

ランダム現象と状態推定
—確率空間からカルマンフィルターまで—

吉田 勝俊

平成 30 年 10 月 1 日

ま え が き

著者の研究室では、人間のゆらぎを測定したり、計算したりしている。当然ながら、確率論の助けが必要になるのだが、善良なる機械系の学生諸君に、確率論の言葉は通じない。ほとんど習う機会がないからだ。例えば、「線形システムにガウス入れたら出力もガウスでしょ」みたいな言葉は通じないのが普通である。かといって、確率論を本格的に勉強してもらう時間もない。それが主題ではないのだから。

そこで、しばらくは、教科書や解説論文に付箋をつけて、まずここ読んで、次ここ読んで...、みたいな抜粋方式で凌いできたのだが、あまり効果的ではなかった。いつまでたっても、学生諸君の誤解が尽きないのである。いっそのこと、まとめて講義してしまったほうが楽のように思えてきて、実際そうするようになり、本書ができた。

このような本書の特徴として、まず、極めて初等的である。かといって、理論に背を向けるでもなく、初等的な題材を選んで理解を促す。例えば、初学者が苦勞する確率論のキーコンセプトに、確率空間 (Ω, \mathcal{F}, P) がある。しかしこれは、題材を選べばさほど難しくない。皆さん、ガチャ (カプセル販売機) はご存知だろうか。本書では、このガチャを題材に確率空間を自作してみせる。これで雰囲気だけでも掴んでおくと、数学チックな論文を読むのが楽になるだろう。もう 1 つの関門は、確率変数 $X(\omega)$ だと思う。これは単なる測定器だと思うと、ごく普通な感覚で扱える (可測性とかも)。得られた測定データをソート (並び換え) する感じで、確率分布関数が導入され、その傾きをとることで確率密度関数が見えてくる。これを物体の密度と見なしたときの重心が、期待値である。以上の議論を多次元化する過程で、独立性や相関性などの概念が身に付いていく。以上が、著者の研究室の確率論速習コースであり、2 章から 6 章をこれに充てる。

つづく 7 章以降は、大学院の輪講用に追加した内容である。カルマンフィルタの構成を学習目標にした。余談だが、同業者の飲み会などで、研究内容を自己紹介すると、ほぼ決って「ああ、カルマンフィルタとかですか？」的な反応が返ってくるのがずっと気になっていた。やっぱ知らないのはモグリかなあと、最近ようやく一念発起して、20 年来、積読状態だった片山先生の名著¹⁾ を読んでみた。著者の実力では、到底、素手では読めず、行間を埋める数学を次々に補充していったのだが、それが 7 章から 10 章までの内容である。特に、確率変数をベクトル、共分散を内積とみなす下りは、直感を交えながら初等的に書けたような気がする。11 章以降で、ようやく、カルマンフィルタの構成に入る。内容的には、片山先生の計

算^{1), 2)}をなぞっただけの学習ノートに過ぎないが、歩幅はかなり狭くできたように思う。その結果、学生諸君にも解読可能なレベルに落ち着いたようだ。

また、本書の連動企画として、本書のグラフ作成などに使用した Python プログラムを、以下の本書レポジトリにあげておく。

- https://github.com/ktysd/test_prob-tut

開発環境は Jupyter Notebook とした。改変自由で好きなようにご活用いただきたい(無保証です)。実は、執筆当初は、Scilab や C++ で数値例を作成していたのだが、最近、Python に乗り換えた。きっかけは新任の若手教員だ。ことあるごとにパイソンパイソンいうので、あるとき思わず「俺もパイソンにしようかな」と口走ってしまった。「しましょーしましょー！」と必死の形相で押し切れ、以来、彼との公用語は Python になっている(笑)。アラフィフの手習いで、しばらくまごついたが、異質の処理を同一言語で書けるのが非常に便利。今流行りの理由を実感している。なんと、確率空間 (Ω, \mathcal{F}, P) は Python で書ける！(事象の加法族なんかも生成できちゃう@o@)

2018 年 8 月

著者しるす

目 次

1. 導入 — 状態推定とノイズの小話

1.1 変位の測定と速度の推定	1
1.2 状態推定と濾波	5

2. 確率空間とは

2.1 ランダム現象と確率空間	7
2.2 見本空間 Ω	7
2.3 事象の代数 \mathcal{F}	8
2.4 確 率 P	10
2.5 確率変数 $X(\omega)$	13
2.6 数 値 例	16
例題の解答例	17

3. ランダム現象の例

3.1 射 撃 実 験	18
3.2 確 率 変 数	19
3.3 事 象 と 確 率	21
3.4 見本空間のタイプ	24
3.5 数 値 例	25

4. 確 率 分 布

4.1 分布関数と密度関数	26
4.2 代表的な確率分布	30

4.3 数 値 例	32
---------------------------	----

5. 多変数の確率分布

5.1 確率ベクトル	33
5.2 2次元の分布	33
5.3 独 立 性	36
5.4 多次元の分布	40
例題の解答例	42

6. 期 待 値

6.1 確率変数の期待値	43
6.2 確率ベクトルの期待値	46
6.3 共分散と相関	47
6.4 条件付き期待値	49
6.5 ベクトルと行列の期待値	51
例題の解答例	52

7. 確率ベクトルの変換

7.1 確率変数の加法とスカラ倍	53
7.2 ガウス確率変数の加法とスカラ倍	54
7.3 共分散行列と多変数ガウス分布	56
7.4 ガウス確率ベクトルの線形変換	57
例題の解答例	61

8. 推定問題と直交射影

8.1 最 小 二 乗 法	62
8.2 確率変数の線形代数	64
8.3 ベクトルの近似	67

8.4	確率変数の推定	69
8.5	数 値 例	72

9. 確率ベクトルの推定

9.1	条件付き期待値	73
9.2	確率ベクトルの推定	74
9.3	確率ベクトルの線形回帰	75
	例題の解答例	79

10. 確 率 過 程

10.1	確 率 過 程	81
10.2	確率的な状態方程式	83
10.3	期待値と共分散行列の推移則	85
10.4	モンテカルロ法による数値例	88

11. カルマンフィルタ

11.1	線形確率システム	91
11.2	状態推定問題	92
11.3	カルマンフィルタの導出	93
11.4	カルマンフィルタ	99
11.5	定常カルマンフィルタ	101
11.6	数 値 例	103

12. 連続時間カルマンフィルタ

12.1	連続時間系の離散時間近似	106
12.2	Kalman–Bucy フィルタ	110
12.3	定常 Kalman–Bucy フィルタ	112
12.4	数 値 例	114

例題の解答例	115
--------	-----

13. LQG 制御への応用

13.1 制御入力を考慮したカルマンフィルタ	116
13.2 LQG レギュレータ	118
13.3 分離定理	119
13.4 数値例	120

14. パラメータ推定への応用と拡張

14.1 カルマンフィルタによるパラメータ推定	122
14.2 拡張カルマンフィルタ (EKF および EKBF)	126
14.3 アンサンブルカルマンフィルタ (EnKF)	130

付 録	136
-----	-----

A.1 集合論の概要	136
------------	-----

引用・参考文献	140
---------	-----

1

導入 — 状態推定とノイズの小話

ある機械振動系の状態推定と制御を例に，本書の動機を述べる．

1.1 変位の測定と速度の推定

図 1.1 のように，1 自由度機械振動系の計測と制御を考える． m は質量， k はばね定数， c

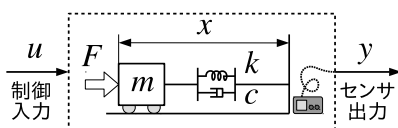


図 1.1 1 自由度機械振動系の計測と制御

は減衰係数とし，質点の実際の変位を x で表す． F は質点を受ける外力である．床との摩擦など， F 以外の外力は無視できるとする．

簡単のため， $m = 1 \text{ kg}$ とし， $u = F$ の場合を考える（制御入力外力そのもの）．このとき，この系の運動方程式は $\ddot{x} = -kx - c\dot{x} + u$ となる． $x_1 := x$ ， $x_2 := \dot{x}$ とおいて一階化すると，次の状態方程式を得る．

$$\dot{x} = Ax + Bu, \quad x := \begin{pmatrix} x \\ \dot{x} \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 1 \\ -k & -c \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (1.1)$$

1.1.1 変位の測定

(1) ノイズがない場合 図 1.1 のセンサを変位センサとする．もし，このセンサには誤差がなく，センサ出力 y と実際の変位 x が等しいなら，出力は，

$$y = Cx = x_1 = x, \quad C = (1, 0) \quad (1.2)$$

と表せる．これを，制御工学では，観測方程式または出力方程式と呼ぶ[†]．

図 1.2a は，このような出力 y の一例である．条件は $k = c = 1$ ， $x(0) = (1, 0)^T$ とし，制御はなし $u = 0$ とした．

[†] センサが速度センサの場合は， $C = (0, 1)$ で表せる．

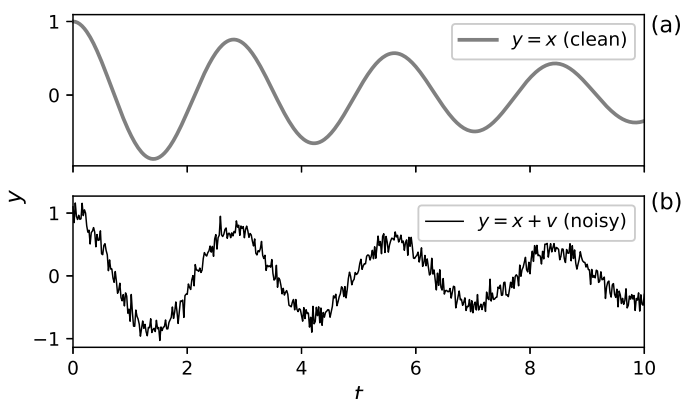


図 1.2 センサ出力 y の例: (a) ノイズなし, (b) ノイズあり

(2) ノイズがある場合 次に, 何らかの原因で測定値にノイズが混入し, 図 1.2b のようなセンサ出力が得られた場合を考える. こうした状況は, 次のように表わせる.

$$y = Cx + v = x + v \quad (1.3)$$

右辺に, ノイズを表す v を追加した. C は (a) と共通である.

1.1.2 数値微分 (ND) による速度の推定

図 1.2 の振動波形から分かるように, この振動系は不足減衰 (減衰振動) の条件にある. その結果, $y = x = 0$ を横切るオーバーシュートが, 複数回生じている. このオーバーシュートを解消するために, 観測値 y に基づくフィードバック制御器を付加することを試みる.

まず, K をフィードバックゲインとして, 観測値をそのままフィードバックすると,

$$u := -Ky = -Kx \quad (1.4)$$

となる. すなわち, 変位 x のフィードバックが実現する. しかし, これではオーバーシュートは解消できない. なぜなら, 解消のためには, 振動系の減衰特性を強める必要がある. それには, 速度 \dot{x} のフィードバックが必要である.

しかし, この制御系のセンサから速度 \dot{x} の情報は得られない. 得られるセンサ出力は変位 $y = x$ のみである. したがって, 速度フィードバックを実現するには, 変位 $y = x$ の情報から, 速度 \dot{x} を推定する必要がある. 一般に, 測れていない状態量を, 測れた状態量から推定することを, 状態推定という.

ここでは, 最も単純に, 数値微分 (ND: numerical differentiation) によって速度を推定してみる. いま, $y(t)$ のサンプル列 y_0, y_1, y_2, \dots が得られたとする. すなわち,

$$y_i := y(t_i), \quad t_i = t_0 + i\Delta t \quad (i = 0, 1, 2, \dots) \quad (1.5)$$

が得られたとする． Δt はサンプリング間隔である．このとき，時刻 $t = t_i$ の速度 \dot{y}_i の推定値を次のように定める．

$$\dot{y}_i \approx \frac{\Delta y}{\Delta t}(t_i) := \frac{y(t_i) - y(t_{i-1})}{\Delta t} \quad (t = 1, 2, \dots) \quad (1.6)$$

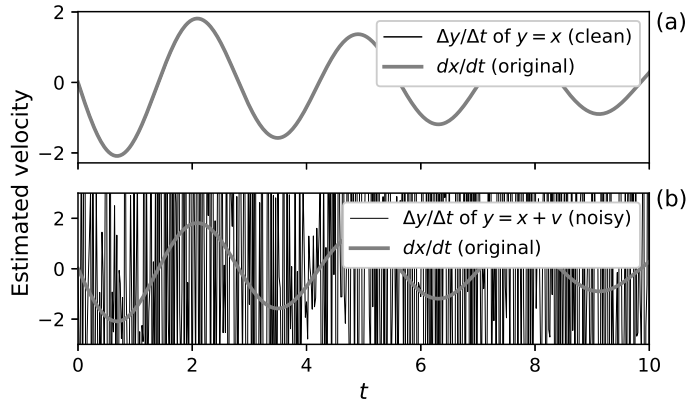


図 1.3 数値微分による速度推定: (a) ノイズなし, (b) ノイズあり．黒が真値，水色と赤が数値微分推定値

(1) ノイズがない場合 ノイズがない図 1.1a の y について，(1.6) で推定した速度の推定値を，図 1.3a に水色示す． $\Delta t = 0.25$ s とした．水色の推定値は，黒の真値とよく一致している．

このように，測定値がノイズを含まない場合は，数値微分 (1.6) の方法は有効である．

(2) ノイズがある場合 同様にして，ノイズのある図 1.1b の y から推定した速度を図 1.3b に赤で示す．図 1.1b の測定値 y (赤) の擾乱が，図 1.3b の速度推定値 $\Delta y/\Delta t$ (赤) では，大きく増幅されている．

これは，数値微分 (1.6) が，一般にハイパスフィルタの性質を持つためである．その結果，図 1.3b の速度推定値からは，真値の傾向を知ることすら不可能となっている．

▶▶ 信号の低周波成分を除去し，高周波成分を通過させるものをハイパスフィルタという．その逆をローパスフィルタという．

このように，測定値がノイズを含む場合 (特に周波数が高いもの)，数値微分 (1.6) による速度推定は，容易に破綻する．

演習 1.1 数値微分すると高周波ノイズが増幅されるメカニズムを，振動波形とその時間微分波形をスケッチしながら考察せよ．

1.1.3 推定値による制御

数値微分による速度の推定値 $\Delta y/\Delta t$ を用いて，近似的な微分制御を行う．すなわち，制御入力を，

$$u = -K\dot{y} \approx -K \frac{\Delta y}{\Delta t} \quad (1.7)$$

とする．このときのブロック線図を図 1.4 に示し，制御結果を図 1.5 に示す．

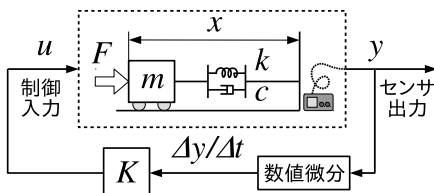


図 1.4 数値微分による微分制御

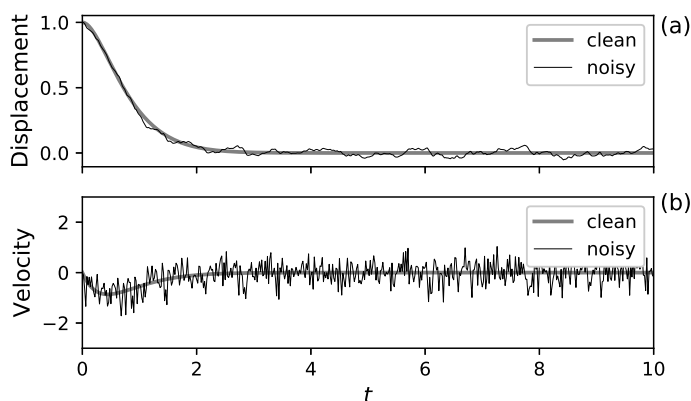


図 1.5 数値微分による微分制御の結果: 水色がノイズなし，赤がノイズあり ($K = 0.45$)

