

状態空間モデルを用いた時系列解析に関する一考察

- 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 -

早稲田大学 *栗田修平, 吉田隆弘, 松嶋敏泰

1 はじめに

金融・経済などの時系列データの多くは、ガウス性や線形性が仮定されないため、データにそれらの仮定を置いて解析を行うカルマンフィルタ等の方法では満足な予測が行えない場合が多い。

それに対して、近年では、非線形性・非ガウス性を仮定した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]などの非線形フィルタ[2]が提案されている。モンテカルロフィルタは、分布をその分布に従って発生させたと見なせる多数の実現値によって近似するモンテカルロ法の1つであるが、多数の実現値による分布の近似精度が悪くなれば、予測精度も悪くなるという問題がある[3], [4]。

そこで本研究では、分布を精度良く表現するための多数の実現値による分布の近似法を提案し、従来よりも精度の高い予測を行うことを目的とする。

2 状態空間モデルと事後分布

本稿で用いる状態空間モデル[1]は、以下の2式により定義される。

$$\begin{aligned} x_t &= f(x_{t-1}, v_t), &< \text{システムモデル} > & (1) \\ y_t &= g(x_t, w_t). &< \text{観測モデル} > & (2) \end{aligned}$$

ここで、 $t (t = 1, 2, \dots)$ は時点を表し、

- x_t : t 時点の状態 (観測不可)。
- y_t : t 時点の観測値。
- v_t : t 時点の白色雑音。確率分布 $p(v)$ に従う。
- w_t : t 時点の白色雑音。確率分布 $p(w)$ に従う。

である。また、 f, g は関数を表し、状態 x_0 の分布は既知であるとする。

今、 T 時点までの観測値 $Y_T = \{y_1, \dots, y_T\}$ が与えられたもとで、 $T+1$ 時点の状態 x_{T+1} の条件付分布 $p(x_{T+1}|Y_T)$ を予測することを考える。 $p(x_{T+1}|Y_T)$ は、

$$p(x_{T+1}|Y_T) = \int p(x_{T+1}|x_T) p(x_T|Y_T) dx_T \quad (3)$$

により求まる。また、(3)式右辺の事後分布と呼ばれる分布 $p(x_T|Y_T)$ は、ベイズの定理により、

$$p(x_T|Y_T) = \frac{p(y_T|x_T)p(x_T|Y_{T-1})}{\int p(y_T|x_T)p(x_T|Y_{T-1})dx_T} \quad (4)$$

から求まる。ここで、関数 f, g が線形であり、かつ、 $p(v), p(w)$ 、状態 x_0 の分布に対してガウス性の仮定を置いた場合には、カルマンフィルタが適用可能となり、(3), (4)式の積分は解析的に解ける。

3 モンテカルロフィルタ

モンテカルロフィルタ[1]は、状態 x_t の分布などをその分布自身から発生させた m 個の実現値の集合で表現することにより、(3), (4)式の積分計算を近似的に行うアルゴリズムである。ここで、分布 $p(x_t|Y_{t-1})$ を集合 $\{x_{t-1}^{(1)}, \dots, x_{t-1}^{(m)}\}$ 、分布 $p(x_t|Y_t)$ を集合

$\{x_{t-1}^{(1)}, \dots, x_{t-1}^{(m)}\}$ 、分布 $p(v)$ を集合 $\{v_t^{(1)}, \dots, v_t^{(m)}\}$ によって表現するものとする。

モンテカルロフィルタは以下の2つのステップからなる。今、状態 x_0 の分布を近似する実現値の集合 $\{x_{0|0}^{(1)}, \dots, x_{0|0}^{(m)}\}$ は既知であり、 $\{x_{t-1|t-1}^{(1)}, \dots, x_{t-1|t-1}^{(m)}\}$ を既に保持しているものとする。

[モンテカルロフィルタ]	
ステップ1 $v_t^{(i)} (i = 1, \dots, m)$ を $p(v)$ に従う乱数から発生させ、(5)式より $x_{t t-1}^{(i)} (i = 1, \dots, m)$ を得る。	$x_{t t-1}^{(i)} = f(x_{t-1 t-1}^{(i)}, v_t^{(i)}) \quad (5)$
ステップ2 (6)式により、 $x_{t t-1}^{(i)} (i = 1, \dots, m)$ の重み $a_t^{(i)} (i = 1, \dots, m)$ を求め、重みに比例した確率で $\{x_{t t-1}^{(1)}, \dots, x_{t t-1}^{(m)}\}$ の中から m 個サンプリングして、選ばれた実現値を $x_{t t}^{(i)} (i = 1, \dots, m)$ とする。	$a_t^{(i)} = p(y_t x_{t t-1}^{(i)}) \quad (6)$

