

ハミルトン「時系列解析」の前半から後半へ

目次

参考文献	1
ハミルトン「時系列解析」の前半から後半へ	1
ベイズ推測へ	2
カルマンフィルタ——みえないものを追いかける手続き	4

参考文献

- [1] James D. Hamilton. Time Series Analysis. Princeton University Press, 1994.
- [2] 沖本竜義. 経済・ファイナンスデータの計量時系列分析. 朝倉書店, 2010.

ハミルトン「時系列解析」の前半から後半へ



ハミルトン「時系列解析」[1]の前半をふりかえってみます。

- 1章では差分方程式を導入しました，そもそも時系列というのは各時点の値がその時点より過去の値に左右されているような値の列ですね。「現時点の値」を「過去何ステップかの値」の式で表す差分方程式はそれを最も素朴に表現したモデルといえるでしょう——もっとも差分方程式といったときにはノイズのない決定論的なモデルになりますが，システムが安定（任意の時刻で値が有界）である条件も導出しましたね．つまり， p 次の差分方程式が安定である条件は，それを 1 次のベクトル差分方程式で表現したときの係数行列の全ての固有値の絶対値が 1 未満であることでした．
- 2章ではラグ演算子——いわば，入力された時系列の 1 ステップ昔を覗くメガネ——を導入しました．私たちはラグ演算子によって「現在の値」を「一番最初の値」の式でかくことが可能になりました．この形式もまた時系列の性質を知るのに有用です．1 章とは別のアプローチで同じ安定条件を導出しましたね．
- 3章では定常性，エルゴード性を導入し，ARMA 過程を導入しました．
- 4章では ARMA 過程の将来の値を予測しました．「この時系列の値は将来どうなるの？」というのは私たちが目の前の時系列に対して最も興味があることの一つでしょう．この章ではいよいよその間に，過去の値を全て観測している理想的な場合，過去 m ステップしか観測できていない場合の順で取り組みました．何次のモデルを仮定すべきかに関連して，Box-Jenkins 法にも触れましたね．
- 5章では観測値から ARMA 過程そのものを推定する方法を扱いました．正規ホワイトノイズを仮定して尤度を最大化するという手続きによりました．

- 6 章ではそれまで自己共分散関数によって時系列を特徴付けていたのとうってかわって、母集団スペクトルによって時系列を特徴付ける流儀を扱いましたね。もっとも両者はどちらかが特定されていればもう片方にかき換えられるわけですが。
- 7 章では漸近分布論を扱いました。これは標本サイズが大きくなったときに推測や検定の結果がが想定通りに正しくなっていくかを調べるとき必要な概念ですね。
- 8 章は既に扱っていた線形回帰モデルの最小 2 乗推定の性質を掘り下げました。まずオーソドックスな最小 2 乗法を扱い、一般化最小 2 乗法を扱いましたね。
- 10 章では単変量 ARMA 過程を多変量 ARMA 過程に拡張しました。
- 時系列が多変量になったとき私たちは「この変数とこの変数は関係しているの？」という興味を抱くでしょう。11 章では多変量 ARMA 過程を用いてそのような問いに答えていきました。グレンジャー因果検定やインパルス応答関数を取り扱いましたね。



——こうしてみると、前半を終えただけでも私たちは目の前の時系列に対して既に色々なことが推測できるように思えます。将来の値を予測したり、変数間にグレンジャー因果があるかを推測したり、あるとしたらその効果の大きさはどれくらいなのか推測したり——ただし往々にしてノイズが正規ホワイトノイズであるような ARMA 過程を仮定することにはなりますが。

となると、12 章からは何を扱うのでしょうか？やはり仮定を緩和していくとか？



それも後の章では出てくるけど、12 章でまずやることは違うかな。12 章でやることは、12 章の冒頭から天下りの的にいえば、「推測に事前知識を反映させること」みたいだからね。



事前知識？

ベイズ推測へ



そうだな、12 章冒頭の例だけど、任意の時刻 t に $y_t \sim \text{i.i.d } N(\mu, \sigma^2)$ が観測されるとするよ。 $y = (y_1, \dots, y_T)^\top$ を観測したときに $\theta = (\mu, \sigma^2)^\top$ をどのように推測すべきだろうか？