図 1.5 上段が質点の変位 x ，同じく下段が速度 \dot{x} の推定値である．水色がノイズが無いときの結果，赤がノイズがあるときの結果である．ノイズがある赤の結果に着目すると，特に，速度 \dot{x} の正負が激しく切り替わっている．これは，ノイズの影響で，制御中の質点が小刻みに激しく振動していることを表している．

1.2 状態推定と濾波

このようなノイズが引き起こす不具合を解消するために、濾波 (フィルタリング) という信号処理がよく用いられる。その目的は、図 1.2b のような測定結果の擾乱成分を除去して、図 1.2a のような真値を推定することである。

まだ学習前なので説明は省くが、11 章以降で導入するカルマンフィルタという推定技術を用いると、驚くほどよい結果が得られる。

図 1.6 は、このカルマンフィルタによって、ノイズに乱された y から速度 \dot{x} を推定し、この推定された速度をフィードバックしたときの制御結果である。

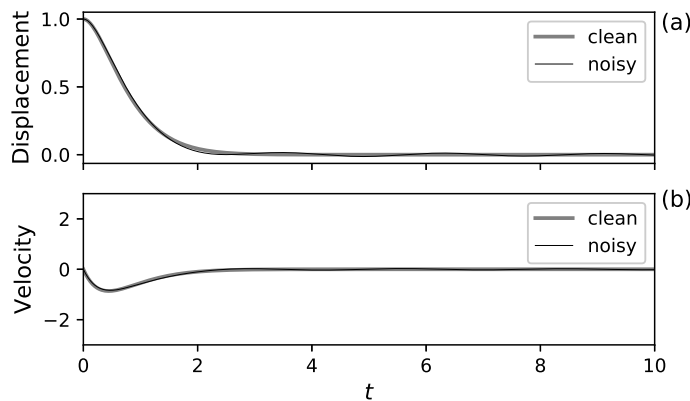


図 1.6 Kalman-Bucy フィルタによる制御結果: 水色がノイズなし, 赤がノイズあり

ノイズが無いとき (水色) と、有るとき (赤) の波形が、ほとんど重なっており、測定値 y を乱していたノイズの影響が、ほとんど除去されてしまったことが分かる[†]。

それでは、さきほどとの違いは何かというと、図 1.5 は四則演算ベースの推定であり、図 1.6 は現代確率論ベースの推定なのである。ドローンやロボットの姿勢制御とか自己位置推定には後者の技術が投入されている。意外なところでは、天気予報などもそうである。

図 1.6 のような高度な推定技術は、使うこと自体は簡単だ。多くの解析ソフトに搭載されているからだ。しかし、現代確率論ベースなだけに、高校～大学 1 年の確率・統計レベルの知識だけでは歯が立たない。ヘルプに出てくる数式すらほとんど理解できないのではないかな。これは無理もない。微積分なしに大学の力学書は読めない、それと同じことである。

[†] 正直にいうと、これほどよい結果になるのは、モデル化誤差がゼロの条件で解いた架空の例題だからである。実験ではモデル化誤差が避けられないので、相応の誤差が出る。

そこで，本書では，工学部 1～2 年生程度の知識を前提に，現代確率論に入門し，推定理論の基礎を学び，カルマンフィルタに到達することを試みる．多くの読者にとっては，ほとんど新規の学習内容になると思うが，初めて微積分に接したときのことを思い出してほしい．最初は面くらうかもしれないが，いつの間にか普通に使えるようになる．

2 確率空間とは

ミニ玩具のカプセル販売機 (以下, ガチャ) を題材に, 現代確率論の難所を早々にクリアする. 確率空間 (Ω, \mathcal{F}, P) と確率変数 $X(\omega)$ を自作するイメージが持てれば目標達成である.

2.1 ランダム現象と確率空間

ガチャを楽しむ顧客にとって, 次に何が出てくるかは予測不可能だ. このような, 試行ごとに結果が変化してしまい, 結果を予測困難な現象を, ランダム現象という. 現代確率論では, 様々なランダム現象を, 確率空間 (Ω, \mathcal{F}, P) という共通の器に収めて考察する[†].

以下, 確率空間 (Ω, \mathcal{F}, P) の簡単な例として, ガチャを表す Ω, \mathcal{F}, P を試作してみる.

2.2 見本空間 Ω

2.2.1 試行

実験 1 回分の実施を, 試行という. 例えば, ガチャからカプセルを 1 つ取り出す操作は試行である. いま, ガチャからカプセルを N 回取り出したところ, 次のような結果が得られたとしよう.

$$\text{犬, 鳥, 犬, 人, 人, 鳥, 犬, } \dots, \text{人} \quad (2.1)$$

このような複数回の試行の並びを試行列という. 試行の回数 N を試行数という.

2.2.2 見本

こうした試行列を紙の上に表記する場合, 字画が多いと面倒なので, 適当な記号や数字が割り当てられる. 例えば, 次のようにしてみよう.

$$a: \text{犬}, \quad b: \text{鳥}, \quad c: \text{人}$$

[†] パボリス³⁾ p15, 津野⁴⁾ p3 など.

結果を識別する記号 a, b, c を見本という．見本を使うと，試行列は簡潔に，

$$a, b, a, c, c, b, a, \dots, c \quad (2.2)$$

と書ける．

2.2.3 見本空間

次に，確率論では，全ての見本を収録した見本帳 Ω を作る．例えば，ミニ玩具が a, b, c の 3 種類だけのとき，ガチャの見本帳を次のように表す．

$$\Omega := \{a, b, c\}$$

このような全ての見本を列挙した集合 (見本帳) Ω を見本空間という．

▶▶ 数学ではしばしば「集合」を「空間」と言い表す．ただし「空間」というときには，何かが動く土台というニュアンスが込められることが多い．

2.3 事象の代数 \mathcal{F}

集合論に不慣れな読者は，A.1 節を予習してから，ここに戻ってほしい．

2.3.1 事象

続いて，確率論では，見本の様々なクラス分けを設定する．どのように設定するかは用途による．例えば，哺乳類と 2 足歩行に興味があるなら，次のように設定するとよいだろう．

$$\begin{aligned} \text{哺乳類: } A_1 &= \{\text{犬, 人}\} = \{a, c\}, \\ \text{2 足歩行: } A_2 &= \{\text{鳥, 人}\} = \{b, c\} \end{aligned} \quad (2.3)$$

得られたクラスのそれぞれを事象という．大学の選択科目のように，各人が複数のクラスに属してよいとする．また，事象どうしの集合演算の結果 (表 2.1) もまた，事象であると約束する．いかなる事象も， Ω の部分集合をなす．

事象 A と共通部分を持たない事象を， A の背反事象という．例えば，余事象 \overline{A} は必ず A の背反事象となる．

演習 2.1 (2.3) の A_1, A_2 を，見本空間 $\Omega = \{a, b, c\}$ 上の事象とする．和事象 $A_1 \cup A_2$ ，積事象 $A_1 \cap A_2$ ，差事象 $A_1 \setminus A_2$ ，余事象 $\overline{A_1}$ を求めよ．

▶▶ 余事象は，見本空間 (= 全体集合) を決めないと定義できない．

表 2.1 事象の演算 (集合演算)

集合演算	選ばれる見本	名称
集合和 $A_1 \cup A_2$	いずれかに含まれる	和事象
集合積 $A_1 \cap A_2$	両方に含まれる	積事象
集合差 $A_1 \setminus A_2$	$:= A_1 \cap \bar{A}_2$	差事象
補集合 \bar{A}_1	A_1 に含まれない ($:= \Omega \setminus A_1$)	余事象

2.3.2 事象の代数

一般に，集合の集合を，族という．したがって，事象を集めた集合は，事象の族と呼ばれる．例えば，冒頭で着目した事象 A_1, A_2 を要素とする集合，

$$F = \{ \text{哺乳類, 2足歩行} \} = \{A_1, A_2\} = \{\{a, c\}, \{b, c\}\} \quad (2.4)$$

は事象の族である．

ここで，興味ある事象だけを並べた族 (2.4) には，各種集合演算の結果が含まれていない．例えば，事象の積 $A_1 \cap A_2 = \{c\}$ もまた事象とみなす約束だが，(2.4) の F には含まれない．そこで，明らかに足りない事象として，

- 和事象 $A_1 \cup A_2 = \{a, b, c\} = \Omega$
- 積事象 $A_1 \cap A_2 = \{c\}$
- 余事象 $\bar{A}_1 = \{b\}$, $\bar{A}_2 = \{a\}$

を追加してみる．その結果，事象の族は，

$$F' = \{\{a, c\}, \{b, c\}, \{a\}, \{b\}, \{c\}, \Omega\} \quad (2.5)$$

まで膨らむ．

▶▶ (1人ぼっちのクラス) $\{a\}, \{b\}, \dots$ は1人だけのクラス(事象)を表す．中括弧なしの a, b, \dots は個人(見本)を表す．両者の意味は異なる．

しかし，この F' においても，2段階の集合演算では閉じない．例えば，

$$\overline{A_1 \cup A_2} = \overline{\{a, b, c\}} = \{\} := \emptyset \quad (\text{空事象} \quad \text{くうじしょう}) \quad (2.6)$$

は F' に含まれない．こうした多段階の集合演算をも網羅しようとなると，事象の族は，

$$\mathcal{F} = \{\{a, c\}, \{b, c\}, \{a, b\}, \{a\}, \{b\}, \{c\}, \Omega, \emptyset\} \quad (2.7)$$

まで膨らむ．この \mathcal{F} は， A_1, A_2 に対する任意回の集合演算に対して閉じている． \mathcal{F} は， A_1, A_2 から理論上想定しうる全ての事象のリストを与える．

2.3.3 σ 代数

(2.7) の \mathcal{F} は，任意回の集合演算で閉じるように F を膨らませたものである．これを， F から生成した σ 加法族または σ 代数と呼ぶ．

ある集合族 \mathcal{F} が σ 代数となるためのチェックリストが整備されている .

定義 2.1 集合 Ω の部分集合を集めた族 \mathcal{F} を考える . 次の条件を満足する \mathcal{F} を , Ω 上の σ 代数という .

$$(\sigma 1) \quad \Omega \in \mathcal{F}.$$

$$(\sigma 2) \quad A \in \mathcal{F} \implies \bar{A} \in \mathcal{F}.$$

$$(\sigma 3) \quad A_1, A_2, \dots, A_n \in \mathcal{F} \implies A_1 \cup A_2 \cup \dots \cup A_n \in \mathcal{F}.$$

(n は任意の自然数)

このような \mathcal{F} は , 大きくも小さくもとれるが , そのなかで最小なものを $\sigma(F)$ と書く .

▶▶ 最小の集合とは , 要素数が最小の集合ということ .

例題 2.1 $\Omega := \{a, b, c\}$ 上の族 $G := \{\{a, c\}, \{b\}\}$ から生成した最小の σ 代数 $\sigma(G)$ を求めよ . $\{\{a, c\}, \{b\}\} = \{\text{哺乳類}, \text{鳥類}\}$

演習 2.2 $\Omega := \{a, b, c\}$ 上の族 $H := \{\{a, b\}, \{b\}\}$ から生成した最小の σ 代数 $\sigma(H)$ を求めよ .

最後に , 最大の σ 代数についても述べておこう . (2.4) の F から生成される最大の σ 代数は , Ω のべき集合というものになる . Ω のべき集合とは , Ω のすべての部分集合からなる族のことで , ガチャの $\Omega = \{a, b, c\}$ については , 次のようになる .

$$\mathcal{F}_\Omega = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \Omega, \emptyset\} \quad (2.8)$$

2.4 確 率 P

事象 A が生起する確率を $P(A)$ と書く . その値の設定方法には , 大きく 2 種類ある . 見本に基づく方法と , 標本に基づく方法である .

2.4.1 見本に基づく方法 (論理的解釈)

大雑把にいうと , 実際の試行によらずに確率を設定する方法である . 以下 , 集合 A の要素数を $\#A$ と表記する .

古典的には , ラプラス (1749 ~ 1827) の方法がある . ガチャの例では ,

$$\begin{aligned} P(\{a\}) &:= \frac{\#\{a\}}{\#\Omega} = \frac{1}{3}, & P(\{b\}) &:= \frac{\#\{b\}}{\#\Omega} = \frac{1}{3}, \\ P(\{c\}) &:= \frac{\#\{c\}}{\#\Omega} = \frac{1}{3} \end{aligned} \quad (2.9)$$

とできる． $\#\Omega = 3$ は見本の総数であり， $\#\{a\} = 1$ は事象 $\{a\}$ に含まれる見本の数である．この定め方は，

- 結果の発現傾向について知識がない場合，等確率を仮定するのが最も合理的である．という仮説に基いている．しかし，ラプラスの方法は，歴史的に様々な批判にさらされてきたようである（実験と合わないなど）．

そこで，力学を始めとする確率論以外の知見を動員して，例えば，

$$P(\{a\}) := \frac{1}{4}, \quad P(\{b\}) := \frac{2}{4}, \quad P(\{c\}) := \frac{1}{4} \quad (2.10)$$

などと確率を仮置きし，これを作業仮説とみなす立場もある．

以上の方法論は，実測に基いていないため，妥当性に疑問が残る．しかし，ランダム現象を数値シミュレーションするような場合は，ランダム性の源をこの立場で設けるしかない．その意味で，淘汰された古い方法論というわけではない．

2.4.2 その他の事象の確率

(2.9) では，ガチャの事象 $\{a\}, \{b\}, \{c\}$ について確率を定めたが，(2.7) の \mathcal{F} にエントリーされた事象はまだ 5 つある．これらの確率を矛盾なく割り出すためのルールが整備されている．

算法 2.1 (確率の性質) 確率 $P(\cdot)$ は一般に次の性質を持つ．

(P1) 任意の事象 A に対して， $0 \leq P(A) \leq 1$.

(P2) 特に， $P(\Omega) = 1$.

(P3) 背反事象 A_1, A_2, \dots に対して，

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots .$$

以上の帰結として，次の性質も示される．

(P4) A の余事象 \bar{A} に対して， $P(\bar{A}) = 1 - P(A)$

(P5) $P(\emptyset) = 0$. ($\emptyset := \{\}$ は空集合)

例題 2.2 (P2)(P3) を前提として (P4) を示せ．(ヒント： $\Omega = A \cup \bar{A}$)

演習 2.3 (P4) を前提として (P5) を示せ．

(2.7) のガチャの事象 \mathcal{F} に話を戻す．(P2)(P5) より $P(\Omega) = P(\{a, b, c\}) = 1$ と $P(\emptyset) = 0$ が定まる．さらに，事象 $\{a\}$ と $\{b\}$ は背反だから，

$$P(\{a, b\}) = P(\{a\} \cup \{b\}) = P(\{a\}) + P(\{b\}) = \frac{2}{3} \quad (2.11)$$

を得る．

演習 2.4 (2.9) を前提に, (2.7) の \mathcal{F} の残りの事象 $\{a, c\}$ と $\{b, c\}$ の確率を求めよ.