この2つのステップを繰り返すことにより、逐次的に $x_{t|t}^{(i)}, x_{t+1|t}^{(i)}, x_{t+2|t}^{(i)}, \dots (i = 1, \dots, m)$ が求まる。ステップ2のサンプリングを行う操作をリサンプリングと呼び、乱数を用いて m 個の集合 $\{x_{t|t-1}^{(1)}, \dots, x_{t|t-1}^{(m)}\}$ の中から重複を許して m 個サンプリングを行う。

4 従来研究

リサンプリングを繰り返していくと、時点が進むにつれて、分布を表現する実現値 $x_{t|t}^{(i)}, x_{t+1|t}^{(i)}$ の種類(多様性)が無くなっていくこと[2]や、 $\{x_{t|t-1}^{(1)}, \dots, x_{t|t-1}^{(m)}\}$ による予測が観測値 y_t と外れた時にステップ2で求まる集合 $\{x_{t|t}^{(1)}, \dots, x_{t|t}^{(m)}\}$ による事後分布の近似精度が悪くなることが指摘されており[3]、従来[1]のリサンプリング方法を改良する研究が行われている[3], [4]。

ここで、モンテカルロフィルタにおける実現値の集合は、図1のように m 個の点で $1/m$ の確率をもつ離散分布によって分布を近似して表現していると言える。

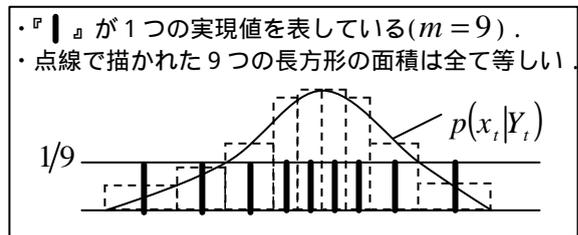


図1 9個の実現値による分布の近似

この考え方からステップ2を見直すと、ステップ2とは、事後分布 $p(x_t|Y_t)$ の近似分布として、ステップ1で求めた m 個の点で $a_t^{(i)} (i = 1, \dots, m)$ を正規化した値の確率をもつ離散分布を、 m 個の点で $1/m$ の確率をもつ離散分布によって近似して表現するために、 m 個の実現値を求め直している操作であると言える。

従来法[1]では、この m 個の点を $a_t^{(i)} (i = 1, \dots, m)$ に比例する確率に従う乱数を用いてリサンプリングしていたが、ステップ2を行う目的が事後分布 $p(x_t|Y_t)$ を近似する各点で $1/m$ の確率をもつ離散分布を求め

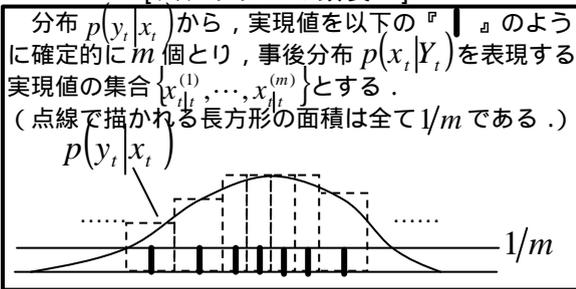
ることであるので、必ずしも乱数を用いる必要はない。

そこで本研究では、事後分布 $p(x_t|Y_t)$ を近似する実現値の集合を、乱数を用いずに確定的に求める（図1のようなヒストグラムの m 個の代表点を、計算により求める）方法を提案する。確定的に実現値を求めると、実現値の多様性が減ることや、予測が外れた時に事後分布の近似精度が悪くなることを防ぐことができると考えられる。

5 提案法

ここで、重みを求める際に用いた分布 $p(y_t|x_t)$ を用いて、分布 $p(y_t|x_t)$ から、図1のように確定的に m 個の実現値をとり、得られた実現値で事後分布を表現する実現値の集合 $\{x_{t|t}^{(1)}, \dots, x_{t|t}^{(m)}\}$ とする方法（ステップ2の改良）を提案する。ただし、ステップ1は従来法[1]と同様である。

【ステップ2の改良】



ここでさらに、実現値の集合 $\{x_{t|t-1}^{(1)}, \dots, x_{t|t-1}^{(m)}\}$ とその重み $a_t^{(i)}$ ($i=1, \dots, m$) を用いることにより、事後分布 $p(x_t|Y_t)$ が、以下のような経験分布関数、

$$\frac{1}{\sum_{j=1}^m a_t^{(j)}} \sum_{i=1}^m a_t^{(i)} I(x, x_{t|t-1}^{(i)}) \quad (7)$$