そのモデルはもはや各時刻の値が独立なので全く時系列らしくないですが……しかし、そのような推測は 5 章で扱いましたよ。そのモデルの下で時刻 $1, \dots, T$ にどんな値が観測されるかは確率ベクトル $Y = (Y_1, \dots, Y_T)^\top$ です。 Y の確率密度関数は $\theta = (\mu, \sigma^2)^\top$ を用いてかけます。実際には $y = (y_1, \dots, y_T)^\top$ が観測されているのでこの点で Y の確率密度関数 (y の関数でなく θ の関数と考えると尤度関数) が最大になってほしいです。そのようになる $\hat{\theta}$ を選びます…というものでしたよね。最尤推定です。



そうだね。その選び方がどれくらいよいかは、たくさんのパラレルワールドで y を観測してそれぞれで $\hat{\theta}$ を推定したときに、平均的にどれくらい真の θ とずれるか $E(\hat{\theta} - \theta)(\hat{\theta} - \theta)^\top$ で評価されるね。

じゃあ少し違う状況を考えるよ。いやモデルは同じなんだけど、観測者は事前に θ に対してある程度の知識というか信念をもっていたとする。「 μ がゼロ未満であることはほとんどないと思う」といった具合に。



なんと、事前に知識が……いえでも、現実にはそのように事前にある程度の見通しがある状況もありそうですね。しかし、最尤推定ではそのような知識を取り込むことは難しそうですね。尤度関数の最大点を取るだけですから……「 μ がゼロ未満であることはほとんどないと思う」と考えているなら、 μ がゼロ未満である場所で尤度関数を薄くしておく必要がありそうです。しかし、どのように薄くすればよいのやら。それに、そうやって尤度関数の最大点が少しずれたとして、その点をどれくらい信用してよいのか……。



うん、もはや、「観測したデータをもとにただ 1 点の推測値 $\hat{\theta}$ を出す」という枠組みだと扱いづらくなってくるよね。だから、ここでは θ も確率変数であると考えよ。 θ に対して事前に想定している確率分布 $f(\theta)$ があって、観測したデータ y をもとにその確率分布を $f(\theta|y)$ に更新する、という手続きをとる。



θ も確率変数であると考えろ？……これまでは神様がモデルというサイコロを振って観測値が出てくると考えていましたが、そのモデル (のパラメータ)すらもサイコロを振った結果で決まっていたということなのですか？では、モデルから観測値を出す神様は実は下請けの神様にパラメータを発注していて、下請けの神様がサイコロを振ってパラメータを出して納品していたと……神様界にもそんな下請け構造が……。



ちょっと意味がわからないかな。

カルマンフィルタ——みえないものを追いかける手続き



12 章でベイズ推測を導入して、13 章はカルマンフィルタ？ 何ですかそれは？



カルマンフィルタとは「各時刻の観測値が、観測できない『状態』から生成されるというモデルを仮定した下で、状態の分布をトラッキングする手続き」かな。

仮定するモデルについてもっとちゃんと説明するよ。状態空間モデルというんだけど、各時刻の観測値 $y_t \in \mathbb{R}^n$ は以下のモデルから生成されると考えるよ。

$$\xi_{t+1} = F\xi_t + v_{t+1} \quad (13.1.1)$$

$$y_t = A^\top x_t + H^\top \xi_t + w_t \quad (13.1.2)$$

上式のうち $v_t \in \mathbb{R}^r$ と $w_t \in \mathbb{R}^n$ はそれぞれ 0 次の自己共分散行列が Q, R であるようなベクトルホワイトノイズだ。ベクトルホワイトノイズだから 1 次以上の自己共分散行列は零行列ね。

ノイズ以外の部分は……2 つ目の (13.1.2) 式から説明すると、「 y_t は x_t の線形変換と ξ_t の線形変換の和からできる」という式だよ。このうち $x_t \in \mathbb{R}^k$ は予め決定している外生変数だ。例えば y_t が日ごとのデータだとしたら x_t は曜日を one-hot エンコーディングしたベクトルとかね。この x_t はない教科書が多いと思うんだよ。a. 曜日ごとの周期性とかが必要な場合でも ξ_t に押し付けることができると思うし。

それで肝心なのは $\xi_t \in \mathbb{R}^r$ の方だ。これが観測できない「状態」。観測値を生み出す特徴ベクトルといった方がわかりやすいかな？ この ξ_t が時間発展すると考えるよ。1 つ目の (13.1.1) 式がそれだね。初期状態 ξ_0 の分布は適当に決め打っておくことになるよ。

^a 例えば以下の教科書がそうである。

- G. Petris 他, R によるベイズ動的線型モデル, 朝倉書店, 2013.
- 樋口知之 他, データ同化入門 (予測と発見の科学), 朝倉書店, 2019.
- 馬場真哉, 時系列分析と状態空間モデルの基礎: R と Stan で学ぶ理論と実装, プレアデス出版, 2018.



