( \left[ \varphi(x_{t|t-1}^{(i)})^2 - 2\varphi(x_{t|t-1}^{(i)}) \int \varphi(z) K(x_{t|t-1}^{(i)}, dz) \right. \right. \\
&\quad \left. \left. + \left( \int \varphi(z) K(x_{t|t-1}^{(i)}, dz) \right)^2 \right] \mid \Sigma_{t-1} \right) \\
&= \frac{1}{N} \left( (\pi_{t-1|t-1}^N, K\varphi^2) - (\pi_{t-1|t-1}^N, K\varphi)^2 \right) \leq \frac{1}{N} (\pi_{t-1|t-1}^N, K\varphi^2) \stackrel{(1)}{\leq} \frac{\|\varphi\|^2}{N}
\end{aligned}$$

Nierówność oznaczona (1) wynika z nierówności Schwarz'a, w kolejnym przejściu szacowania błędu estymacji korzystamy z nierówności Minkowskiego

$$\begin{aligned}
&E \left[ ((\pi_{t|t-1}^N, \varphi) - (\pi_{t|t-1}, \varphi))^2 \right]^{1/2} \\
&\leq E \left[ ((\pi_{t|t-1}^N, \varphi) - (\pi_{t-1|t-1}^N, K\varphi))^2 \right]^{1/2} \\
&\quad + E \left[ ((\pi_{t-1|t-1}^N, K\varphi) - (\pi_{t-1|t-1}, \varphi))^2 \right]^{1/2} \\
&\leq \frac{\sqrt{c_{t|t-1}}}{\sqrt{N}} \|\varphi\|,
\end{aligned}$$

gdzie  $c_{t|t-1} = (1 + \sqrt{c_{t-1|t-1}})^2$ .  $\square$

*Dowód lematu 2.* Dla oszacowania różnicy między estymatorem ważonym a estymatorem otrzymanym po dokonaniu resamplingu cząsteczek, wykorzystuje się fakt, iż  $(\pi_{t|t}, \varphi) = (\pi_{t|t-1}, g\varphi)(\pi_{t|t-1}, g)^{-1}$ . Wówczas

$$\begin{aligned}
(\tilde{\pi}_{t|t}^N, \varphi) - (\pi_{t|t}, \varphi) &= \frac{(\pi_{t|t-1}^N, g\varphi)}{(\pi_{t|t-1}^N, g)} - \frac{(\pi_{t|t-1}, g\varphi)}{(\pi_{t|t-1}, g)} + \frac{(\pi_{t|t-1}^N, g\varphi)}{(\pi_{t|t-1}, g)} - \frac{(\pi_{t|t-1}, g\varphi)}{(\pi_{t|t-1}, g)} \\
&\left| \frac{(\pi_{t|t-1}^N, g\varphi)}{(\pi_{t|t-1}^N, g)} - \frac{(\pi_{t|t-1}, g\varphi)}{(\pi_{t|t-1}, g)} \right| = \frac{(\pi_{t|t-1}^N, g\varphi) \left| (\pi_{t|t-1}, g) - (\pi_{t|t-1}^N, g) \right|}{(\pi_{t|t-1}^N, g)(\pi_{t|t-1}, g)} \\
&\leq \frac{\|\varphi\| \int g(dy \mid x) \pi_{t|t-1}^N(dx) \left| (\pi_{t|t-1}, g) - (\pi_{t|t-1}^N, g) \right|}{(\pi_{t|t-1}^N, g)(\pi_{t|t-1}, g)} \\
&= \frac{\|\varphi\| (\pi_{t|t-1}^N, g) \left| (\pi_{t|t-1}, g) - (\pi_{t|t-1}^N, g) \right|}{(\pi_{t|t-1}^N, g)(\pi_{t|t-1}, g)} \\
&= \frac{\|\varphi\| \left| (\pi_{t|t-1}, g) - (\pi_{t|t-1}^N, g) \right|}{(\pi_{t|t-1}, g)}.
\end{aligned}$$

Na podstawie lematu 1 szacujemy, że  $E \left[ (\pi_{t|t-1}^N, g) - (\pi_{t|t-1}, g) \right]^2 \leq \frac{c_{t|t-1}}{N} \|g\|^2$  oraz  $E \left[ (\pi_{t|t-1}^N, g\varphi) - (\pi_{t|t-1}, g\varphi) \right]^2 \leq \frac{c_{t|t-1}}{N} \|g\|^2 \|\varphi\|^2$ . Następnie korzystając z nierówności Minkowskiego otrzymujemy oszacowanie