算法 2.1 による確率計算は, 特に (2.10) のような, 要素数のカウントで確率を割り出せないような場合に役立つ.

演習 2.5 等確率でない (2.10) を前提に, (2.7) の \mathcal{F} の残りの事象 $\{a, c\}$, $\{b, c\}$, $\{a, b\}$, $\Omega = \{a, b, c\}$, $\emptyset = \{\}$ の確率を求めよ.

2.4.3 試行に基づく方法 (頻度解釈, 統計的解釈)

あるランダム現象の無限試行列 $a_1, a_2, a_3, \dots \in \Omega$ を要素とする集合:

$$\Pi := \{a_1, a_2, \dots, a_n, \dots\} \quad (2.12)$$

を統計学では, 母集団という. 一般に $\Pi \neq \Omega$ である. 例えば, ガチャでは,

$$\Omega = \{a, b, c\} \quad (\text{見本空間})$$

$$\Pi = \{a, b, a, c, c, b, a, \dots\} \quad (\text{母集団})$$

のような違いとなる. また, 母集団の一部を抽出した部分集合,

$$\Pi_n = \{a_1, a_2, \dots, a_n\} \subset \Pi \quad (2.13)$$

を, 標本という. このなかで, 試行結果 a_i が事象 $A \subset \Omega$ に含まれる試行の個数を $N(A)$ と書くとき,

$$P_n(A) := \frac{N(A)}{\#\Pi_n} = \frac{N(A)}{n} \quad (2.14)$$

を, 事象 A の標本 Π_n における相対度数という.

例題 2.3 ガチャの標本 $\Pi_7 := \{a, b, a, c, c, b, a\}$ について, 事象 $A_1 = \{a, c\}$, $A_2 = \{b, c\}$ の相対度数をそれぞれ求めよ.

演習 2.6 (2.7) の \mathcal{F} の残りの事象について, 相対度数を求めよ.

以上の統計学をベースに, 確率を, 相対度数の極限,

$$P(A) := \lim_{n \rightarrow \infty} P_n(A) \quad (2.15)$$

として定める. このような確率の解釈を, 頻度解釈または統計的解釈という.

▶▶ (混同しやすい用語) 見本空間 Ω を, 標本空間と呼ぶ流儀がある. 語感としては, 標本が降り積る土台が標本空間 Ω である.

2.5 確率変数 $X(\omega)$

以上，ガチャのようなランダム現象は確率空間 (Ω, \mathcal{F}, P) で表せる． Ω は試行結果の見本帳であり，そのクラス分けを事象と呼んだ． \mathcal{F} は集合演算で閉じた事象の族， $P(A)$ は事象 $A \in \mathcal{F}$ の確率であった．しかしながら，この確率空間 (Ω, \mathcal{F}, P) には，工学的な実験には付き物の，測定値を収める器がまだ存在しない．

例えば，ガチャの各見本について，全長 x [cm] と重量 y [g] を測定したところ，下表の結果が得られたとしよう．

見本 ω	全長 x [cm]	重量 y [g]
a : 犬	3.5	22
b : 鳥	2.8	22
c : 人	3.5	25

このとき，例えば $x = 3.5$ となる確率はどのように考えればよいのだろうか？

▶▶ 値に重複が見られる例としたが，こうしても確率論は破綻しない．ただし，「可測性」と呼ばれる実用上の問題が起こる (2.5.4 節で述べる) ．

2.5.1 確率変数

そこで，まず，上記の「測定」を次のような関数でモデル化する．

$$\begin{cases} x = & X(a) = 3.5, & X(b) = 2.8, & X(c) = 3.5 \\ y = & Y(a) = 22, & Y(b) = 22, & Y(c) = 25 \end{cases} \quad (2.16)$$

このような，各見本 ω の測定値 x, y を記憶させた関数 X, Y を確率変数と呼ぶ．

▶▶ $x = \sin(t)$ という関数関係において， \sin は関数， x は t における関数値である．同様にして， $x = X(\omega)$ において， X は関数， x は ω における関数値である．

▶▶ 見本の関数を確率変数と呼ぶのは，確率論の歴史的慣習である．

参考まで，上記の確率変数 $x = X(\omega)$ は次のようにプログラミングできる．

```
def X( ):          #Python の場合
    if == a:
        val = 3.5
    if == b:
        val = 2.8
    if == c:
        val = 3.5
    return val
```

以上の測定を通して，(2.2) の試行列 $a, b, a, c, c, b, a, \dots, c$ からは，

$$\begin{cases} x &= X(a), X(b), X(a), X(c), X(c), X(b), X(a), \dots, X(c) \\ &= 3.5, 2.8, 3.5, 3.5, 3.5, 2.8, 3.5, \dots, 3.5 \\ y &= Y(a), Y(b), Y(a), Y(c), Y(c), Y(b), Y(a), \dots, Y(c) \\ &= 22, 22, 22, 25, 25, 22, 22, \dots, 25 \end{cases} \quad (2.17)$$

のようなデータ列が得られる。

2.5.2 測定値の生起確率

早速, $P(x = 3.5)$ のような確率を導入したいわけだが, 確率 $P(A)$ は事象に対する量である。そこで, 測定値 $x = X(\omega)$ に対応する事象を, $X^{-1}(x)$ と書き, 次のように逆算する。

$$X^{-1}(x) := \{\omega \in \Omega \mid X(\omega) = x\} \quad (2.18)$$

例えば, (2.16) の X について $x = 3.5$ とすると,

$$X^{-1}(3.5) := \{\omega \in \Omega \mid X(\omega) = 3.5\} = \{a, c\} \quad (2.19)$$

となる。もう 1 つの測定値 $x = 2.8$ については,

$$X^{-1}(2.8) := \{\omega \in \Omega \mid X(\omega) = 2.8\} = \{b\} \quad (2.20)$$

となる。

演習 2.7 同様に, $Y^{-1}(22)$ と $Y^{-1}(25)$ を求めよ。

以上より, $x = 3.5$ となる確率は, ラプラスの確率 (2.9) では,

$$P(X^{-1}(3.5)) = P(\{a, c\}) = \frac{2}{3} \quad (2.21)$$

となる。表記短縮のため, よく用いられる略記

$$P(X = x) := P(X^{-1}(x)) \quad (2.22)$$

を使うと, 先の例は,

$$P(X = 3.5) = P(\{a, c\}) = \frac{2}{3} \quad (2.23)$$

と書ける。

演習 2.8 確率 $P(X = 2.8)$, $P(Y = 22)$, $P(Y = 25)$ を求めよ。

2.5.3 確率変数から生成される σ 代数

$X(\omega)$ が取り得る値 $x = 3.5, 2.8$ が表す事象は $\{a, c\}, \{b\}$ であった．これらから生成される σ 代数を, $\sigma(X)$ と書き, 確率変数 X から生成される σ 代数という．

実際に計算してみよう．定義は,

$$\sigma(X) := \sigma(\{\{a, c\}, \{b\}\}) \quad (2.24)$$

であった．ひとまず, 括弧内の集合族に, $\Omega = \{a, b, c\}, \emptyset = \{\}$ を追加してみると,

$$F = \{\{a, c\}, \{b\}, \{a, b, c\}, \{\}\} \quad (2.25)$$

となる．これは, 各事象の余事象を含む．また, 全てのペアの和集合を含む．というわけで, これでよい．すなわち, $\sigma(X) = F = \{\{a, c\}, \{b\}, \Omega, \emptyset\}$ である．

演習 2.9 同様に $\sigma(Y)$ を求めよ．

2.5.4 確率変数の可測性

X から生成した $\sigma(X)$ の事象は全て, 測定値 $x = X(\omega)$ に関する条件で表せる．これを表 2.2 に示す． $\sigma(X)$ が含む事象を○で示し, 対応する測定値 x の条件を示した．同じく $\sigma(Y)$ についても示した．

表 2.2 確率変数から生成される事象と測定値の関係

事象	$\sigma(X)$	測定値 x の条件	$\sigma(Y)$	測定値 y の条件
$\{\}$	○	$x \neq 2.8, 3.5$	○	$y \neq 22, 25$
$\{a\}$				
$\{b\}$	○	$x = 2.8$		
$\{c\}$			○	$y = 25$
$\{a, b\}$			○	$y = 22$
$\{b, c\}$				
$\{a, c\}$	○	$x = 3.5$		
$\{a, b, c\}$	○	$x = 3.5 \text{ or } 2.8$	○	$y = 22 \text{ or } 25$

一般に, ある事象を, 測定値側の条件で識別できることを, 可測であるという．例えば, 表 2.2 の X は, その値 x に関する条件で,

$$\mathcal{F}_1 = \{\{a, c\}, \{b\}, \Omega, \emptyset\} \quad (= \sigma(X)) \quad (2.26)$$

の全ての事象を識別できる．この事実を, X は \mathcal{F}_1 -可測である, と言い表す．(これは, \mathcal{F}_1 を X から生成したのだから当然である)

可測でない状況は次のようにして起る．例えば, 2 足歩行 $A_2 = \{b, c\}$ に興味があったとしよう．まず, 対応する σ 代数を生成し, 見るべき事象をリストアップする．

$$\mathcal{F}_2 = \sigma(A_2) = \{\{b, c\}, \{a\}, \Omega, \emptyset\} \quad (2.27)$$

次に，測定値として見本の全長 x を採用したところ， x のデータ列を分析しても 2 足歩行 A_2 か否かは判別できないことが分かった．なぜなら，表 2.2 の $\sigma(X)$ においては， $A_2 = \{b, c\}$ の部分が空欄なので， A_2 を表す X の値は存在しない．そこで，測定値を重量 y に変更しても，表 2.2 の $\sigma(Y)$ の $A_2 = \{b, c\}$ は空欄のままなので，やはり A_2 は識別できない．ようするに，全長や重量を測定しても「2 足歩行」は識別できない．

一般に，確率変数 X の値 (測定データの値) によって， \mathcal{F} の事象が全て識別できることを， X は \mathcal{F} -可測である，と言い表す．

上述の例では， X も Y も \mathcal{F}_2 -可測でない．これを解消するには， A_2 が識別可能となるような新たな確率変数 (測定方法) を導入しないといけない．

演習 2.10 \mathcal{F}_2 -可測となるような新しい確率変数 $Z(\omega)$ の例を作れ．

2.6 数 値 例

- Ch02.ipynb

例題の解答例

例題 2.1 p10

初期状態を $\mathcal{F}_0 = F = \{\{a, c\}, \{b\}\}$ とする．これに，定義 2.1 p10 のチェックリストが求める事象を追加していく．まず， $(\sigma 1)$ より， Ω が必要なので追加する．

$$\mathcal{F}_1 = \mathcal{F}_0 \cup \{\Omega\} = \{\{a, c\}, \{b\}, \{a, b, c\}\}$$

次に， $(\sigma 2)$ を充足するため， \mathcal{F}_1 の余事象を求めると，

$$\overline{\{a, c\}} = \{b\}, \quad \overline{\{b\}} = \{a, c\}, \quad \overline{\{a, b, c\}} = \{\} = \emptyset$$

となる．このうちで \mathcal{F}_1 にないのは \emptyset であるから，これを追加すると，次のようになる．

$$\mathcal{F}_2 = \mathcal{F}_1 \cup \{\emptyset\} = \{\{b\}, \{a, c\}, \Omega, \emptyset\}$$

定義 2.1 p10 でチェックすると， \mathcal{F}_2 は Ω を含むから $(\sigma 1)$ を満す．全ての余事象を含むから $(\sigma 2)$ を満す．全ての組み合わせの和事象を含むから $(\sigma 3)$ も満す．さらに， \mathcal{F}_2 の要素がどれか 1 つでも欠けると，チェックを通らなくなるから， \mathcal{F}_2 は最小である．したがって， $\sigma(F) = \mathcal{F}_2$ である．

例題 2.2 p11

(P2) より $P(\Omega) = 1$ が成立する．また，余事象の定義より $\Omega = A \cup \overline{A}$ が成立し， A と \overline{A} は背反だから，(P3) より，

$$1 = P(\Omega) = P(A \cup \overline{A}) \stackrel{(P3)}{=} P(A) + P(\overline{A})$$

したがって， $P(\overline{A}) = 1 - P(A)$ ．

例題 2.3 p12

事象 $A_1 = \{a, c\}$ に該当する試行の個数は，試行結果が a または c となった回数だから $N(A_1) = 5$ である．同じく $A_2 = \{b, c\}$ については， b または c となった回数だから $N(A_2) = 4$ である．ゆえに，

$$P_n(A_1) = \frac{N(A_1)}{7} = \frac{5}{7}, \quad P_n(A_2) = \frac{N(A_2)}{7} = \frac{4}{7}$$

となる．

3

ランダム現象の例

前章のガチャの確率空間 (Ω, \mathcal{F}, P) は見本数が 3 だったので、実際に見本を見ながら事象や確率を議論できた。いわば、測定値 (確率変数) は補助的な存在であった。しかし、工学的には、測定値 (確率変数の値) のほうが主役で、見本空間 Ω はブラックボックスという状況も少なくない。本章では、測定値を出発点とする定式化を概観する。

3.1 射撃実験

図 3.1 は、原点から発射された弾丸の軌跡をシミュレートしたものである。弾丸は $x = 100$ で標的 (壁) に着弾し、図の右端の小円は着弾点を表す。この軌跡は、次の運動方程式を数值的に解いて求めた。

$$m\ddot{x} = -c|\dot{x}|\dot{x}, \quad m\ddot{y} = -c|\dot{y}|\dot{y} - mg \quad (3.1)$$

x は発射点からの距離で、 y は高さである。各右辺の第 1 項は、速度の 2 乗に比例する抗力を表す。 m は質量で、 g は重力加速度である。初期条件は、原点を初期位置として、初速度 800m/s を仰角 10^{-3}rad で与えた。

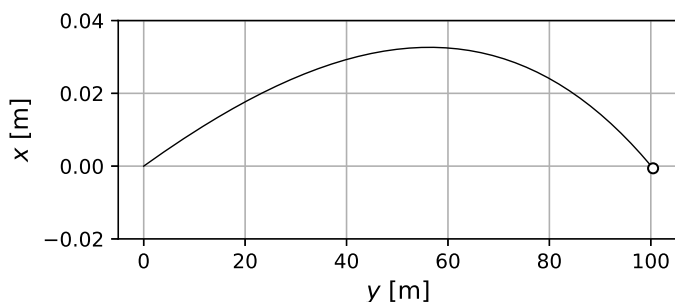


図 3.1 弾丸の軌跡

この理想的なシミュレーションの世界では、パラメータ m, c, g と初期条件を変えない限り、何度発射しても、この弾道が再現される。このような再現性のある挙動は、決定論的であるという[†]。