観測できない「状態」が観測値を生み出す……そうですね、例えば私が毎日動物園にパンダさんのようすをみにいくとします。私に観測できるのは、パンダさんが遊び回っているか、そっぽを向いているかのどちらくらいでしょう。しかし、パンダさんのそのようすの裏には「今日はいい気分だ」「今日は疲れがたまっている」「今日はいらいらしている」「今日は寂しい」などのもっと複雑な「状態」があるに違いありません。そのようなパンダさんの状態の分布の推移をこそモデリングしようというのが「状態空間モデル」なのですね？



う、うん……パンダの毎日のふるまいを分析する用途に適切かどうかは別として、状態空間モデルの考え方はそれで合っていると思うよ。

まあそれで重要なのは、状態空間モデルは以下が満たされるとき状態 ξ_t の分布を完全に解析的に追いかけることができるということなんだよね。

- ・ 初期状態 ξ_0 が多変量正規分布にしたがう。
- ・ ベクトルホワイトノイズ v_t, w_t が多変量正規分布にしたがう。
- ・ 変換 F, H が線形変換である (教科書の書き方では転置をとっているし最初から行列であることが前提になっているけどね)。

このとき ξ_t の分布を追いかける手続きをカルマンフィルタとよぶ。ちなみにこのときのモデルを特に動的線型モデルというよ^a。

^a G. Petris 他, R によるベジアン動的線型モデル, 朝倉書店, 2013. の 41 ページ。



状態の分布を完全に追いかけることができる……とは？



初期状態 ξ_0 の分布は決め打っておくとして、時刻 t までに観測できる情報 $y_t \equiv (y_t, \dots, y_1, x_t, \dots, x_t)$ が手に入ったとき、現在の状態 ξ_t の分布が完全に求まるってことだね。確率密度関数 $p(\xi_t | y_t)$ を具体的に求めてみれば、きちんと求まることがわかると思うよ。



$p(\xi_t | y_t)$ を求めるのですか……といわれましても、どうやって求めればいいのかやら……とりあえず、ベイズの定理を適用してみましょう。 y_t というかたまりは y_t, x_t, y_{t-1} とほぐしてみましょう。観測値 y_t を切り出したいので。条件付き分布の項からは、なくてもいい条件は取り除いておきましょう。

$$\begin{aligned} p(\xi_t | y_t) &= p(\xi_t | y_t, x_t, y_{t-1}) \\ &= \frac{p(y_t | \xi_t, x_t, y_{t-1}) p(\xi_t | x_t, y_{t-1})}{p(y_t | x_t, y_{t-1})} \\ &= \frac{p(y_t | \xi_t, x_t) p(\xi_t | y_{t-1})}{p(y_t | x_t, y_{t-1})} \end{aligned} \quad (\text{ア})$$

上式をみると……分母は正規化定数と考えればいいので放っておきましょう。分子の $p(y_t | \xi_t, x_t)$ は (13.1.2) 式から直ちに求まりますね。だって、 ξ_t と x_t が確定している下では y_t の確率的な成分はベクトルホワイトノイズ w_t だけですから。よってこのときの y_t の分布は以下の多変量正規分布です。

$$y_t | \xi_t, x_t \sim N(A^\top x_t + H^\top \xi_t, R) \quad (\text{イ})$$

となると、分子の $p(\xi_t | y_{t-1})$ の方が問題ですね。(13.1.1) 式から直ちに求まるの

は $p(\xi_t|\xi_{t-1})$ であって $p(\xi_t|y_{t-1})$ ではありませんので。うーん…… ξ_t, ξ_{t-1} の同時密度関数 $p(\xi_t, \xi_{t-1}|y_{t-1}) = p(\xi_t|\xi_{t-1}, y_{t-1})p(\xi_{t-1}|y_{t-1}) = p(\xi_t|\xi_{t-1})p(\xi_{t-1}|y_{t-1})$ を ξ_{t-1} について積分して周辺化してみましょうか……。

$$p(\xi_t|y_{t-1}) = \int p(\xi_t|\xi_{t-1})p(\xi_{t-1}|y_{t-1})d\xi_{t-1} \quad (\text{ウ})$$

