状態空間モデルを用いた時系列解析に関する研究

モンテカルロフィルタにおけるリサンプリング方法について

G97H043-9　桑田 修平
指　　導　松嶋 敏泰 教授

1　研究目的

近年, 金融・経済などの分野で得られる時系列データには非ガウス性, 非線形性が見られたり, データ中に急激な構造変化などが見られたりする場合が多く, 状態空間モデルに従来用いられてきたカルマンフィルタなどでは満足な結果が得られなくなった.

このような非線形・非ガウス型の状態空間モデルに有効な解析方法として, モンテカルロフィルタ [1], ブートストラップフィルタ [2] などの非線形フィルタがある.

モンテカルロフィルタは, MCMC法などのように分布を多数の実現値により近似して, 一期先予測・フィルタを繰り返していく逐次型のアルゴリズムであるが, アルゴリズム中にあるリサンプリング方法が分布の予測精度に影響すること指摘されている [3].

そこで本研究では,Kullback-Liebler 情報量（以下,K-L 情報量）用いて分布の近似精度を評価することにより, アルゴリズム中にあるリサンプリング方法を改良し, 予測精度を改善することを目的とする.

2　準備

2.1　定義

本研究では, 以下の式により定義される状態空間モデルを対象とする.

$$x_n = F(x_{n-1}, w_n). \quad \text{(システムモデル)} \qquad (1)$$

$$y_n = H(x_n, v_n). \quad \text{(観測モデル)} \qquad (2)$$

n : 期.
y_n : n 期の観測値.
x_n : n 期の状態.
w_n : n 期の白色雑音（システムノイズ）.
v_n : n 期の白色雑音（観測ノイズ）.
F,H : 非線形の関数.
$Y^n = (y_1, y_2, \ldots, y_n)$.
θ : (1),(2) 式に含まれる未知パラメータの集合.

＜仮定＞

① 状態 x_0 が従う分布 $p_0(x)$ が既知である.

② システムノイズ, 観測ノイズが従う分布 $q(w)$,$r(v)$ がそれぞれ既知である.

③ 状態 x_n と観測値 y_n が与えられると, 観測ノイズ v_n が一意に決まり, y_n に関して微分可能な関数 G を用いて以下のように表される.

$$v_n = G(y_n, x_n) \qquad (3)$$

2.2　問題設定

ここで、観測値 Y^{n-1} が与えられたもとで観測不可の状態 x_n を推定する（$p(x_n|Y^{n-1})$ を推定する）ことを考える. モンテカルロフィルタでは, 状態 x_n の一期先予測分布 $p(x_n|Y^{n-1})$ とフィルタ分布（事後分布）$p(x_n|Y^n)$ を以下のように m 個の実現値で近似する.

$\{p_n^{(1)}, p_n^{(2)}, \ldots, p_n^{(m)}\}$:一期先予測分布.
$\{f_n^{(1)}, f_n^{(2)}, \ldots, f_n^{(m)}\}$:フィルタ分布.

3　従来法

モンテカルロフィルタは, 以下の３つのステップを n に関して繰り返し行うことにより, 逐次的に $p(x_n|Y^{n-1})$,$p(x_n|Y^n)$ を近似的に求めるアルゴリズムである. はじめに, $\{f_{n-1}^{(1)}, f_{n-1}^{(2)}, \ldots, f_{n-1}^{(m)}\}$ が既知であるとする.

＜ＳＴＥＰ１＞

・$q(w)$ から m 個の実現値 $w_n^{(j)} (j = 1, 2, \ldots, m)$ を発生させ, 以下の (4) 式によって一期先予測分布 $p(x_n|Y^{n-1})$ の実現値 $p_n^{(j)} (j = 1, 2, \ldots, m)$ を求める.

$$p_n^{(j)} = F\left(f_{n-1}^{(j)}, w_n^{(j)}\right). \qquad (4)$$

＜ＳＴＥＰ２＞

・以下の (5) 式により, 重み $\alpha_n^{(j)} (j = 1, 2, \ldots, m)$ をそれぞれの $p_n^{(j)} (j = 1, 2, \ldots, m)$ に対して計算する.

$$\alpha_n^{(j)} = p(y_n|p_n^{(j)}) = r(G(y_n, p_n^{(j)})) \left|\frac{\partial G}{\partial y_n}\right|. \qquad (5)$$

＜ＳＴＥＰ３＞(リサンプリング)

・$\alpha_n^{(j)} (j = 1, 2, \ldots, m)$ をそれぞれの $p_n^{(j)} (j = 1, 2, \ldots, m)$ に割り当てて, $p_n^{(j)}$ を $\alpha_n^{(j)}$ の確率でサンプルすることにより, $f_n^{(j)} (j = 1, 2, \ldots, m)$ を求める.

ここで,$p_n^{(j)} (j = 1, 2, \ldots, m)$ が観測値 y_n から大きく外れていた場合には, ＜ＳＴＥＰ３＞により得られる $f_n^{(j)} (j = 1, 2, \ldots, m)$ の, 真のフィルタ分布に対する近似が悪くなってしまい, その結果,$n+1$ 期以降の予測精度が悪くなってしまうことが考えられる [2],[3].

この問題に対処するために, $p_n^{(j)}$ を $l \times m$ 個発生させて, そこから m 個リサンプリングを行う方法 [3] や,

$|v_n^{(j)}|$ がある定められた閾値より大きくなるような $f_{n-1}^{(j)}$ に対して, $n-1$ 期に戻って $f_{n-1}^{(j)}$ をリサンプリングし直す方法 [2] などが考えられているが, 分布自体の近似精度を考慮した方法は考えられていない.