$$\begin{aligned} E \left[ \left( (\tilde{\pi}_{t|t}^N, \varphi) - (\pi_{t|t}, \varphi) \right)^2 \right]^{1/2} &\leq E \left[ \left( \frac{(\pi_{t|t-1}^N, g\varphi)}{(\pi_{t|t-1}^N, g)} - \frac{(\pi_{t|t-1}, g\varphi)}{(\pi_{t|t-1}, g)} \right)^2 \right]^{1/2} \\ &+ E \left[ \left( \frac{(\pi_{t|t}^N, g\varphi)}{(\pi_{t|t}, g)} - \frac{(\pi_{t|t-1}, g\varphi)}{(\pi_{t|t-1}, g)} \right)^2 \right]^{1/2} \\ &\leq \frac{\|\varphi\|}{(\pi_{t|t-1}, g)} E \left[ \left( (\pi_{t|t-1}, g) - (\pi_{t|t-1}^N, g) \right)^2 \right]^{1/2} \\ &+ \frac{E \left[ \left( (\pi_{t|t-1}^N, g\varphi) - (\pi_{t|t-1}, g\varphi) \right)^2 \right]^{1/2}}{(\pi_{t|t-1}, g)} \\ &\leq \frac{2\sqrt{c_{t|t-1}} \|g\| \cdot \|\varphi\|}{(\pi_{t|t-1}, g)\sqrt{N}} \leq \sqrt{\tilde{c}_{t|t}} \frac{\|\varphi\|}{\sqrt{N}}, \end{aligned}$$

Z założenia, wiemy że  $\|g\| \leq \infty$ . □

W dowodzie lematu 3 wykorzystamy z nierówności Minkowskiego

$$\begin{aligned} E \left[ \left( (\pi_{t|t}^N, \varphi) - (\pi_{t|t}, \varphi) \right)^2 \right]^{1/2} &\leq E \left[ \left( (\pi_{t|t}^N, \varphi) - (\tilde{\pi}_{t|t}^N, \varphi) \right)^2 \right]^{1/2} \\ &+ E \left[ \left( (\tilde{\pi}_{t|t}^N, \varphi) - (\pi_{t|t}, \varphi) \right)^2 \right]^{1/2} \end{aligned}$$

oraz z faktu, że  $E((\pi_{t|t}^N, \varphi) | \Sigma_t) = (\tilde{\pi}_{t|t}^N, \varphi)$ . Na podstawie przyjętych definicji oraz założeń szacujemy

$$\begin{aligned} E \left[ \left( (\pi_{t|t}^N, \varphi) - (\tilde{\pi}_{t|t}^N, \varphi) \right)^2 \right] &\leq E \left[ \frac{1}{N^2} \left( \sum_{i=1}^N (N_t^{(i)} \varphi(x_{t|t-1}^{(i)}) - N_t^{(i)} \tilde{w}_t^{(i)} \varphi(x_{t|t-1}^{(i)})) \right)^2 \right] \\ &\leq \frac{1}{N^2} E \left[ \sum_{i=1}^N (N_t^{(i)} - N_t^{(i)} \tilde{w}_t^{(i)}) \left| \varphi(x_{t|t-1}^{(i)}) \right| \right]^2 \\ &\leq C \frac{\max_i \left| \varphi(x_{t|t-1}^{(i)}) \right|^2}{N} \leq C \frac{\|\varphi\|^2}{N} \end{aligned}$$

$$E \left[ \left( (\pi_{t|t}^N, \varphi) - (\pi_{t|t}, \varphi) \right)^2 \right]^{1/2} \leq E \left[ \left( (\pi_{t|t}^N, \varphi) - (\tilde{\pi}_{t|t}^N, \varphi) \right)^2 \right]^{1/2}$$

$$+ E \left[ \left( (\tilde{\pi}_{t|t}^N, \varphi) - (\pi_{t|t}, \varphi) \right)^2 \right]^{1/2} \leq \sqrt{c_{t|t}} \frac{\|\varphi\|}{\sqrt{N}},$$

gdzie  $c_{t|t} = (\sqrt{\tilde{c}_{t|t}} + \sqrt{C})^2$ , co kończy dowód twierdzenia 1.  $\square$

Estymator otrzymany metodą filtru cząsteczkowego wykazuje zbieżność średniokwadratową niezależnie od wymiaru badanego procesu  $n_x$ . Wskaźnik zbieżności szacowanego błędu estymacji  $E \left[ (\pi_{t|t}^N, \varphi) - (\pi_{t|t}, \varphi) \right]^2$  jest rzędu  $\frac{1}{N}$ . Niemniej (dla ustalenia precyzji zbieżności) należy zauważyć, że liczba cząsteczek  $N$  zależy od  $c_{t|t}$ , która z kolei zależy od  $n_x$ . Wadą metody jest fakt, iż w sytuacji, gdy optymalny filtr skojarzony z badanym układem wykazuje długą pamięć (wykazuje zależność od stanu początkowego), następuje kumulacja błędów. To w konsekwencji powoduje wzrost  $c_{t|t}$  w czasie. Wówczas do uzyskania zamierzonej dokładności estymacji  $(\pi_{t|t}^N, \varphi)$  potrzebna jest duża liczba generowanych cząsteczek.