……まとめると、 ξ_t の条件付き密度関数 $p(\xi_t|y_t)$ は (ア) になるので、これを求めるには (イ) と (ウ) の密度関数の積を正規化すればいいわけですが、(ウ) に 1 つ前の時刻の条件付き密度関数 $p(\xi_{t-1}|y_{t-1})$ が出てきてしまいましたね……。



それで問題ないよ。最初の時刻の $p(\xi_1|y_1)$ がどうなるか考えてみて。先にここていくつかの表記を導入しておこうか。 y_τ が手に入っている条件下での ξ_t の期待値と分散共分散行列を以下のように表記することにしよう。

$$\begin{aligned}\hat{\xi}_{t|\tau} &\equiv E(\xi_t|y_\tau) \\ P_{t|\tau} &\equiv E[(\xi_t - \hat{\xi}_{t|\tau})(\xi_t - \hat{\xi}_{t|\tau})^\top]\end{aligned}$$

初期状態 ξ_0 の分布は $N(\hat{\xi}_{0|0}, P_{0|0})$ としようか。



はあ…… $p(\xi_1|y_1)$ がどうなるかというと、(ア) に当てはめれば $p(y_1|\xi_1, x_1)$ と $p(\xi_1)$ の積の正規化ですね。前者は (13.1.2) 式より以下ですね。

$$y_1|\xi_1, x_1 \sim N(A^\top x_1 + H^\top \xi_1, R) \quad (\text{イ'})$$

$p(\xi_1)$ はまだ何も観測しない下での ξ_1 の密度関数ですが、(13.1.1) 式より $\xi_1 = F\xi_0 + v_1$ ですから、これは多変量正規分布にしたがうベクトルの線形変換の和であり、多変量正規分布にしたがいますね^a。周辺化するまでもないです。

$$\begin{aligned}\xi_0 &\sim N(\hat{\xi}_{0|0}, P_{0|0}) \\ F\xi_0 &\sim N(F\hat{\xi}_{0|0}, FP_{0|0}F^\top) \\ v_1 &\sim N(0, Q) \\ \xi_1 = F\xi_0 + v_1 &\sim N(F\hat{\xi}_{0|0}, FP_{0|0}F^\top + Q) \equiv N(\hat{\xi}_{1|0}, P_{1|0}) \quad (\text{ウ'})\end{aligned}$$

となると、(イ') と (ウ') の密度関数が具体的にわかるので積をとることができますね。それぞれ多変量正規分布ですが正規化因子はさしあたり無視しましょう。 ξ_1

に依存しない、密度を定数倍しかしない項も無視していきましょう。

$$\begin{aligned}
p(\xi_1|y_1) &\propto \exp\left(-\frac{1}{2}(y_1 - A^\top x_1 - H^\top \xi_1)^\top R^{-1}(y_1 - A^\top x_1 - H^\top \xi_1)\right) \\
&\quad \times \exp\left(-\frac{1}{2}(\xi_1 - \hat{\xi}_{1|0})^\top P_{1|0}^{-1}(\xi_1 - \hat{\xi}_{1|0})\right) \\
&\propto \exp\left(-\frac{1}{2}\xi_1^\top \left(HR^{-1}H^\top + P_{1|0}^{-1}\right)\xi_1 \right. \\
&\quad \left. + \frac{1}{2}\left((y_1 - A^\top x_1)^\top R^{-1}H^\top + \hat{\xi}_{1|0}^\top P_{1|0}^{-1}\right)\xi_1 \right. \\
&\quad \left. + \frac{1}{2}\xi_1^\top \left(P_{1|0}^{-1}\hat{\xi}_{1|0} + HR^{-1}(y_1 - A^\top x_1)\right)\right) \quad (\text{工})
\end{aligned}$$