ところが、同様の試行を物理的に行うと、例えば、図 3.2 のような結果になるはずだ。試行ごとに結果が変化している。これはすなわち、ランダム現象である。

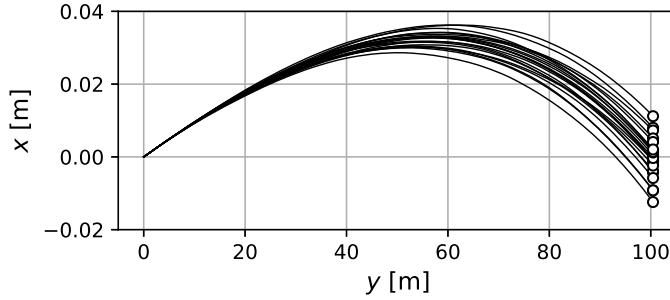


図 3.2 弾丸の軌跡 (試行 21 回)

3.2 確 率 変 数

図 3.2 のランダムな着弾高さ y の標本は、

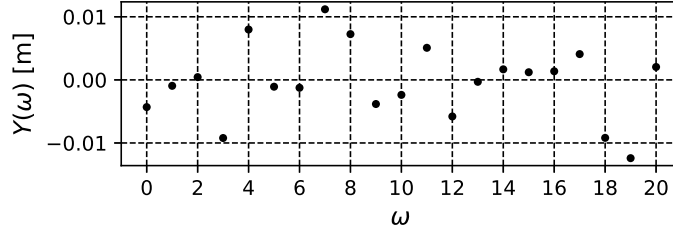
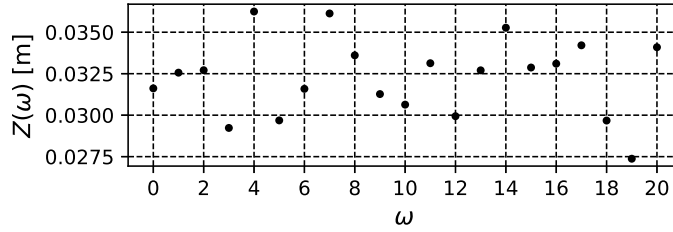
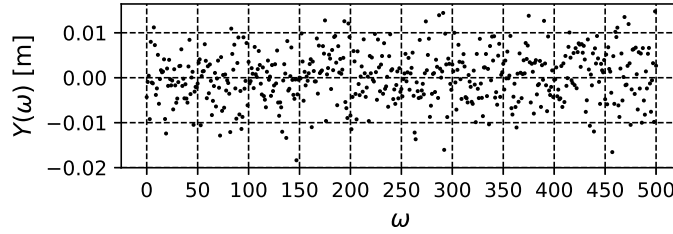
$$\begin{aligned} \Pi_n &= \{y_0, y_1, \dots, y_n\} \\ &= \left\{ \begin{array}{lll} -0.0043143, & -0.0009421, & 0.0004355, \\ -0.0092159, & 0.0079840, & -0.0010818, \\ -0.0012494, & 0.0111891, & 0.0072529, \\ -0.0038258, & -0.0023774, & 0.0050846, \\ -0.0057984, & -0.0003055, & 0.0016732, \\ 0.0012103, & 0.0013663, & 0.0040884, \\ -0.0091895, & -0.0124091, & 0.0020454 \end{array} \right\} \end{aligned} \quad (3.2)$$

となった ($n = 21$)。全て値が異なるので、見本帳 Ω には少なくとも 21 個の見本が必要である。そこで、 y_i の添字をそのまま見本空間

$$\Omega := \{0, 1, 2, \dots, 20\} \quad (3.3)$$

とし、確率変数を、

[†] 確定的ともいう。

図 3.3 確率変数 $Y(\omega)$ ($x = 100$)図 3.4 確率変数 $Z(\omega)$ ($x = 60$)図 3.5 確率変数 ($x = 100$, $\#\Omega = 501$)

$$y = Y(\omega) = y_{\omega} \quad (\omega = 0, 1, \dots, 20) \quad (3.4)$$

としてみる．これを図 3.3 に示す．横軸は見本番号 $\omega \in \Omega$ ，縦軸は対応する着弾点の観測値 $y = Y(\omega)$ である．

図 3.3 と同じ試行について， $x = 60$ での通過高さを測定値とした確率変数 $Z(\omega)$ を図 3.4 に示す． $Y(\omega)$ と $Z(\omega)$ は同一の見本空間 (= 実験) 上で定めた確率変数だが，測定方法が違うので関数形は異なる．

同様にして，標本数を 501 としたときの結果を図 3.5 に示す．この場合も標本 $\Pi_{501} = \{y_0, \dots, y_{501}\}$ に重複は生じなかったので， y_i の添字を $\omega := i$ としている．

見本空間 Ω とそれ上の確率変数 $Y(\omega)$ ， $Z(\omega)$ の具体的な役割については，次のようなイメージを持っておくとよい．

- Ω は試行結果の全種類を収録した見本帳．
- $\omega \in \Omega$ は見本番号．(番号ではなく文字列でもよい)
- 同じ実験 Ω の同じ見本 $\omega \in \Omega$ から，性質の異なる測定値 $X(\omega), Y(\omega), Z(\omega), \dots$ が収集される．これらの違いは測定箇所や使用センサの違いなどによる．

このように，確率変数の関数形はギザギザしていて捉え所がないが，測定値が実際そうなのだからしかたがない．

3.3 事 象 と 確 率

前章のガチャと本章の射撃実験との違いは，見本空間 Ω の大きさにある．ガチャでは $\#\Omega = 3 < \infty$ だったが，以下に述べるように，射撃実験では理論上 $\#\Omega = \infty$ となる．

3.3.1 初等的確率の限界

2.5.2 節 p14 の方法をそのまま適用してみる．例えば，着弾点が $y = Y(0) = -0.0043143$ となる確率 $P(Y = -0.0043143)$ は，

$$P(Y = -0.0043143) := P(Y^{-1}(-0.0043143)) = P(\{0\}) \quad (3.5)$$

となる．その値は，ラプラスの方法 (2.9) により， $\#\Omega = 21$ の場合，

$$P(\{0\}) := \frac{\#\{0\}}{\#\Omega} = \frac{1}{21} \quad (3.6)$$

となる．と結論づけてしまってよいのだろうか？

残念ながらそうはできない．ガチャでそうできたのは，分母の 3 が確定したからである．これに対して，着弾点の可能性は高々 21 点では網羅できない．実際，501 点まで増やした標本 Π_{501} にも重複点は存在しない．さらに試行を増やしても，着弾点がそれまでの着弾点に一致する状況は，ほとんど起らないだろう．結局のところ，将来起りうる着弾点を全て網羅するには， $\#\Omega \rightarrow \infty$ の極限を考えるしかない．

というわけで，2.4.3 節の頻度解釈に乗り換えて，相対度数を求めると，

$$P_n(\{0\}) := \frac{N(\{0\})}{\#\Pi_n} = \frac{N(\{0\})}{n} \quad (3.7)$$

となる．ここで若干，楽観的に，

- 同じ着弾点は有限回しか繰り返さない (限りなくは繰り返さない)

と仮定しても，分子 $N(\{0\})$ が有限であるから，

$$P(\{0\}) := \lim_{n \rightarrow \infty} \frac{N(\{0\})}{n} = \frac{\text{有限}}{\infty} = 0 \quad (3.8)$$

となる．すなわち，発生頻度が有限な事象については，頻度の大小によらず，常に確率が 0 になってしまう．

以上，「将来起りうる」測定値が無数ある場合，測定値が定数に一致する確率は 0 になる．すなわち，定点 a に着目した確率 $P(Y = a)$ は，測定値が連続値であるようなランダム現象に対しては，常に同じ値をとるので意味をなさない．

3.3.2 事 象

そこで，起りうる測定値 y の範囲をピンポイントではなく，実数直線 R 上の区間で指定することにする．次のように表記する．

$$y \in [a, b] \quad \text{または} \quad a \leq y \leq b \quad (3.9)$$

▶▶ $[a, b]$ を閉区間といい， $y \in [a, b]$ は $a \leq y \leq b$ を意味する．同じく (a, b) を開区間といい， $y \in (a, b)$ は $a < y < b$ を意味する．同様に， $[a, b)$ や $(a, b]$ を半開区間という．

ここで改めて，測定値が区間に入る見本だけを集めた Ω の部分集合，

$$\begin{aligned} \{Y \in [a, b]\} \quad \text{または} \quad \{a \leq Y \leq b\} \\ := Y^{-1}([a, b]) = \{\omega \in \Omega \mid a \leq Y(\omega) \leq b\} \quad \subset \Omega \end{aligned} \quad (3.10)$$

を事象と定める．2.3.1 節で既に述べたことが異なる事象 $A_i := \{a_i \leq Y \leq b_i\}$ ($i = 1, 2, \dots$) を，表 2.1 のように集合演算した結果もまた事象であると約束する．

さて，図 3.3 の確率変数 $Y(\omega)$ の見本空間は $\Omega = \{0, 1, 2, \dots, 20\}$ であったが，そのなかで，例えば $|Y(\omega)| \leq 10^{-3}$ に該当する事象は，

$$\{|Y| < 10^{-3}\} = \{1, 2, 13\} \quad (3.11)$$

となる．これらを図 3.6 に赤○で示す．同じく， $\#\Omega = 501$ での結果を図 3.7 に示す．

3.3.3 確 率

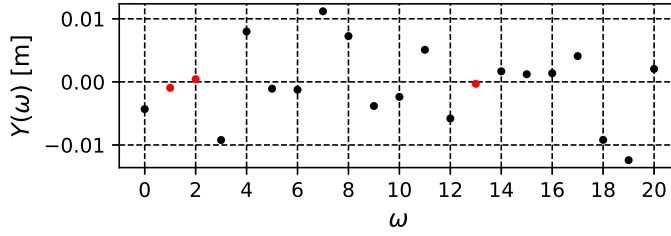
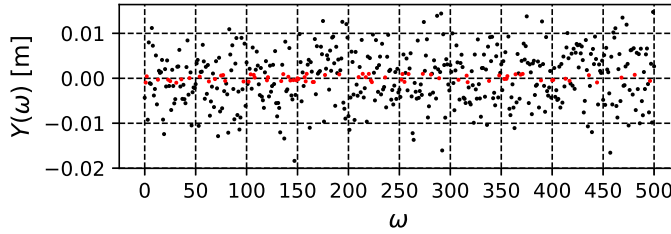
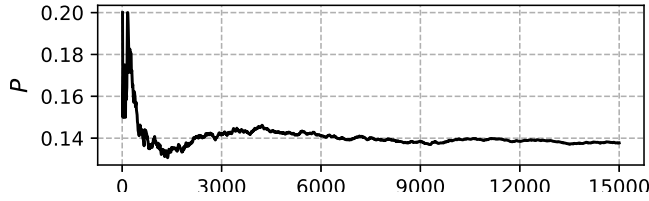
事象が実現する確率は，厳密には $P(\text{事象})$ と書くべきだが，表記短縮のためこれを略して $P(\text{条件})$ とも書くことにする．

事象 $\{a \leq Y \leq b\}$ が実現する確率は，ラプラスの確率を使うと，

$$P(a \leq Y \leq b) = \frac{\#\{a \leq Y \leq b\}}{\#\Omega} \quad (3.12)$$

となる． $\#A$ は集合 A の要素数を表す．

図 3.8 は，図 3.6 や図 3.7 と同様の試行結果から求めた確率 $P(|Y| \leq 10^{-3})$ の数値例である．横軸は分母となる全見本数 $\#\Omega$ であり，縦軸は得られた確率である．全見本数 $\#\Omega$ が


 図 3.6 事象 $\{Y \leq 10^{-3}\}$ の抽出 ($\#\Omega = 21$)

 図 3.7 事象 $\{Y \leq 10^{-3}\}$ の抽出 ($\#\Omega = 501$)

 図 3.8 確率 $P(Y \leq 10^{-3})$ の数値例

8000 を超えたあたりから，確率の計算値がある値に漸近していくように見える．区間で分子をカウントする (3.12) の定義では，分母と同時に分子も増えるから，このようなことが起きる．

実用的には，十分に漸近したと思われる計算値を，確率の近似値とすることが多い．右端の $\#\Omega = 15000$ においてもなお，可能な測定値が出尽したとは言い難いので，こうした数値計算値はあくまで近似値である．射撃の想定に戻ると，この (架空の) 実験においては，およそ 14% の弾丸が $y = 0 \pm 1\text{mm}$ の範囲に命中することが分かる．

また，区間の確率を考えるうえでは，次の性質が重要である．

算法 3.1 $\#\Omega = \infty$ の状況では，(1 点や 2 点の確率は 0 なので)

$$\begin{aligned} P(a \leq Y \leq b) &= P(a \leq Y < b) \\ &= P(a < Y \leq b) = P(a < Y < b) \end{aligned} \quad (3.13)$$

3.4 見本空間のタイプ

本章の射撃実験のような問題では，見本空間 Ω の要素は列挙できない．すなわち Ω はブラックボックスとなるが，理論上どのような見本空間を想定しうるか，いくつか例を挙げておく．

3.4.1 有限集合

測定値が有限個の値しか取らない場合の見本空間は有限集合となる．

$$\Omega := \{\omega_1, \omega_2, \dots, \omega_n\} \quad (n < \infty) \quad (3.14)$$

上述の数値例は厳密にいうと全てこれに該当する．なぜなら，コンピュータの数値は有限桁の小数なので，とりうる値は有限個である．また，統計学においても，無限回の試行を実施するのは不可能だから，有限な標本 Π_n から見本を抽出することになり，この場合に該当する．

3.4.2 可算無限集合

測定値が整数値や有理数値をとるとき，理論上起りうる全ての見本は，次のような形式で表せる．

$$\Omega := \{\omega_1, \omega_2, \dots\} \quad (3.15)$$

この見本空間の要素数は無限になるが，まだ番号付けは可能である（シリアル番号を振れる）． N を自然数の全体とすると，同じことは，

$$\Omega := \{\omega_i \mid i \in N\} \quad (3.16)$$

と書ける．

なお，このような番号付け可能な無限を可算無限という．

3.4.3 非可算無限集合

これに対して，測定値が実数のときは，(3.16) の見本空間では全てを網羅できない．そこで， R を実数の全体として，

$$\Omega := \{\omega_s \mid s \in R\} \quad (3.17)$$

というタイプの見本空間が導入される．

この見本空間は非常に大きくて，シリアル番号で全ての見本を網羅することはできない．このタイプの無限を非可算無限という．なにやらいかめしいが，力学的なランダム現象の見本空間はほとんどこのタイプであろう．

例えば，(3.1) のような運動方程式の解は実数である．それがランダムに揺らいだ解もまた実数である．ということは，図 3.2 の弾道を理論解析する際には，見本空間を非可算無限として想定しておかないと，可能性が漏れることになる．

3.5 数 値 例

- Ch03.ipynb

4

確率分布

ランダム現象における試行のばらつきを関数で表す．

4.1 分布関数と密度関数

図 3.5 の確率変数のグラフは無作為で理解しがたいが，見本 ω を測定値 $Y(\omega)$ の小さい順に並べると，だいぶ傾向が見やすくなる．図 3.5 に対して，実際にそうしたものを図 4.1 に示す． ω' は測定値が小さい順に振り直した見本番号である．

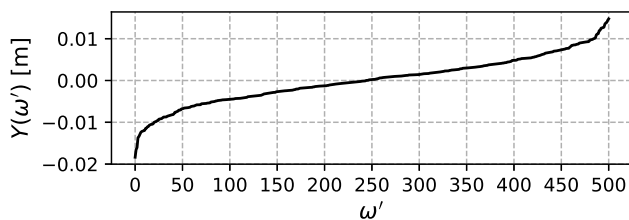


図 4.1 見本を昇順にソートした確率変数 ($x = 100$, $\#\Omega = 501$)

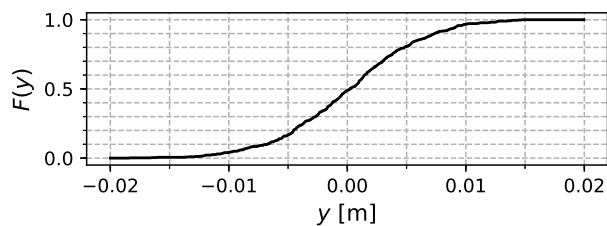


図 4.2 確率分布関数 ($x = 100$, $\#\Omega = 501$)

4.1.1 分布関数

図 4.1 の傾向を定量化するために, y を閾値とする次の事象を導入する.

$$\{Y \leq y\} \quad (4.1)$$

これは, 測定値が y 以下となる見本を集めた事象である.