により近似されることを考え、(7)式から得られる情報も用いて事後分布を近似することを考える。ただし、 $I(x, a)$ は、 $x < a$ の時 $I(x, a) = 0$ 、その他の時 $I(x, a) = 1$ となる関数である。(7)式の情報を用いることにより、事後分布 $p(x_t|Y_t)$ の近似精度は良くなると考えられる。

具体的には、改良の方法により求まる実現値の集合 $\{x_{t|t}^{(1)}, \dots, x_{t|t}^{(m)}\}$ の平均値が、(7)式により近似される事後分布 $p(x_t|Y_t)$ の平均値 $\{E_t^{POS}$ とおく) に一致するように、 $\{x_{t|t}^{(1)}, \dots, x_{t|t}^{(m)}\}$ を $\{x'_{t|t}^{(1)}, \dots, x'_{t|t}^{(m)}\}$ へと変換する方法（ステップ2の改良）を提案する。

【ステップ2の改良】

改良により求まる実現値の集合 $\{x_{t|t}^{(1)}, \dots, x_{t|t}^{(m)}\}$ を、(8)式により $\{x'_{t|t}^{(1)}, \dots, x'_{t|t}^{(m)}\}$ へと変換する。ここで、 E_t^{POS} は(9)式により求める。

$$x'_{t|t}^{(i)} = x_{t|t}^{(i)} + (E_t^{POS} - \frac{\sum_{i=1}^m x_{t|t}^{(i)}}{m}) \quad (8)$$

$$E_t^{POS} = \frac{\sum_{i=1}^m a_t^{(i)} x_{t|t-1}^{(i)}}{\sum_{j=1}^m a_t^{(j)}} \quad (9)$$

6 シミュレーションと結果

シミュレーション結果の例として以下に2つ載せる。

$$x_t = x_{t-1} + v_t, \quad v_t \sim N(0, 6), \quad (10)$$

$$y_t = x_t + w_t, \quad w_t \sim N(0, 1). \quad (11)$$

以上の2式から時系列データ y_t を500個作成する ($x_0 \sim N(0, 2)$ である)。ここで、 $\{x_{t|t-1}^{(1)}, \dots, x_{t|t-1}^{(m)}\}$ の平均値と y_t との平均2乗誤差により従来法[1]と提案法との比較を行う(実験)。2例目として、(10), (11)式において、 $v_t \sim N(0, 2)$, $w_t \sim N(0, 5)$, $x_0 \sim N(0, 3)$ とした場合の結果を以下に載せる(実験)。

表1 シミュレーション結果(実験)

実験	平均2乗誤差(m=10)	平均2乗誤差(m=100)
従来法[1]	9.756983	8.281971
提案法(改良)	8.872578	8.267287
提案法(改良)	10.111279	8.273681

表2 シミュレーション結果(実験)

実験	平均2乗誤差(m=10)	平均2乗誤差(m=100)
従来法[1]	9.859728	8.852446
提案法(改良)	11.134615	10.642045
提案法(改良)	9.773427	8.791337

7 考察

改良は、 $p(w)$ の分散の値に予測精度が大きく影響されることが表1、表2から分かる。これは、分布 $p(y_t|x_t)$ の分散が、 $p(w)$ の分散の値によって変化し、事後分布の近似精度が変化してしまうためであると考えられる。また、改良は、(9)式の値 E_t^{POS} の近似精度に依存して予測精度が変化すると考えられ、実際に、実現値の数 m が小さい場合には予測精度が落ちることが表1より確認できる。

8 まとめ

本研究では、ステップ2において求まる事後分布を近似する実現値の集合を、乱数を用いずに確定的に求める方法を提案することにより事後分布の近似精度を良くして、予測精度を上げることを考えた。

今後は、本稿で用いた方法以外にも確定的に実現値を求める方法が考えられるので、実現値の求め方の違いが予測精度にどう影響するか見ていきたい。

参考文献

- [1] 北川 源四郎, "一般化状態空間モデルと自己組織化の方法", 人工知能学会誌, vol.16, no.2, pp.300-307, 2001.
- [2] N.J. Gordon, et al, "Novel approach to non-linear/non-Gaussian Bayesian state estimation", IEE PROCEEDINGS-F, vol.140, no.2, pp.107-113, 1993.
- [3] T. Higuchi, "MONTE CALRO FILTER USING THE GENERIC ALGORITHM OPERATIONS", J. Statist. Comput. Simul., vol.59, pp.1-23, 1997.
- [4] A. Doucet, et al, "Sequential Monte Carlo Methods to Train Neural Network Models", Neural Computation, vol.12, pp.955-993, 2000.