(工) は平方完成して多変量正規分布 $\exp\left(-\frac{1}{2}(\xi_1 - \hat{\xi}_{1|1})^\top P_{1|1}^{-1}(\xi_1 - \hat{\xi}_{1|1})\right)$ にできますね。 $P_{1|1}^{-1} = HR^{-1}H^\top + P_{1|0}^{-1}$ ですから、係数比較で平均ベクトル $\hat{\xi}_{1|1}$ を求めましょう。

$$\begin{aligned}
p(\xi_1|y_1) &\propto \exp\left(-\frac{1}{2}(\xi_1 - \hat{\xi}_{1|1})^\top \left(HR^{-1}H^\top + P_{1|0}^{-1}\right)(\xi_1 - \hat{\xi}_{1|1})\right) \\
&\propto \exp\left(-\frac{1}{2}\xi_1^\top \left(HR^{-1}H^\top + P_{1|0}^{-1}\right)\xi_1 \right. \\
&\quad \left. + \frac{1}{2}\left(\hat{\xi}_{1|1}^\top (HR^{-1}H^\top + P_{1|0}^{-1})\right)\xi_1 \right. \\
&\quad \left. + \frac{1}{2}\xi_1^\top \left((HR^{-1}H^\top + P_{1|0}^{-1})\hat{\xi}_{1|1}\right)\right) \quad (\text{オ})
\end{aligned}$$

これで(オ)と(工)を比較すれば $\hat{\xi}_{1|1}$ が求まるはずです^b。

$$\begin{aligned}
&\hat{\xi}_{1|1}^\top (HR^{-1}H^\top + P_{1|0}^{-1}) = (y_1 - A^\top x_1)^\top R^{-1}H^\top + \hat{\xi}_{1|0}^\top P_{1|0}^{-1} \\
\Rightarrow \hat{\xi}_{1|1}^\top &= \left((y_1 - A^\top x_1)^\top R^{-1}H^\top + \hat{\xi}_{1|0}^\top P_{1|0}^{-1}\right)(HR^{-1}H^\top + P_{1|0}^{-1})^{-1} \\
\Rightarrow \hat{\xi}_{1|1} &= (HR^{-1}H^\top + P_{1|0}^{-1})^{-1}\left((y_1 - A^\top x_1)^\top R^{-1}H^\top + \hat{\xi}_{1|0}^\top P_{1|0}^{-1}\right)^\top \\
\Rightarrow \hat{\xi}_{1|1} &= (HR^{-1}H^\top + P_{1|0}^{-1})^{-1}\left(HR^{-1}(y_1 - A^\top x_1) + P_{1|0}^{-1}\hat{\xi}_{1|0}\right) \quad (\text{カ})
\end{aligned}$$

求まりました！つまり、 $p(\xi_1|y_1)$ の分布は共分散行列が $P_{1|1}^{-1} = HR^{-1}H^\top + P_{1|0}^{-1}$ で平均ベクトルが(カ)である多変量正規分布です！……初期状態の分布を多変量正規分布としましたが、時刻1にかけて時間発展し、時刻1の観測を受けてベイズの定理を適用してもなお、状態は多変量正規分布にしたがうのですね！

^a 多変量正規分布にしたがう確率ベクトルは線形変換しても多変量正規分布にしたがいますし、多変量正規分布にしたがう確率ベクトルどうしを足しても多変量正規分布にしたがいます。多変量正規分布のモーメント母関数を用いて証明できます。

^b (オ)と(工)の2行目を比較していますが3行目を比較すれば一発で(カ)にたどり着きます。式変形では観測ノイズの共分散行列 R が対称行列であること、対称行列の逆行列も対称行列である

こと、転置の線形性などを利用しています、



そうだね、それに、部長がいま $p(\xi_1|y_1)$ を求めたのと全く同じ要領で、任意の t に対する $p(\xi_t|y_t)$ が求まるよ。



え？ あっ、なるほど、いま $p(\xi_1|y_1)$ を求めたことで以下がわかりましたね。

- 初期状態 ξ_0 の分布を多変量正規分布 $N(\hat{\xi}_{0|0}, P_{0|0})$ とする。
- すると時刻 1 の観測をする前の ξ_1 の分布も多変量正規分布になる。(ウ')
- すると時刻 1 の観測をした後の ξ_1 の分布も多変量正規分布になる。(エ)(オ)(カ)