次に y を等間隔にとって, 事象の見本数 $\#\{Y \leq y\}$ をカウントする. 例えば, 図 4.1 の結果を, y の刻みを,

$$y = -0.02, -0.015, -0.01, -0.005, 0, 0.005, 0.015, 0.02 \quad (4.2)$$

としてカウントすると, 表 4.1 のような結果が得られる.

表 4.1 累積度数と分布関数

y	$\#\{Y \leq y\}$	$F(y) := P(Y \leq y)$
-0.02	0	0.00e+00
-0.015	3	5.99e-03
-0.01	21	4.19e-02
-0.005	85	1.70e-01
0	244	4.87e-01
0.005	404	8.06e-01
0.01	485	9.68e-01
0.015	501	1.00e+00
0.02	501	1.00e+00

$X_e Y = X \times 10^Y$

表の 2 列目は測定値が y 以下となる見本数で, 累積度数と呼ぶ. また, 表の 3 列目は事象 $\{Y \leq y\}$ の確率であり, これを改めて,

$$F(y) := P(Y \leq y) \quad (4.3)$$

と書き, 確率分布関数または単に分布関数と呼ぶ.

y の刻み幅をさらに細かく ($\Delta y = 4 \times 10^{-5}$) して求めた分布関数を図 4.2 に示す.

算法 4.1 (分布関数の性質) 分布関数 $F(x)$ は一般に次の性質を持つ^{†2}.

(F1) $F(x)$ は単調増加.

(F2) $F(-\infty) = 0, \quad F(\infty) = 1.$

(F3) $P(a \leq x \leq b) = F(b) - F(a).$

演習 4.1 (F1) ~ (F3) を満足する $F(x)$ の概形を図示せよ.

^{†2} パボリス³⁾ p73 など.

4.1.2 ヒストグラム

図 4.2 の分布関数 $F(y)$ では，測定値の頻度が見えにくい．そこで， y 軸上に分点：

$$y_i := y_0 + i\Delta y \quad (i = 0, 1, 2, \dots, n) \quad (4.4)$$

をとり，隣り合う分点から連続的な背反事象：

$$A_i := \{y_i \leq y < y_{i+1}\} \quad (i = 0, 1, 2, \dots, n-1) \quad (4.5)$$

をとる．

事象 A_i に入る見本数 $\#A_i$ をカウントする．これらの列 $\#A_0, \#A_1, \dots$ を度数分布といい，そのグラフをヒストグラムという．表 4.1 と同じデータから求めた度数分布を表 4.2 の $\#A_i$ 列に示す．

表 4.2 ヒストグラムと近似的な密度関数

i	y_i	y_{i+1}	$\#A_i$	$P(A_i)$	$p^*(A_i)$
0	-0.02	-0.015	3	5.99e-03	1.20e+00
1	-0.015	-0.01	18	3.59e-02	7.19e+00
2	-0.01	-0.005	64	1.28e-01	2.55e+01
3	-0.005	0	159	3.17e-01	6.35e+01
4	0	0.005	160	3.19e-01	6.39e+01
5	0.005	0.01	81	1.62e-01	3.23e+01
6	0.01	0.015	16	3.19e-02	6.39e+00
7	0.015	0.02	0	0.00e+00	0.00e+00

$X \text{e} Y = X \times 10^Y$

次に，表の $P(A_i)$ 列は， $\#A_i$ を $\#\Omega$ で割って確率として表示したものであり，これを相対度数分布という．図 4.3 はそのグラフだが，これを相対度数ヒストグラムという．各矩形の底辺の範囲をビンと呼び，事象の区間を表す．矩形の高さが，測定値が各区間に入る確率 $P(A_i)$ である．

4.1.3 密 度 関 数

図 4.3 の矩形面積が確率 $P(A_i)$ になるよう高さ調整したものを，図 4.4 に示す．高さ調整は，

$$p^*(A_i) = \frac{P(A_i)}{\Delta y} \quad (i = 0, 1, 2, \dots, n-1) \quad (4.6)$$

とすればよい．具体値を表の $p^*(A_i)$ 列に示す．

大雑把にいうと，図 4.4 のグラフ（矩形の面積が確率を表すヒストグラム）形状を，確率密度関数または単に密度関数という．

正確には，試行数 $\rightarrow \infty$ ，ビンの幅 $\Delta y \rightarrow 0$ としたときに浮び上がる関数形：

$$p^*(A_i) \rightarrow p(\bar{y}_i) \quad (i = 0, 1, 2, \dots, n-1) \quad (4.7)$$

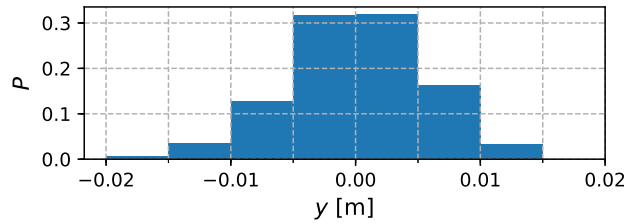


図 4.3 相対度数ヒストグラム ($x = 100$, $\#\Omega = 501$) .
高さが確率を表す .

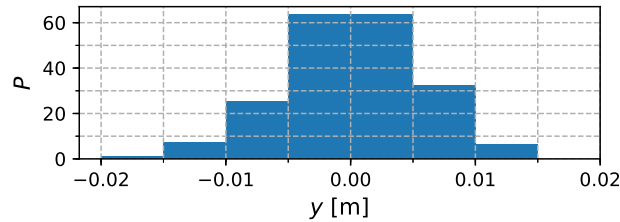


図 4.4 近似的な確率密度関数 ($x = 100$, $\#\Omega = 501$) .
面積が確率を表す .

を, 確率密度関数という . ここに, \bar{y}_i は区間の代表値である . なお, 比較的小さな n に対して近似的な密度関数を求めるときは, 中央値 :

$$\bar{y}_i := \frac{y_i + y_{i+1}}{2} \quad (i = 0, 1, 2, \dots, n-1) \quad (4.8)$$

で表示するとグラフの左右の偏りが防げる .

▶▶ 「試行数 $\rightarrow \infty$ 」は, 同じ測定値が (有限回しか) 繰り返さない試行については「見本数 $\#\Omega \rightarrow \infty$ 」と同義になる .

算法 4.2 (密度関数の性質) 密度関数 $p(x)$ は一般に次の性質を持つ^{†3} .

- (p1) $p(x) \geq 0$.
- (p2) $\int_{-\infty}^{\infty} p(x) dx = 1$.
- (p3) $P(a \leq X \leq b) = \int_a^b p(x) dx$.
- (p4) 分布関数 $F(x)$ が連続関数のとき, $p(x) = \frac{dF(x)}{dx}$.

演習 4.2 性質 (p1) ~ (p4) を図示せよ .

^{†3} パボリス³⁾ p76, p77 など .

4.2 代表的な確率分布

広く認知された典型的な確率分布をいくつか挙げておく．

4.2.1 一 様 分 布

次の密度関数を持つ分布を，一様分布という．分布形の一例を図 4.5 に示す．

$$p_U(x) := \begin{cases} \frac{1}{b-a} & (a \leq x \leq b) \\ 0 & (\text{otherwise}) \end{cases} \quad (4.9)$$

この分布は，区間 $[a, b]$ 上を均等に分布する．一様分布の身近な例として，各種プログラミング言語で提供される疑似乱数がある．

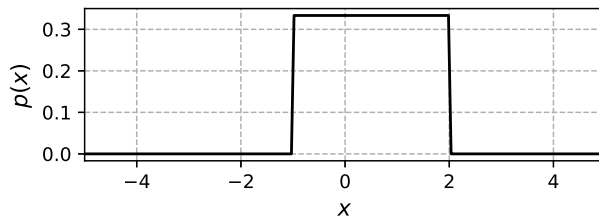


図 4.5 一様分布の密度関数の例 ($a = -1, b = 2$)

演習 4.3 $p_U(x)$ の非ゼロ部分の高さを，なぜ $1/(b-a)$ とするのか，理由を述べよ．

4.2.2 ガウス分布 (正規分布)

次の密度関数を持つ分布を，ガウス分布または正規分布という．分布形を図 4.6 に示す．

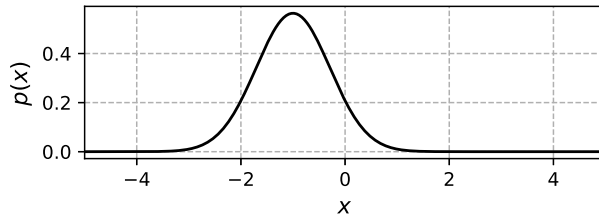
$$p_G(x) = \mathcal{N}(x; m, v) := \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x-m)^2}{2v}\right) \quad (4.10)$$

分布形の中心を定める m を平均値，広がり $v > 0$ を分散という．実現象にはこの分布にフィットするものが多いとされる．

ガウス分布を有する確率変数 X は，ガウス型[†]であるという．文脈に応じて，ガウ斯的，ガウス性という言い方もする．ガウス型の確率変数を略して，ガウス確率変数という．

演習 4.4 m と v を変えると， $p_G(x)$ の形状はどのように変化するか，略図で示せ．

[†] 「がうすがた」

図 4.6 ガウス分布の密度関数の例 ($m = -1, v = 0.5$)

4.2.3 ヘヴィータイル分布

ヘヴィータイル分布 (heavy-tailed distribution) とは、分布形の裾野が、ガウス分布より桁違いに厚いものをいう。近年、災害の発生頻度、経済データ、人間の平衡維持などがこのタイプの分布を持つことが指摘されている。

関数形を陽に書き下せるヘヴィータイル分布の一つに、ラプラス分布がある。分布形を図 4.7 に示す。

$$p_L(x) := \frac{1}{\sqrt{2v}} \exp\left(-\sqrt{\frac{2(x-m)^2}{v}}\right) \quad (4.11)$$

m は平均値, $v > 0$ は分散である。

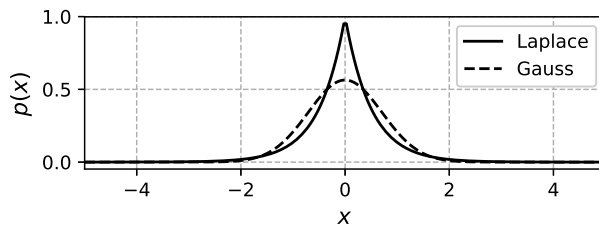
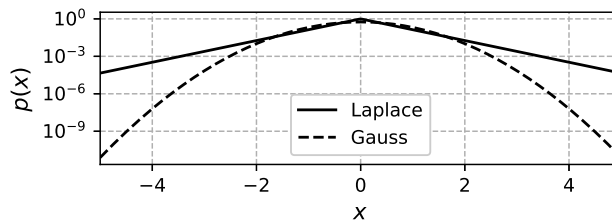
図 4.7 ラプラス分布の密度関数の例 ($m = 0, v = 0.5$)

図 4.8 図 4.7 の対数表示

縦軸を対数表示した図 4.8 を見るとよく分かるが、ラプラス分布は正規分布より桁違いに厚い裾野を持っている。例えば、 $x = 4$ において両者の確率密度を比較すると、ラプラス分

布は正規分布より約 5283 倍も高い密度を有している．両者を混同すると恐ろしいことになる．測定データの平均と分散を求め，これらと正規分布型の理論より「数万年に 1 度」と見積もられた事象 ($x = 4$) があったとしよう．ここでもし，本当の分布がラプラス分布だったとしたら，実際の頻度は「10 年に 1 度」程度になる．

4.3 数 値 例

- Ch04.ipynb

5 | 多変数の確率分布

分布関数と密度関数を多変数に拡張すると、独立性という概念が浮上してくる。

5.1 確率ベクトル

確率空間 (Ω, \mathcal{F}, P) において、 d 個の確率変数 (測定値) をとる。

$$X_1(\omega), X_2(\omega), \dots, X_d(\omega) \quad (5.1)$$

これらを成分とするベクトルを構成する。

$$X(\omega) := (X_1(\omega), X_2(\omega), \dots, X_d(\omega))^T \quad (5.2)$$

このベクトルを、 d 次元確率ベクトルという。すなわち、確率ベクトルとは、特定の見本 ω を代入すると、対応する d 次元測定値を返すベクトル値関数である。

▶▶ 確率ベクトルは、確率変数を並べたベクトルであり、確率 P を並べたものではない。

例えば、2 章のガチャの例でいうと、見本 ω の測定値を並べたベクトル、

$$X(\omega) = \begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix} = \begin{pmatrix} \text{全長 [m]} \\ \text{重量 [g]} \end{pmatrix} \quad (5.3)$$

を考え、これをガチャの確率ベクトルとするわけである。

以下、表記上、 $X(\omega)$ の ω を略して X と書くが、背後に ω がいることを忘れないでほしい。

5.2 2 次元の分布

簡単のため、2 次元 $X := (X_1, X_2)^T$ の場合を例にとる。

5.2.1 事象

確率変数 X では、実数軸上の区間から事象 $\{a \leq X \leq b\}$ を定義した。これの自然な拡張として、各 X_i について同様な区間 $D_i := [a_i, b_i]$ をとり、対応する事象をとる。

$$A_i = X_i^{-1}(D_i) = \{X_i \in D_i\} \\ := \{\omega \in \Omega \mid a_i \leq X_i(\omega) \leq b_i\} \quad (i = 1, 2) \quad (5.4)$$

ここで,

$$A_1 \cap A_2 = \{\omega \in \Omega \mid a_1 \leq X_1(\omega) \leq b_1, a_2 \leq X_2(\omega) \leq b_2\} \quad (5.5)$$

であるから, 積事象 $A_1 \cap A_2$ は, 測定値 $x = X(\omega)$ が, 実平面 \mathbb{R}^2 上の矩形領域 $[a_1, b_1] \times [a_2, b_2]$ に入る事象を表す. これらを含めた全事象の族を \mathcal{F} とする.

▶▶ 集合の掛け算表記 $[a_1, b_1] \times [a_2, b_2]$ の意味するところは, **A.1.5** 節で確認せよ.

5.2.2 結合確率

任意の2つの事象 $A, B \in \mathcal{F}$ に対して, 結合確率 $P(A, B)$ を

$$P(A, B) := P(A \cap B) \quad (5.6)$$

のように定義する. $P(A, B)$ は事象 A, B が同時に起る確率を表すので, 同時確率とも呼ばれる. この定義より,

$$P(A, B) = P(A \cap B) = P(B \cap A) = P(B, A) \quad (5.7)$$

となることに注意する.