Zbieżność algorytmu dla funkcji nieograniczonych można pokazać przy dodatkowych ograniczeniach nałożonych na funkcje przejścia i gęstości  $g$

Z4. Dla ustalonej realizacji zmiennej obserwowanej  $y_{1:N}$   $g(y_t | x_t) < \infty$ ,  $K(x_t | x_{t-1}) < \infty$ ,  $t = 1, \dots, N$ .

Z5. Dla ustalonej realizacji zmiennej obserwowanej  $y_{1:N}$  funkcja  $\varphi$  spełnia warunek

$$\sup_{x_t} |\varphi(x_t)|^4 g(y_t | x_t) < C(y_{1:t}).$$

Przy powyższych założeniach można pokazać, że czwarty moment estymowanej wartości oczekiwanej jest ograniczony [23].

$$(9.9) \quad E \left| \int \varphi(x_t) \pi_{t|t}^N(dx_t) - \int \varphi(x_t) \pi_{t|t}(dx_t) \right|^4 \leq \frac{C_\varphi}{N^2}$$

gdzie  $C_\varphi$  – zależy od  $\varphi$  oraz od  $y_{1:N}$ .

Korzystając z nierówności Rosenthala nierówność (9.9) można udowodnić dla przypadków bardziej ogólnych, w przestrzeniach  $L^p$ , dla  $p > 1$  [23].

**10. Podsumowanie.** Celem naszej pracy było zainteresowanie czytelnika metodą filtru cząsteczkowego. Prezentowana metoda łączy zagadnienie filtracji stochastycznej z możliwościami metod Monte Carlo. Problem filtracji jest dobrze znany w inżynierii, w teorii przetwarzania sygnałów. W statystycznym wnioskowaniu na temat nieliniowych modeli przestrzeni stanów, gdzie estymacji podlegają warunkowe rozkłady a posteriori  $p(x_{0:T} | y_{1:T})$ ,  $p(x_t | y_{1:t})$ , najczęściej używanymi metodami są algorytm Gibbsa oraz Metropolis i Hastingsa. W praktyce okazuje się, że zastosowanie metody filtru cząsteczkowego, jako sekwencyjnej metody Monte Carlo, w wielu przypadkach może okazać się istotnie efektywniejsze. Aby dokładnie przedstawić

sposób działania metody PF szczegółowo omówiliśmy kolejne etapy algorytmu SIR oraz sposoby zwiększania efektywności jego działania. Większość badań nad filtrem cząsteczkowym dotyczy ustalenia kryteriów wyboru funkcji ważności, metody resamplingu, efektywnego wykorzystanie symulowanej próbki i monitorowanie wydajności próbki. Należy pamiętać, że wiele zagadnień praktycznych nadal wymaga ustalenia podstaw teoretycznych, co może być tematem ciekawych prac. Sekwencyjna metoda Monte Carlo jest atrakcyjną metodą estymacji, pozwalającą na elastyczne modelowanie oraz wprowadzanie realizacji równoległych niezależnie od rodzaju nieliniowości i charakteru rozkładów badanych zmiennych. Obiecujące wyniki, jeżeli chodzi o budowanie bardziej realistycznych i złożonych probabilistyczne modeli, otrzymuje się przez łączenie metody PF z innymi algorytmami np. MCMC, EM.