となると次の時刻も同じことが繰り返せます。

- 時刻 1 の観測をした下での ξ_1 の分布は多変量正規分布である。
- すると時刻 2 の観測をする前の ξ_2 の分布も多変量正規分布になる。
- すると時刻 2 の観測をした後の ξ_2 の分布も多変量正規分布になる。

つまり任意の時刻 t に対して $p(\xi_t|y_{t-1})$, $p(\xi_t|y_t)$ は平均ベクトルと共分散行列が以下の多変量正規分布です！

$$\hat{\xi}_{t|t-1} = F\hat{\xi}_{t-1|t-1}$$

$$P_{t|t-1}^{-1} = FP_{t-1|t-1}^{-1}F^\top + Q$$

$$\hat{\xi}_{t|t} = (HR^{-1}H^\top + P_{t|t-1}^{-1})^{-1} \left(HR^{-1}(y_t - A^\top x_t) + P_{t|t-1}^{-1}\hat{\xi}_{t|t-1} \right) \quad (\text{キ})$$

$$P_{t|t}^{-1} = HR^{-1}H^\top + P_{t|t-1}^{-1} \quad (\text{ク})$$



……と思ったのですが、上式の (キ), (ク) は教科書の (13.2.15), (13.2.16) 式とは異なって見えますね。教科書の (13.2.15), (13.2.16) 式の $\hat{\xi}_{t|t}$ と $P_{t|t}^{-1}$ は時刻 t の観測をする前のそれにインクリメントする形式になっていますが、上式はそうありません。どこで計算ミスしてしまったのでしょうか？



計算ミスではないよ。以下の Woodbury の公式を使ってみよう。

Sherman-Morrison-Woodbury の公式

$A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{m \times n}$, $D \in \mathbb{R}^{m \times m}$ について、 A も $D - CA^{-1}B$ も D も

$A - BD^{-1}C$ も正則なとき、以下の等式が成り立つ。

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}$$



何ですかその公式は……右辺も左辺もややこしくてありがたみがよくわからないのですが…しかし、今回は $(HR^{-1}H^{\top} + P_{t|t-1}^{-1})^{-1}$ の箇所にその公式を適用することができそうですね。 $A = P_{t|t-1}^{-1} \in \mathbb{R}^{r \times r}$ としましょう。 $D = -R \in \mathbb{R}^{n \times n}$ としましょう。 $B = H \in \mathbb{R}^{r \times n}$, $C = H^{\top} \in \mathbb{R}^{n \times r}$ とすれば次元の要件は満たされますね。適用すると以下のようになります。

$$(HR^{-1}H^{\top} + P_{t|t-1}^{-1})^{-1} = P_{t|t-1} - P_{t|t-1}H(R + H^{\top}P_{t|t-1}H)^{-1}H^{\top}P_{t|t-1} \quad (\text{ケ})$$

(ケ) を (ク) の両辺の逆行列をとったものに代入すると……教科書の (13.2.16) 式に他なりません！

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}H(R + H^{\top}P_{t|t-1}H)^{-1}H^{\top}P_{t|t-1} \quad (\text{コ})$$

(ケ) を (キ) にも代入してみましょう。

$$\begin{aligned} \hat{\xi}_{t|t} &= \left(P_{t|t-1} - P_{t|t-1}H(R + H^{\top}P_{t|t-1}H)^{-1}H^{\top}P_{t|t-1} \right) \\ &\quad \times \left(HR^{-1}(y_t - A^{\top}x_t) + P_{t|t-1}^{-1}\hat{\xi}_{t|t-1} \right) \\ &= \hat{\xi}_{t|t-1} + P_{t|t-1}HR^{-1}(y_t - A^{\top}x_t) \\ &\quad - P_{t|t-1}H(R + H^{\top}P_{t|t-1}H)^{-1}H^{\top} \left(P_{t|t-1}HR^{-1}(y_t - A^{\top}x_t) + \hat{\xi}_{t|t-1} \right) \\ &= \hat{\xi}_{t|t-1} + \left(1 - P_{t|t-1}H(R + H^{\top}P_{t|t-1}H)^{-1}H^{\top} \right) P_{t|t-1}HR^{-1}(y_t - A^{\top}x_t) \\ &\quad - P_{t|t-1}H(R + H^{\top}P_{t|t-1}H)^{-1}H^{\top}\hat{\xi}_{t|t-1} \\ &= \hat{\xi}_{t|t-1} + \left(P_{t|t-1}H(R + H^{\top}P_{t|t-1}H)^{-1}R \right) R^{-1}(y_t - A^{\top}x_t) \\ &\quad - P_{t|t-1}H(R + H^{\top}P_{t|t-1}H)^{-1}H^{\top}\hat{\xi}_{t|t-1} \\ &= \hat{\xi}_{t|t-1} + P_{t|t-1}H(R + H^{\top}P_{t|t-1}H)^{-1}(y_t - A^{\top}x_t - H^{\top}\hat{\xi}_{t|t-1}) \quad (\text{サ}) \end{aligned}$$