表記短縮のため, 本来なら $P(\text{事象}, \text{事象})$ と書くべきところを, 略して $P(\text{条件}, \text{条件})$ とも書く.

5.2.3 分布関数

X_i の値が x_i 以下となる確率

$$F_i(x_i) := P(X_i \leq x_i) \quad (5.8)$$

を X_i の周辺分布関数という ($i = 1, 2$).

また, X_1 の値が x_1 以下となり, なおかつ X_2 の値が x_2 以下となる確率

$$F(x_1, x_2) := P(X_1 \leq x_1, X_2 \leq x_2) \quad (5.9)$$

を, X_1 と X_2 の結合分布関数という. 特別な場合を除いて,

$$F(x_1, x_2) \neq F(x_2, x_1) \quad (5.10)$$

であることに注意する.

演習 5.1 次の事象が指定する範囲を, (X_1, X_2) 平面上に図示せよ.

$$\{X_1 \leq x_1\}, \quad \{X_2 \leq x_2\}, \quad \{X_1 \leq x_1\} \cap \{X_2 \leq x_2\}$$

算法 5.1 (結合分布関数の性質) $F(x_1, x_2)$ は一般に次の性質を示す^{†2}.

- | | |
|------------------------------|--------------------------------|
| 1. $F(-\infty, -\infty) = 0$ | 4. $F(\infty, \infty) = 1$ |
| 2. $F(x_1, -\infty) = 0$ | 5. $F(x_1, \infty) = F_1(x_1)$ |
| 3. $F(-\infty, x_2) = 0$ | 6. $F(\infty, x_2) = F_2(x_2)$ |

例題 5.1 (2) と (5) を示せ.

演習 5.2 同様に (1)(3)(4)(6) を示せ.

5.2.4 密度関数

(x_1, x_2) 平面上に分布する測定値の面密度 $p(x_1, x_2)$ を, 結合密度関数という. 特に $F(x_1, x_2)$ が連続関数の場合は,

$$p(x_1, x_2) := \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2} \quad (5.11)$$

のように計算できる. 結合確率 $P(A, B)$ とは異なり, 特別な場合を除いて,

$$p(x_1, x_2) \neq p(x_2, x_1) \quad (5.12)$$

であることに注意する.

この密度関数 $p(x_1, x_2)$ を, x_1 や x_2 で積分した

$$\begin{cases} p_1(x_1) := \int_{-\infty}^{\infty} p(x_1, x_2) dx_2 \\ p_2(x_2) := \int_{-\infty}^{\infty} p(x_1, x_2) dx_1 \end{cases} \quad (5.13)$$

を, 周辺密度関数という. 実際に積分を実行すると,

$$\begin{aligned} p_1(x_1) &= \int_{-\infty}^{\infty} \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2} dx_2 = \frac{\partial F(x_1, \infty)}{\partial x_1} - \frac{\partial F(x_1, -\infty)}{\partial x_1} \\ &= \frac{\partial F_1(x_1)}{\partial x_1} - \frac{\partial 0}{\partial x_1} = \frac{\partial F_1(x_1)}{\partial x_1} \quad \because \text{算法 5.1 (2) (5)} \end{aligned} \quad (5.14)$$

となって, 周辺分布関数 $F_1(x_1)$ の微分に帰着する (算法 4.2 参照).

^{†2} パボリス³⁾ p132 など.

算法 5.2 (結合密度関数の性質) $p(x_1, x_2)$ は一般に次の性質を示す .

1. $p(x_1, x_2) \geq 0$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) dx_1 dx_2 = 1$
3. $P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} p(x_1, x_2) dx_1 dx_2$

5.2.5 2次元のガウス分布

次の結合密度関数で表される X_1, X_2 の分布を , 2 変数ガウス分布という[†] .

$$p(x_1, x_2) := \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} \right. \\ \left. \times \left(\frac{(x_1 - m_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - m_1)(x_2 - m_2)}{\sigma_1\sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2} \right) \right] \quad (5.15)$$

この密度関数の形状は $m_1, m_2, \sigma_1, \sigma_2, \rho$ という 5 つのパラメータで定まるが , m_i を X_i の平均 , $\sigma_i > 0$ を X_i の標準偏差 , ρ を X_1, X_2 の相関係数という . σ_i^2 を X_i の分散という .

算法 5.3 X_i の周辺密度関数 $p_i(x_i)$ は , 平均 m_i , 分散 σ_i^2 のガウス分布となる .

$$p_i(x_i) = \mathcal{N}(x_i; m_i, \sigma_i^2) \\ = \frac{1}{\sigma_i\sqrt{2\pi}} \exp \left(-\frac{(x_i - m_i)^2}{2\sigma_i^2} \right) \quad (i = 1, 2) \quad (5.16)$$

例題 5.2 示せ . ($u_i := (x_i - m_i)/\sigma_i$ において , (5.15) の指数部分を計算する)

5.3 独立 性

2 つの事象が何らかの意味で「無関係」なことを定式化したい .

5.3.1 事象の独立性

そのために確率論では , 独立性の概念を導入する .

定義 5.1 2 つの事象 $A, B \in \mathcal{F}$ の結合確率が ,

[†] パボリス³⁾ p135 など .

$$P(A, B) = P(A)P(B) \quad (5.17)$$

となるとき, A と B は独立であるという.

▶▶ (背反事象 \neq 独立) 背反事象 $A \cap B = \emptyset$ は, 互いに無関係とはいえない. なぜなら, A が起きれば B は起らず, B が起きれば A は起らないという意味で, 互いに依存関係にある. 実際に計算してみると, $P(A), P(B) \neq 0$ のとき,

$$P(A, B) := P(A \cap B) = P(\emptyset) = 0 \neq P(A)P(B) \quad (5.18)$$

となる. すなわち, 確率が 0 でない背反事象 A, B は互いに独立ではない.

5.3.2 条件付き確率

2 つの事象 $A, B \in \mathcal{F}$ について, 次の確率を導入する.

$$P(A|B) := \frac{P(A, B)}{P(B)} \quad (P(B) \neq 0) \quad A|B \text{ の読み方「} A \text{ パイプ } B \text{」} \quad (5.19)$$

これを, 条件 B における A の条件付き確率と呼び, 条件 B のなかで A が発生する確率を表す[†]. 右辺は, ラプラスの確率を使うと,

$$P(A|B) := \frac{P(A, B)}{P(B)} = \frac{\frac{\#(A \cap B)}{\#\Omega}}{\frac{\#B}{\#\Omega}} = \frac{\#(A \cap B)}{\#B} \quad (5.20)$$

のように解釈できる.

演習 5.3 $P(A) = \#A/\#\Omega$ と $P(A|B)$ の違いをベン図で説明せよ.

算法 5.4 (条件付き確率の性質) 条件付き確率は一般に次の性質を示す.

1. 乗法定理: $P(A, B) = P(A|B)P(B)$, $P(B, A) = P(B|A)P(A)$
2. ベイズの定理: $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$

演習 5.4 示せ. (1) は (5.19) の分母を払えばよい. (2) は, (5.7) より $P(A, B) = P(B, A)$ であることを利用して, 乗法定理の 2 式を等値すると得られる.

条件付き確率を使うと, 独立性の意味がより明確になる.

算法 5.5 事象 A, B が互いに独立ならば, 条件付き確率は次の性質を示す.

$$P(A|B) = P(A) \quad (P(B) \neq 0) \quad (5.21)$$

$$P(B|A) = P(B) \quad (P(A) \neq 0) \quad (5.22)$$

すなわち, A の確率に B は影響せず, B の確率に A は影響しない.

[†] 例えば, 男子の留年率 $P(\text{留年} | \text{男子})$ の分母は, 男子の総数. (< 学生総数)

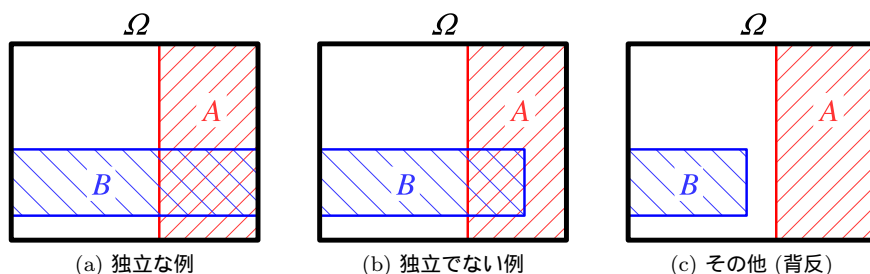


図 5.1 事象の独立性の例

演習 5.5 (5.17) を (5.19) に代入し，算法 5.5 を示せ．

(5.21) の意味を考えてみる．両辺をラプラスの確率で解釈すると，

$$\frac{\#(A \cap B)}{\#B} = \frac{\#A}{\#\Omega} \quad (5.23)$$

となる．この等式の成否を図 5.1 に模式的に示す．黒枠で見本空間 Ω を表す．そこに無数の見本を均等にばら撒いたとしよう．このとき，確率 = 面積の割合 となる．

図 5.1 (a) は A, B が独立な場合で，このような配置に対しては (5.23) が成立する．すなわち，青 B に占める赤 A の割合 $P(A|B)$ と，全体 Ω に占める赤 A の割合 $P(A)$ は一致する．これに対して，図 5.1 (b) の A, B は独立でない．なぜなら，青に占める赤の割合と，全体に占める赤の割合が一致せず，(5.23) は成立しない．

5.3.3 確率変数の独立性

事象の独立性を，確率変数から逆算した事象に適用すると，確率変数の独立性が定式化できる．

定義 5.2 確率変数 X_1, X_2 の値域の任意の区間 D_1, D_2 に対して，

$$P(X_1 \in D_1, X_2 \in D_2) := P(X_1 \in D_1)P(X_2 \in D_2) \quad (5.24)$$

が成立するとき， X_1 と X_2 は独立であるという．

この定義の意味するところは，図 5.1 の黒枠の底辺を X_1 軸，左の側辺を X_2 軸とすることで同様に理解できる．

独立な確率変数 X_1, X_2 の分布関数は，

$$\begin{aligned} F(x_1, x_2) &:= P(X_1 \leq x_1, X_2 \leq x_2) \\ &= P(X_1 \leq x_1)P(X_2 \leq x_2) = F_1(x_1)F_2(x_2) \end{aligned} \quad (5.25)$$

のように，周辺分布の積になる．この両辺を x_1 と x_2 で偏微分すると，

$$\begin{aligned} p(x_1, x_2) &= \frac{\partial^2 F_1(x_1)F_2(x_2)}{\partial x_1 \partial x_2} \\ &= \frac{\partial F_1(x_1)}{\partial x_1} \frac{\partial F_2(x_2)}{\partial x_2} = p_1(x_1)p_2(x_2) \end{aligned} \quad (5.26)$$

となり，結合密度関数は周辺密度関数の積になる．以上をまとめて次の公式を得る．

算法 5.6 互いに独立な確率変数 X_1, X_2 の結合分布は，周辺分布の積で書ける．

1. $F(x_1, x_2) = F_1(x_1)F_2(x_2)$
2. $p(x_1, x_2) = p_1(x_1)p_2(x_2)$

5.3.4 条件付き密度関数

確率変数 X_1, X_2 の結合密度関数 $p(x_1, x_2)$ を， X_2 の周辺密度関数 $p_2(x_2)$ で割った，

$$p(x_1|x_2) := \frac{p(x_1, x_2)}{p_2(x_2)} \quad (5.27)$$

を，条件 $X_2 = x_2$ における X_1 の条件付き密度関数という．

例えば，2 変数ガウス分布 (5.15) について，条件付密度関数 $p(x_1|x_2)$ を求めると，周辺密度関数 (5.16) より，

$$\begin{aligned} p(x_1|x_2) &:= \frac{p(x_1, x_2)}{p_2(x_2)} \\ &= \frac{1}{\sigma_1 \sqrt{1-\rho^2} \sqrt{2\pi}} \times \exp \left[\frac{(x_2 - m_2)^2}{2\sigma_2^2} \right. \\ &\quad \left. - \frac{1}{2(1-\rho^2)} \left(\frac{(x_1 - m_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - m_1)(x_2 - m_2)}{\sigma_1 \sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2} \right) \right] \\ &= \cdots \times \exp \left[- \frac{\frac{(x_1 - m_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - m_1)(x_2 - m_2)}{\sigma_1 \sigma_2} + \rho^2 \frac{(x_2 - m_2)^2}{\sigma_2^2}}{2(1-\rho^2)} \right] \\ &= \cdots \times \exp \left[- \frac{1}{2(1-\rho^2)} \left(\frac{x_1 - m_1}{\sigma_1} - \rho \frac{x_2 - m_2}{\sigma_2} \right)^2 \right] \\ &= \cdots \times \exp \left[- \frac{1}{2(1-\rho^2)\sigma_1^2} \left(x_1 - m_1 - \rho \frac{\sigma_1}{\sigma_2} (x_2 - m_2) \right)^2 \right] \\ &= \mathcal{N} \left(x_1; m_1 + \rho \frac{\sigma_1}{\sigma_2} (x_2 - m_2), \sigma_1^2 (1-\rho^2) \right) \end{aligned} \quad (5.28)$$

すなわち，平均 $m_1 + \rho(\sigma_1/\sigma_2)(x_2 - m_2)$ ，分散 $\sigma_1^2(1-\rho^2)$ の正規分布となる．

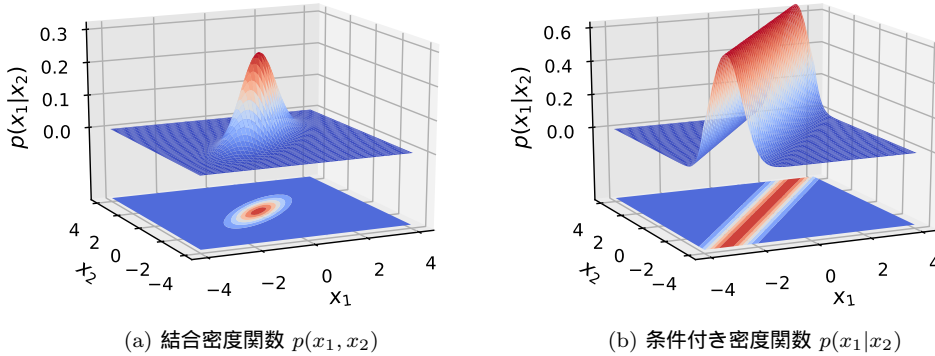


図 5.2 2 変数ガウス分布の結合密度関数と条件付き密度関数 ($m_1 = m_2 = 0, \sigma_1 = \sigma_2 = 1, \rho = 0.8$)

演習 5.6 同様に $p(x_2|x_1)$ を求めよ .

図 5.2 (a) は 2 変数ガウス分布 (5.15) の結合密度関数 $p(x_1, x_2)$ であり, (b) は対応する (5.28) の条件付き密度関数 $p(x_1|x_2)$ である . 例えば, $(x_1, x_2) = (-3, -4)$ の付近 (手前) に着目すると, $p(-3, -4) \approx 8.42 \times 10^{-5}$ であり, $(x_1, x_2) = (-3, -4)$ の付近の測定値は極めて稀にしか発生しないことが分かる .