## 10. Bibliografia

- [1] Anderson B.D.O., Moore J.B., *Optimal Filtering*, Englewood Cliffs, 1979.
- [2] Andrieu Ch., Doucet A., Singh S. S., Tadić V.B, *Particle Methods for Change detection, System Identification, and Control*, Proceeding of the IEEE, **92**, No 3, March 2004.
- [3] Arulampalam S., Maskell S., Gordon N., Clapp T., *A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking*, IEEE Proceedings on Signal Processing, **50** (2): 174-188, 2001.
- [4] Capp O., Godsill S. J., Moulines E., *An overview of existing methods and recent advances in sequential Monte Carlo*, IEEE Proceedings, **95** (5):899-924, 2007.
- [5] Chopin N., *A sequential particle filter for static models*, Biometrika, **89**, 539-551, 2002.
- [6] Chopin N., *Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference*, Ann. Statist., **32**, 2385-2411, 2004.
- [7] Crisan D., Doucet A., *Convergence of sequential Monte Carlo methods*, Technical report, University of Cambridge, CUED/F-INFENG/TR381, 2000.
- [8] Doucet A., *On Sequential Simulation – Based Methods for Bayesian Filtering*, 1998.
- [9] Doucet A, S. Godsill, Andrieu C., *On sequential Monte Carlo sampling methods for Bayesian filtering*, Statistics and Computing, **10**, 197-208, 2000.
- [10] Doucet A., de Freitas J. F. G., Gordon N. J., *Sequential Monte Carlo Methods in Practice*, Springer-Verlang: New York 2001.
- [11] Doucet A., Briers M., Sncal S., *Efficient block sampling strategies for sequential Monte Carlo*, Journal of Computational and Graphical Statistics, **15** (3), 693-711, 2006.
- [12] Fearnhead P., *Sequential Monte Carlo methods in filter theory*, Ph.D. Thesis, Merton College, University of Oxford, 1998.
- [13] Geweke J., *Bayesian inference in econometric models using Monte Carlo integration*, Econometrica **57**, 1317-1339, 1989.
- [14] Gordon N.J, Salmond N.J, Smith A. F. M, *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, IEE Proceedings-F, **140** (2), pp. 107-113, 1993 (first application in signal processing).

- [15] Kitagawa G, Gersch G, *Smoothness Priors Analysis of Time Series*, Springer, Lecture Notes in Statistics, **116**, 1996.
  - [16] Liu J.,S, Chen R., *Sequential Monte Carlo methods for dynamic systems*, Journal of the American Statistical Association, **93** (433), 1032-1044, 1998.
  - [17] Liu J.,S., *Monte Carlo Strategies in Scientific Computing*, Springer Series in Statistics, 2001.
  - [18] Pajor A, *Procesy zmienności stochastycznej SV w bayesowskiej analizie finansowych szeregów czasowych*, Monografie: Prace Doktorskie, Akademia Ekonomiczna w Krakowie, Nr 2, Kraków 2003.
  - [19] Pitt M., K., Shephard N., *Filtering via simulation: auxiliary particle filters*, Journal of the American Statistical association, **94**, 590-599, 1999.
  - [20] Schn T B, *Estimation of Nonlinear Dynamic System- Theory and Applications*, PhD thesis No 998, Department of Electrical Engineering, Linkping University, Sweden, Feb. 2006.
  - [21] Wang X., Chen R., Guo D., *Delayed – Pilot Sampling for Mixture Kalman Filter with Applications in Fading Channels*, IEEE Transactions on Signal Processing, **50**, 241-253, 2002.
  - [22] V.S. Zaritskii, V.B. Setnik, L.I. Shimelevich, *Monte Carlo Technique in Problems of Optimal Data Processing*, Automation and Remote Control, **12**, pp. 95-103, 1975.
  - [23] Xiao-Li Hu, Thomas B. Schn and Lennart Ljung. *A Basic Convergence Result for Particle Filtering*. IEEE Transactions on Signal Processing, **56** (4):1337-1348, Apr. 2008.
- Istnieje strona poświęcona SMC/PF w Cambridge,  
<http://www-sigproc.eng.cam.ac.uk/smc/>

Katarzyna Brzozowska–Rup  
 Katedra Matematyki Politechniki Świętokrzyskiej  
 ul. Tysiąclecia Państwa Polskiego 7  
 25-314 Kielce  
 e-mail: brzozows@poczta.fm

Antoni Leon Dawidowicz  
 Instytut Matematyki Uniwersytetu Jagiellońskiego  
 ul. Prof. Łojasiewicza 6  
 Kraków  
 e-mail: Antoni.Leon.Dawidowicz@im.uj.edu.pl

---

### The field of sequential Monte Carlo methods

**Abstract.** This paper provides an introduction to the field of sequential Monte Carlo methods which are also known as particle filters methods. The best known algorithm to solve the problem of non-linear non-Gaussian filtering is the Extended Kalman Filter (EKF) but in settings where the dynamics are significantly non-linear or the noise intensities are high, the EKF can perform quite poorly. Particle filtering methods are powerful tools for online estimation and tracking in nonlinear and non-Gaussian dynamical systems. The basic idea is to approximate the probability density function  $p(x_{0:t} | y_{1:t}, \theta)$  by a discrete cloud of points  $\{x_{0:t}^{(j)}\}_{j=1}^N$ ,  $t = 1, \dots, T$ , called particles. They commonly consist

of three steps: (1) drawing samples in the state-space of the system, (2) computing proper importance weights of each sample and (3) resampling. These methods are becoming increasingly popular in economics and finance so the objective of this paper is to explain the basic assumptions of the methodology and provide references to relevant literature.

**Keywords:** state-space models, hidden Markov model, optimal filtering, sequential Monte Carlo, sequential importance sampling, resampling.

*(wpłynęło 29 kwietnia 2009 r.)*