教科書の (13.2.15) 式に一致します！

しかし、(キ), (ク) も計算ミスではなかったのですね。それでも Woodbury の公式を使って (コ), (サ) に変形するのはやはり共分散行列と平均ベクトルが「観測する前のそれにインクリメント」となるのがわかりやすいからなのでしょう。



形のわかりやすさ以上の意味があるよ。(キ), (ク) の場合と (コ), (サ) の場合で逆行列をとるべき行列のサイズはどうなっているだろう？ どちらの場合でも式の中にインバースがあるから、実際に計算していくにあたって逆行列の計算をしないとイケないよね。



逆行列をとるべき行列のサイズですか？ ……(キ), (ク) の場合は $r \times r$, (コ), (サ) の場合は $n \times n$ ですね。両方で異なっていますね……。



だよ。それで、カルマンフィルタを用いる状況下では、観測ベクトルの次元数 n よりも状態ベクトルの次元数 r を大きくとることが多い。逆行列の計算は行列のサイズに対して $\mathcal{O}(N^3)$ だから、逆行列をとるべき行列のサイズは小さいほどよい。だから普通は (コ), (サ) の形式で状態の平均ベクトルと共分散行列の更新式を表示することが多いはずだよ。



なるほど。ときに (コ), (サ) が求まったのはよいのですが、私たちはこれを求めてどうすればいいのでしょうか。



次の時刻 $t+1$ の状態がどうなるか予測してみたら？ (コ), (サ) に基づくとどうなるだろう。



次の時刻 $t+1$ の状態の予測ですか。といっても時刻 t の状態を (13.1.1) 式にしたがって線形変換するだけです。結局多変量正規分布のままです。

$$\begin{aligned} P_{t+1|t} &= F P_{t|t} F^\top + Q \\ &= F P_{t|t-1} F^\top - F P_{t|t-1} H (R + H^\top P_{t|t-1} H)^{-1} H^\top P_{t|t-1} F^\top + Q \quad (\text{シ}) \\ \hat{\xi}_{t+1|t} &= F \hat{\xi}_{t|t} \\ &= F \hat{\xi}_{t|t-1} + F P_{t|t-1} H (R + H^\top P_{t|t-1} H)^{-1} (y_t - A^\top x_t - H^\top \hat{\xi}_{t|t-1}) \quad (\text{ス}) \end{aligned}$$



うん。ここで $K_t \equiv F P_{t|t-1} H (R + H^\top P_{t|t-1} H)^{-1}$ とおくと以下のようにかけるよ。この K_t をカルマンゲインとか利得行列とかよぶ。

$$\hat{\xi}_{t+1|t} = F \hat{\xi}_{t|t-1} + K_t (y_t - A^\top x_t - H^\top \hat{\xi}_{t|t-1}) \quad (\text{ス'})$$



カルマンゲイン？



(ス') においてカルマンゲイン K_t がかかっている $(y_t - A^\top x_t - H^\top \hat{\xi}_{t|t-1})$ の部分って、「時刻 $t-1$ までの観測に基づいて y_t を予測したときの誤差」(予測分布の期待値を予測値としたときの) に他ならないよね。だから、カルマンゲインは「予測誤差を受けて状態の予測をどれだけ修正するか」という係数(行列)になる。カルマンゲイン K_t は例えば観測ノイズの共分散行列 R のノルムが大きいと小さくなる

よね.「観測のノイズが大きいなら予測誤差が出てもあまり修正しない」というのは理解しやすいよね.