これに対して, $p(-3|-4) := p(-3, -4)/p_2(-4)$ は, ごく稀に $x_2 = -4$ 付近の測定値が得られた場合に ($p_2(-4) = 1.34 \times 10^{-4}$), もう 1 つの測定値が $x_1 = -3$ 付近にくる確率密度である . その値は $p(-3|-4) \approx 0.629$ であり, そうなる頻度は (この例では) 比較的高いことが分かる .

算法 5.7 (条件付き密度関数の性質) 一般に次の性質を示す .

1. 条件 x_2 によらず, $\int_{-\infty}^{\infty} p(x_1|x_2)dx_1 = 1$
2. 乗法定理: $p(x_1, x_2) = p(x_1|x_2)p_2(x_2)$

5.4 多次元の分布

d 次元確率ベクトル $X = (X_1, \dots, X_d)^T$ を考える .

5.4.1 密度関数

d 個の確率変数 X_1, \dots, X_d が共存するので, 密度関数は d 変数となる .

$$p(x_1, \dots, x_d) \quad (5.29)$$

5.4.2 周辺密度関数

$d = 4$ で例示すると,

$$p_2(x_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2, x_3, x_4) dx_1 dx_3 dx_4 \quad (5.30)$$

$$p_{13}(x_1, x_3) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2, x_3, x_4) dx_2 dx_4 \quad (5.31)$$

$$p_{134}(x_1, x_3, x_4) = \int_{-\infty}^{\infty} p(x_1, x_2, x_3, x_4) dx_2 \quad (5.32)$$

のようなものを考える.

5.4.3 独立性

同じく, $d = 4$ で例示すると,

- X_1 と X_2, X_3, X_4 が独立 $p(x_1, x_2, x_3, x_4) = p_1(x_1)p_{234}(x_2, x_3, x_4)$
- X_1, X_3 と X_2, X_4 が独立 $p(x_1, x_2, x_3, x_4) = p_{13}(x_1, x_3)p_{24}(x_2, x_4)$

5.4.4 条件付き密度関数

同じく,

$$p(x_1, x_2, x_3 | x_4) := \frac{p(x_1, x_2, x_3, x_4)}{p_4(x_4)} \quad (5.33)$$

$$p(x_1, x_3 | x_2, x_4) := \frac{p(x_1, x_2, x_3, x_4)}{p_{24}(x_2, x_4)} \quad (5.34)$$

$$p(x_3 | x_1, x_2, x_4) := \frac{p(x_1, x_2, x_3, x_4)}{p_{124}(x_1, x_2, x_4)} \quad (5.35)$$

などとなる.

例題の解答例

例題 5.1 p35

$$\begin{aligned}
 (2) \text{ について, } F(x_1, -\infty) &:= P(\{X_1 \leq x_1\} \cap \{X_2 \leq -\infty\}) \\
 &= P(\{X_1 \leq x_1\} \cap \emptyset) = P(\emptyset) = 0
 \end{aligned}$$

$$\begin{aligned}
 (5) \text{ について, } F(x_1, \infty) &:= P(\{X_1 \leq x_1\} \cap \{X_2 \leq \infty\}) \\
 &= P(\{X_1 \leq x_1\} \cap \Omega) = P(\{X_1 \leq x_1\}) =: F_1(x_1)
 \end{aligned}$$

例題 5.2 p36

(5.15) の指数部分 E は, $u_i := (x_i - m_i)/\sigma_i$ と置くと,

$$\begin{aligned}
 E &= -\frac{1}{2(1-\rho^2)}(u_1^2 - 2\rho u_1 u_2 + u_2^2) \\
 &= -\frac{1}{2(1-\rho^2)}((u_1 - \rho u_2)^2 + u_2^2 - \rho^2 u_2^2) = -\frac{(u_1 - \rho u_2)^2}{2(1-\rho^2)} - \frac{u_2^2}{2}
 \end{aligned}$$

周辺密度関数を置換積分で整理していくと, $dx_i = \sigma_i du_i$ より,

$$\begin{aligned}
 p_2(x_2) &:= \int_{-\infty}^{\infty} p(x_1, x_2) dx_1 = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp(E) dx_1 \\
 &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{(u_1 - \rho u_2)^2}{2(1-\rho^2)}\right) \sigma_1 du_1 \cdot \exp\left(-\frac{u_2^2}{2}\right) \\
 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{1-\rho^2}\sqrt{2\pi}} \exp\left(-\frac{(u_1 - \rho u_2)^2}{2(1-\rho^2)}\right) du_1 \cdot \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{u_2^2}{2}\right) \\
 &= \underbrace{\int_{-\infty}^{\infty} \mathcal{N}(u_1; \rho u_2, (1-\rho^2)) du_1}_{=1} \cdot \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{u_2^2}{2}\right)
 \end{aligned}$$

6 | 期 待 値

確率変数の値 (測定値) のばらつき具合を、いくつかの指標で表す。

6.1 確率変数の期待値

以下、確率変数 $X(\omega)$ を略して X と表記する。

6.1.1 密度関数 $p(x)$ が既知の場合

確率変数 X の確率密度関数を $p(x)$ とする。 $p(x)$ を x 軸上に分布する物質の線密度とみなしたときの重心：

$$m_X = \langle X \rangle := \int_{-\infty}^{\infty} x p(x) dx \quad \left(= \frac{\int_{-\infty}^{\infty} x p(x) dx}{\int_{-\infty}^{\infty} p(x) dx = 1} \right) \quad (6.1)$$

を、 X の期待値または平均値という。以下、 $\langle \cdots \rangle$ を平均操作と呼ぶことにする。 m_X は X の値のばらつきの重心であるから、ばらつきの中心を表す指標になる。

演習 6.1 $p(x)$ が一様分布 (4.9) であるときの期待値を求めよ。

同様に、スカラー値関数 $f(\cdot)$ を介した、

$$\langle f(X) \rangle := \int_{-\infty}^{\infty} f(x) p(x) dx \quad (6.2)$$

を、 $f(X)$ の期待値または平均値という。特に、 $f(x) = x^n$ としたもの

$$m_X^{(n)} := \langle X^n \rangle = \int_{-\infty}^{\infty} x^n p(x) dx \quad (n = 1, 2, 3, \dots) \quad (6.3)$$

を n 次モーメントと呼ぶ。これを期待値との差で計算したもの、

$$\begin{aligned} s_X^{(n)} &:= \langle (X - m_X)^n \rangle \\ &= \int_{-\infty}^{\infty} (x - m_X)^n p(x) dx \quad (n = 1, 2, 3, \dots) \end{aligned} \quad (6.4)$$

を n 次中心モーメントと呼ぶ。特に、 $n = 2$ のときの中心モーメント、

$$\sigma_X^2 := s_X^{(2)} = \langle (X - m_X)^2 \rangle \quad (6.5)$$

を分散という．分散は平均値からの「ばらつき」を表す指標としてよく用いられる．

同じく「ばらつき」を表す量として，分散の正の平方根 $\sigma_X := \sqrt{\sigma_X^2}$ もよく用いられる．これを標準偏差という．標準偏差は「ばらつき」を確率変数と同じ単位で表す．例えば， X [m] の分散の単位は σ_X^2 [m²] だが，標準偏差の単位は σ_X [m] となる．

演習 6.2 $p(x)$ が一様分布 (4.9) であるときの分散を求めよ．

算法 6.1 (ガウス分布の期待値と分散) (4.10) のガウス分布 $p(x) = \mathcal{N}(x; m, v)$ の期待値と分散は，

$$\begin{aligned} \langle X \rangle &= \int_{-\infty}^{\infty} x p(x) dx = m, \\ \langle (X - m)^2 \rangle &= \int_{-\infty}^{\infty} (x - m)^2 p(x) dx = v \end{aligned} \quad (6.6)$$

6.1.2 密度関数 $p(x)$ が未知の場合

確率変数 $x = X(\omega)$ の試行列 x_1, x_2, \dots から標本

$$\Pi_n := \{x_1, x_2, \dots, x_n\} \quad (6.7)$$

をとる． n は試行回数である．このとき，

$$\langle X \rangle_n := \frac{1}{n} \sum_{i=1}^n x_i \quad (6.8)$$

を， X の標本平均という．同様に，適当なスカラー値関数 $f(\cdot)$ を介した，

$$\langle f(X) \rangle_n := \frac{1}{n} \sum_{i=1}^n f(x_i) \quad (6.9)$$

を， $f(X)$ の標本平均という．

単に平均と言わずに「標本」をつけるのは，標本を変えると $\langle X \rangle_n$ の値が変化するためである．例えば，標本 Π_n を別の標本 $\Pi'_n := \{x_{n+1}, x_{n+2}, \dots, x_{2n}\}$ に置き換えると， $\langle X \rangle_n$ は変化する．また標本数 n を変えても $\langle X \rangle_n$ は変化する．

ところで，分散 σ_X^2 に対応する標本平均は，

$$\langle (X - m)^2 \rangle_n := \frac{1}{n} \sum_{i=1}^n (x_i - m)^2, \quad m = \langle X \rangle_n \quad (6.10)$$

であり、これを標本分散というが、 n が小さいときは理論値 σ_X^2 より若干小さな値をとる。そこで、分母を補正した、

$$\langle (X - m)^2 \rangle_n^* := \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2 \quad (6.11)$$

を用いることがあり、これを不偏分散という。標本数 n が十分大きければ両者はほぼ等しい。

6.1.3 標本平均の性質

標本平均 $\langle X \rangle_n$ は、理論的な平均 $\langle X \rangle$ の近似値を与える。次の定理が成立する。

算法 6.2 (大数の法則)^{†1} 期待値 $\langle X \rangle$ が存在する確率変数 X について^{†2},

$$\langle X \rangle_n \rightarrow \langle X \rangle \quad (n \rightarrow \infty) \quad (6.12)$$

また、平均値 $\langle X \rangle$ まわりの $\langle X \rangle_n$ のばらつきについては、次の定理が成立する。

算法 6.3 (中心極限定理) X を平均 m 、分散 σ の確率変数とする。 X の試行列から複数の異なる標本 Π_n^1, Π_n^2, \dots をとり、それらの標本平均

$$m^1 = \langle X \rangle_n^1, m^2 = \langle X \rangle_n^2, \dots \quad (6.13)$$

を考える。 m^1, m^2, \dots はランダムな試行列となり、その分布は、 $n \rightarrow \infty$ の極限で平均 m 、分散 σ^2/n のガウス分布に近づく。

▶▶ 算法 6.3 では、 X をガウス型としていない。ようするに、どんな分布の確率変数でも、いっぱい足すと、ガウス分布に近づくということ。

以上をまとめると、標本平均 $\langle X \rangle_n$ の n を十分に大きくとると、

- 算法 6.2 により、 X の相異なる試行列から計算した標本平均の重心は、どれも理論値 $\langle X \rangle$ に近づく。
- さらに算法 6.3 で、 $\sigma^2/n \rightarrow 0$ ($n \rightarrow \infty$) だから、標本平均の分散は 0 に近づく。

ということが分かる。すなわち、期待値が存在する確率変数の標本平均は、標本数 n を増やしていくと、理論値の近くに集積していく。

^{†1} たいすうの法則と読む

^{†2} 期待値が定まらない分布もあるので仮定が必要。例えば、コーシー分布は極限の取り方で期待値が変化する。

6.2 確率ベクトルの期待値

簡単のため, 2 次元 $X := (X_1, X_2)^T$ の場合を例にとる.

6.2.1 多変数の期待値

単独の確率変数との違いとして, 密度関数を結合密度関数 $p(x_1, x_2)$ とし, 積分を 2 重積分とする. すなわち, $f(X_1, X_2)$ に対する平均操作を,

$$\langle f(X_1, X_2) \rangle := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) p(x_1, x_2) dx_1 dx_2 \quad (6.14)$$

で定める. 代表的な期待値 (広義) とその記号や名称を表 6.1 に示す.

表 6.1 代表的な期待値 (広義)

名称	記号 1	記号 2	定義
X_i の期待値 (狭義)	$E[X_i]$	m_i	$\langle X_i \rangle$
X_i の n 次モーメント	—	$m_i^{(n)}$	$\langle X_i^n \rangle$
X_i の n 次中心モーメント	—	$s_i^{(n)}$	$\langle (X_i - m_i)^n \rangle$
X_i の分散 ($= s_i^{(2)}$)	$V[X_i]$	σ_i^2	$\langle (X_i - m_i)^2 \rangle$
X_i の標準偏差	$SD[X_i]$	σ_i	$\sqrt{\sigma_i^2}$
X_i, X_j の結合モーメント	—	m_{ij}	$\langle X_i X_j \rangle$
X_i, X_j の共分散	$\text{Cov}[X_i, X_j]$	C_{ij}	$\langle (X_i - m_i)(X_j - m_j) \rangle$
X_i, X_j の相関係数	—	ρ_{ij}	$C_{ij}/(\sigma_i \sigma_j)$

相関係数 ρ_{ij} は, 標準偏差 σ_i, σ_j で規格化した確率変数 $X'_i := X_i/\sigma_i$, $X'_j := X_j/\sigma_j$ の共分散を表す.

$$\rho_{ij} = \text{Cov}[X'_i, X'_j] \quad (6.15)$$

なお, ベクトル $X = (X_1, X_2)^T$ に対する平均操作は, 各成分に対する平均操作

$$\left\langle \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \right\rangle := \begin{pmatrix} \langle X_1 \rangle \\ \langle X_2 \rangle \end{pmatrix} \quad (6.16)$$

を意味すると約束する. 2 次元の期待値 $m := (\langle X_1 \rangle, \langle X_2 \rangle)^T$ は, $p(x_1, x_2)$ を (x_1, x_2) 平面上に分布する物質の面密度とみなしたときの重心ベクトルを与える.

6.2.2 期待値に関する公式

(1) 加法とスカラ倍 積分の性質 $\iint (f+g) dx_1 dx_2 = \iint f dx_1 dx_2 + \iint g dx_1 dx_2$,
 $\iint \alpha f dx_1 dx_2 = \alpha \iint f dx_1 dx_2$ より, 次の公式が成立する.