4 提案法

そこで本研究では, 予測分布自体の近似精度にあわせて, 近似精度が悪い場合には,[2] に従って1期前に戻り, [3] に従ってサンプリング数を増やすことにより, リサンプリングをやり直す方法を考える. まず, 分布間の距離を測るK-L情報量 ((6) 式) を用いることを考える.

$$\int_{-\infty}^{\infty} \log \frac{g(y_n)}{f(y_n)} g(y_n) dy_n. \tag{6}$$

$g(y_n)$: 真の分布.

$f(y_n)$: 近似した分布.

(6) 式の定数部分を取り除いた以下の (7) 式の値によって予測分布の良さを比較することができる. この値が大きいほど真の分布に近くなる.

$$p(y_n|Y^{n-1}) \cong \frac{1}{m} \sum_{j=1}^{m} \alpha_n^{(j)} \tag{7}$$

ゆえに,(7) 式の値が小さくなる (n 期の予測分布の近似精度が悪くなる) 場合には, $n-1$ 期に戻り,(7) 式の値に見合った l 個の実現値 $p_{n-1}^{(j)}(j=1,2,\ldots,l)$ を取り, それらの実現値から得られる $f_{n-1}^{(j)}(j=1,2,\ldots,l)$ を用いて n 期の予測分布の実現値 $p_n^{(j)}$ を $m+l$ に増やせば, n 期の予測分布の近似精度は上がると考えられる. よって, ＜ＳＴＥＰ３＞の替わりに l の値を固定した以下の＜ＳＴＥＰ３'①＞を提案する.

＜ＳＴＥＰ３'①＞

・n 期の (7) 式の値が $n-1$ 期の値以上の場合は, 通常どおりの＜ＳＴＥＰ３＞を行う.

・n 期の (7) 式の値が $n-1$ 期の値より小さくなった場合には,$n-2$ 期のフィルタ分布の実現値 $f_{n-2}^{(j)}(j=1,2,\ldots,m)$ に対してシステムノイズを発生させ,$n-1$ 期の予測分布の実現値 $p_{n-1}^{(j)}(j=1,2,\ldots,l)$ を求める ((8) 式).

$$p_{n-1}^{(j)} = F(f_{n-2}^{j}, w_{n-1}^{(j)}). \tag{8}$$

$p_{n-1}^{(j)}(j=1,2,\ldots,m)$ に対して、重み $\alpha_{n-1}^{(j)}(j=1,2,\ldots,l)$ をそれぞれ求め ((9) 式), その重みを用いてリサンプリングを行い $n-1$ 期のフィルタ分布の実現値 $f_{n-1}^{(j)}(j=1,2,\ldots,l)$ を求め, 既に求められた実現値 $f_{n-1}^{(j)}(j=1,2,\ldots,m)$ と合わせて, $f_{n-1}^{(j)}(j=1,2,\ldots,m+l)$ とする.

$$\alpha_{n-1}^{(j)} = p(y_{n-1}|p_{n-1}^{(j)}). \tag{9}$$

$f_{n-1}^{(j)}(j=1,2,\ldots,m+l)$ に対して, ＜ＳＴＥＰ１＞～＜ＳＴＥＰ３＞の操作を行い, $f_n^{(j)}(j=1,2,\ldots,m)$ を求める.

5 シミュレーション結果

提案法との比較を行った例を以下に載せる (例は従来法との比較). 今回は, 次式により得られる人工データ 500 個を用いた.

$$y_n \sim N(5,2). \tag{10}$$

推定に用いるモデルは次式を用い, $m=1000, l=10$ (固定値) として, 未知パラメータ $\theta=\beta$ を状態に含めて推定を行った. ここで,$N(a,b)$ は平均 a, 分散 b の正規分布,$U(c,d)$ は (c,d) の一様分布を表す.

$$x_n = x_{n-1} + w_n. \quad w_n \sim N(0,\beta). \tag{11}$$

$$y_n = x_n + v_n. \quad v_n \sim U(0,2). \tag{12}$$

$$x_0 \sim N(4,1). \quad \beta \sim U(0,2). \tag{13}$$

２つの方法により得られた平均２乗誤差を以下に示す.

表１　従来法と提案法の比較

	平均2乗誤差
従来法	2.9978262
提案法	2.42376007

6 考察

従来法に比べて提案法の方 (7) 式の値が大きくなり, フィルタ分布の近似精度の悪化を防いだ結果, 次期の予測分布の近似精度が上がったことが分かった. つまり, 平均２乗誤差も小さくなった.

ここで, 今回は付け足す実現値の数 l を固定値としたが, (7) 式の値の大きさに合わせて l の値を変化させてリサンプリングを行う方法＜ＳＴＥＰ３'②＞も考えられるが, その方法については, 現在考察中である.

7 まとめと課題

今回は, 分布の近似精度を評価し, 近似精度が悪い場合には過去の期に戻り, サンプル数を変化させてリサンプリングをし直すことにより, 予測分布の近似精度の改善を行った. 今後は, 非線形・非ガウス型の状態空間モデルに対しても同様のシミュレーションを行う.

参考文献

[1] G.Kitagawa,"Monte Calro filter and smoother for non-gaussian nonlinear state space models", J.Computational and Graphical Statistics,5,p.1-25,1996

[2] N.J.Gordon et al.,"Novel approach to nonlinear/non-Gaussian Bayesian state estimation", IEE PROCEEDINGS-F,Vol.40,No.2,p.107-113,1993.4

[3] 樋口 知之,"モンテカルロフィルタ (ＭＣＦ) におけるランダムサンプリング", ハイパフォーマンスコンピューティング, 54-2,p.7-14,1994