算法 6.4 (平均の公式)

- (1) $\langle f(X_1, X_2) + g(X_1, X_2) \rangle = \langle f(X_1, X_2) \rangle + \langle g(X_1, X_2) \rangle$
- (2) $\langle \alpha f(X_1, X_2) \rangle = \alpha \langle f(X_1, X_2) \rangle$ (α は定数)

算法 6.4 より，次の公式が従う．

算法 6.5 (分散の公式)

- (1) $V[X_i] = \langle X_i^2 \rangle - m_i^2$
- (2) $\text{Cov}[X_1, X_2] = \langle X_1 X_2 \rangle - m_1 m_2$
- (3) $\rho_{ij} = \text{Cov}[X_i, X_j] / (\sigma_i \sigma_j)$

例題 6.1 (1) を示せ．

演習 6.3 同様に (2) を示せ．

演習 6.4 同様に (3) を示せ．(ヒント：(6.15) の $\langle (X'_i - m'_i)(X'_j - m'_j) \rangle$ に $m'_i = \langle X'_i \rangle = \langle X_i / \sigma_i \rangle = m_i / \sigma_i$ を代入して展開する)

算法 6.6 (相関係数の性質)

$$-1 \leq \rho_{ij} \leq 1 \quad (6.17)$$

例題 6.2 示せ．

(2) 低次元化 1 変数の平均操作は周辺密度で簡単にできる．

算法 6.7 X_i の周辺密度関数 $p_i(x_i)$ が既知のとき， $f(X_i)$ の期待値は，

$$\langle f(X_i) \rangle = \int_{-\infty}^{\infty} f(x_i) p_i(x_i) dx_i \quad (i = 1, 2) \quad (6.18)$$

例題 6.3 示せ．

6.3 共分散と相関

定義 6.1 (無相関) 確率変数 X_1, X_2 が無相関であるとは，共分散が 0 になること

$$\text{Cov}[X_i, X_j] = 0 \quad (6.19)$$

をいう．特に $\sigma_i, \sigma_j \neq 0$ のときは，相関係数が 0 になること

$$\rho_{ij} = 0 \quad (6.20)$$

と同じである．

$\text{Cov}[X_i, X_j] > 0$ のとき， X_i と X_j には正の相関があるといい， $\text{Cov}[X_i, X_j] < 0$ のときは負の相関があるという．このとき， X_i と X_j の間には正または負の比例関係が見い出される．逆に， X_i, X_j が無相関ならば比例関係は見出せない．

6.3.1 独立 \implies 無相関

X_1, X_2 を独立な確率変数とする．このとき結合密度関数は，算法 5.6 より， $p(x_1, x_2) = p_1(x_1)p_2(x_2)$ となるので，

$$\begin{aligned} \text{Cov}[X_1, X_2] &:= \langle (X_1 - m_1)(X_2 - m_2) \rangle \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - m_1)(x_2 - m_2)p_1(x_1)p_2(x_2)dx_1dx_2 \\ &= \int_{-\infty}^{\infty} (x_1 - m_1)p_1(x_1)dx_1 \int_{-\infty}^{\infty} (x_2 - m_2)p_2(x_2)dx_2 \\ &= m_1 - m_1 + m_2 - m_2 = 0 \quad (6.21) \end{aligned}$$

となる．すなわち，次がいえる．

算法 6.8 X_1, X_2 が独立 $\implies X_1, X_2$ は無相関．

しかしながら，その逆は一般には成立しない．

6.3.2 無相関だが独立でない例

よく知られた例を挙げよう．平均 0 分散 1 のガウス確率変数 X に対して， $Y := X^2$ をとる． Y は X の 2 次関数なので，明らかに X とは独立ではない．

しかし，この Y は X と無相関になる．共分散を計算してみよう． X と Y の結合密度関数を $p(x, y)$ と書く．まずは，定義にしたがって，

$$\text{Cov}[X, Y] = \langle XY \rangle - \underbrace{\langle X \rangle}_{=0} \langle Y \rangle = \iint xy p(x, y) dx dy \quad (6.22)$$

までいく．所与の関係式 $y = x^2$ は積分以前の関係式なので，そのまま代入できて，

$$= \iint x^3 p(x, y) dx dy = \langle X^3 \rangle = 0 \quad (6.23)$$

が得られる．最後は，平均 0 のガウス確率変数の奇数次モーメントは 0 になることを用いた．
以上，共分散が 0 なので， X と Y は (独立ではないが) 無相関である．

6.3.3 ガウス確率変数なら，独立 \iff 無相関

ただし，ガウス確率変数に話を限ると逆も成立する． X_1, X_2 が 2 変数ガウス分布 (5.15) に従うとする．無相関 $\rho_{12} = 0$ を仮定して， $\rho = \rho_{12} = 0$ を (5.15) に代入すると，

$$\begin{aligned} p(x_1, x_2) &:= \frac{1}{2\pi\sigma_1\sigma_2} \exp \left[-\frac{1}{2} \times \left(\frac{(x_1 - m_1)^2}{\sigma_1^2} + \frac{(x_2 - m_2)^2}{\sigma_2^2} \right) \right] \\ &= \frac{1}{\sigma_1\sqrt{2\pi}} \exp \left[-\frac{(x_1 - m_1)^2}{2\sigma_1^2} \right] \frac{1}{\sigma_2\sqrt{2\pi}} \exp \left[-\frac{(x_2 - m_2)^2}{2\sigma_2^2} \right] \\ &= p_1(x_1)p_2(x_2) \end{aligned} \quad (6.24)$$

となって，独立性 (算法 5.6) が帰結される． $p_i(x_i)$ は算法 5.3 の周辺分布である．

これと一般論「独立 \implies 無相関」を組み合わせると，

算法 6.9 X_1, X_2 がガウス確率変数ならば，

$$X_1, X_2 \text{ が独立} \xLeftrightarrow[\text{十分}]{\text{必要}} X_1, X_2 \text{ は無相関} .$$

6.4 条件付き期待値

6.4.1 関数としての定義

$Y = y$ を条件とする X の条件付き密度関数 $p(x|y)$ による期待値，

$$E[X|y] = \langle X|y \rangle := \int_{-\infty}^{\infty} x p(x|y) dx \quad (6.25)$$

を考える．変数 y が残されているので， $E[X|y]$ は y の関数である．これを，条件 $Y = y$ における X の条件付き期待値という．

$E[X|y]$ は，確率ベクトル (X, Y) の試行列

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots \quad (6.26)$$

において，与えられた定数 $y_i = y$ を検出したときの相方 x_i の期待値を表す[†]．

算法 6.10 確率変数 X_1, X_2 の結合密度関数 $p(x_1, x_2)$ が 2 変数ガウス分布 (5.15) のとき，条件付き期待値 $E[X_1|X_2]$ の関数形 $\phi(x_2) = E[X_1|x_2]$ は，

$$\begin{aligned}\phi(x_2) &= E[X_1|x_2] = m_1 + \rho_{12} \frac{\sigma_1}{\sigma_2} (x_2 - m_2) \\ &= m_1 + \frac{C_{12}}{\sigma_2^2} (x_2 - m_2)\end{aligned}\quad (6.27)$$

演習 6.5 示せ．(5.28) p39 の条件付き密度 $p(x_1|x_2)$ の期待値 (6.6) を求めればよい．

6.4.2 確率変数としての定義

(6.25) の定義には，実は，論理上の問題がある．なぜなら，確率変数 Y の値が定数 y に一致する確率は 0 なので，実際に $Y = y$ を検出する可能性は無きに等しい．

そこで，次の計算をヒントにする．

$$\begin{aligned}\langle X \rangle &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xp(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xp(x|y)p(y) dx dy \quad \because \text{算法 5.7 p40(2)} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xp(x|y) dx p(y) dy = \int_{-\infty}^{\infty} E[X|y] p(y) dy\end{aligned}\quad (6.28)$$

最後の式は， $\phi(y) = E[X|y]$ と置くと，確率変数 Y の関数 $\phi(Y)$ の期待値

$$\int_{-\infty}^{\infty} \phi(y) p(y) dy = \langle \phi(Y) \rangle = \langle E[X|Y] \rangle\quad (6.29)$$

になっている．

以上を踏まえて，確率変数 $Y(\omega)$ の関数であるところの確率変数

$$E[X|Y](\omega) := \phi(Y(\omega))\quad (6.30)$$

を改めて，条件 Y における X の条件付き期待値と定める．その土台となる関数形 $\phi(y) = E[X|y]$ は，(6.25) から求める．

算法 6.11 (条件付き期待値の性質 1) 定数 a, b, c と実数値関数 $f(x), g(x)$ に対して，

- (1) $E[af(Y) + bg(Y) + c|x] = aE[f(Y)|x] + bE[g(Y)|x] + c.$
- (2) $E[f(X)g(Y)|x] = f(x)E[g(Y)|x].$

[†] 例えば，身長・体重の一覧において，身長 170cm の人の平均体重 $E[\text{体重} | 170\text{cm}]$ など．

条件付き期待値 $E[X|Y]$ について,

$$(3) \quad E[E[X|Y]] := \int_{-\infty}^{\infty} E[X|y] dy = E[X]. \quad \therefore (6.28)$$

例題 6.4 (2) を示せ .

演習 6.6 同様に (1) を示せ .

6.5 ベクトルと行列の期待値

d 次元確率ベクトル $\mathbf{X} = (X_1, \dots, X_d)^T$ を考える . d 個の確率変数が共存するので , 平均操作は d 重積分となる .

$$\begin{aligned} \langle f(X_1, \dots, X_d) \rangle &:= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_d) p(x_1, \dots, x_d) dx_1 \cdots dx_d \end{aligned} \quad (6.31)$$

確率ベクトルの期待値 $\langle \mathbf{X} \rangle$ は , 各成分の平均操作 (6.31) を意味すると約束する .

$$\langle \mathbf{X} \rangle = \left\langle \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} \right\rangle := \begin{pmatrix} \langle X_1 \rangle \\ \vdots \\ \langle X_d \rangle \end{pmatrix} = \begin{pmatrix} m_1 \\ \vdots \\ m_d \end{pmatrix} \quad (6.32)$$

同様に , 行列の期待値 $\langle A \rangle$ も各成分の平均操作であるとする .

$$\langle A \rangle = \left\langle \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \right\rangle := \begin{pmatrix} \langle a_{11} \rangle & \cdots & \langle a_{1n} \rangle \\ \vdots & \cdots & \vdots \\ \langle a_{m1} \rangle & \cdots & \langle a_{mn} \rangle \end{pmatrix} \quad (6.33)$$

条件付き期待値は , 多次元の条件付き密度関数を用いて , 次のようにとる .

$$\begin{aligned} E[X_1|x_2, x_3] &= \int_{-\infty}^{\infty} x_1 p(x_1|x_2, x_3) dx_1 \\ &= \int_{-\infty}^{\infty} x_1 \frac{p(x_1, x_2, x_3)}{p(x_2, x_3)} dx_1 \end{aligned} \quad (6.34)$$

算法 6.12 (条件付き期待値の性質 2) ガウス確率変数 X_1, X_2, X_3 について , X_2, X_3 が独立ならば ,

$$E[X_1|X_2, X_3] = E[X_1|X_2] + E[X_1|X_3] - E[X_1] \quad (6.35)$$

例題 6.5 示せ . 算法 6.10 p50 と , その多次元版である算法 9.2 p74 による .

例題の解答例

例題 6.1 p47

$$\begin{aligned}
 (1) \text{ について, } V[X_i] &= \langle (X_i - m_i)^2 \rangle = \langle X_i^2 - 2m_i X_i + m_i^2 \rangle \\
 &= \langle X_i^2 \rangle - 2m_i \langle X_i \rangle + m_i^2 \langle 1 \rangle = \langle X_i^2 \rangle - m_i^2
 \end{aligned}$$

例題 6.2 p47

任意の t に対して, $S = \{(X_i - m_i)t + (X_j - m_j)\}^2 \geq 0$ は明らか. その平均も当然 $\langle S \rangle = \langle (X_i - m_i)^2 \rangle t^2 + 2 \langle (X_i - m_i)(X_j - m_j) \rangle t + \langle (X_j - m_j)^2 \rangle \geq 0$ となる. すなわち, $\sigma_i^2 t^2 + 2C_{ij}t + \sigma_j^2 \geq 0$ がいえる. これが任意の t で成立するのだから, 判別式は $C_{ij}^2 - \sigma_i^2 \sigma_j^2 \geq 0$ となる. ゆえに $(\rho_{ij})^2 \leq 1$.

例題 6.3 p47

$$\begin{aligned}
 &\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1) p(x_1, x_2) dx_1 dx_2 \\
 &= \int_{-\infty}^{\infty} f(x_1) \int_{-\infty}^{\infty} p(x_1, x_2) dx_2 dx_1 = \int_{-\infty}^{\infty} f(x_1) p_1(x_1) dx_1
 \end{aligned}$$

例題 6.4 p51

(2) について,

$$E[f(X)g(Y)|x] = \int_{-\infty}^{\infty} f(x)g(y)p(y|x)dy = f(x) \int_{-\infty}^{\infty} g(y)p(y|x)dy$$

例題 6.5 p51

まず右辺は, 算法 6.10 p50 の $E[X_1|X_2] = m_1 + \frac{C_{12}}{\sigma_2^2}(x_2 - m_2)$, $E[X_1|X_3] = m_1 + \frac{C_{13}}{\sigma_3^2}(x_3 - m_3)$ より, (右辺) $= m_1 + \frac{C_{12}}{\sigma_2^2}(x_2 - m_2) + \frac{C_{13}}{\sigma_3^2}(x_3 - m_3)$ となる.

次に左辺は, 算法 9.2 p74 で $\mathbf{Y} := X_1$, $\mathbf{X} := (X_2, X_3)^T$ とおく. X_2, X_3 はガウス性で独立と仮定したから $C_{23} = 0$ より,

$$A_0 = \Sigma_{yx} \Sigma_{xx}^{-1} = (C_{12}, C_{13}) \begin{pmatrix} \sigma_2^2 & C_{23} \\ C_{23} & \sigma_3^2 \end{pmatrix}^{-1} = \left(\frac{C_{12}}{\sigma_2^2}, \frac{C_{13}}{\sigma_3^2} \right)$$

となるが, これを $E[X_1|X_2, X_3] = E[\mathbf{Y}|\mathbf{X}] = \bar{\mathbf{y}} + A_0(\mathbf{X} - \bar{\mathbf{X}})$ に代入すると, 右辺と同じ式を得る.

7

確率ベクトルの変換

例えば, X, Y がガウス分布のとき, $X + Y$ などがどういう分布になるのか計算する. 前章までの算法の応用であり, よい計算練習になる. もちろん計算結果も大事で, 後続の章で使われる.

7.1 確率変数の加法とスカラ倍

X, Y を確率変数とする.

7.1.1 加法 $Z = X + Y$ の分布

X, Y の結合密度関数 $p_{XY}(x, y)$ を既知とする. Z の分布関数は,

$$F_Z(z) := P(\{Z \leq z\}) \quad (7.1)$$

であるが, 事象 $\{Z \leq z\} = \{X + Y \leq z\}$ の範囲より,

$$\begin{aligned} F_Z(z) &:= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{z-y} p_{XY}(x, y) dx \right\} dy \\ &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^z p_{XY}(s-y, y) ds \right\} dy \end{aligned} \quad (7.2)$$

となる (中括弧内を $x = s - y$ で置換した). これを z で微分すると,

$$p_Z(z) = \frac{dF_Z(z)}{dz} = \int_{-\infty}^{\infty} p_{XY}(z-y, y) dy \quad (7.3)$$

という公式を得る.

特に, X, Y が独立ならば, $p_{XY}(z-y, y) = p_X(z-y)p_Y(y)$ より,

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(z-y)p_Y(y) dy \quad (7.4)$$

となる. この形式の積分を, 畳み込み積分と呼んだ.

7.1.2 スカラ倍 $Z = aX$ の分布

X の密度関数 $p_X(x)$ を既知とする. まず, X と Z の分布関数の定義は,

$$F_X(x) = P(\{X \leq x\}), \quad F_Z(z) = P(\{Z \leq z\}) \quad (7.5)$$

であった． P は確率である．ここで，事象 $\{Z \leq z\} = \{aX \leq z\}$ は，事象 $\{X \leq a^{-1}z\}$ と同じなので，

$$F_Z(z) = P(\{Z \leq z\}) = P(\{X \leq a^{-1}z\}) = F_X(a^{-1}z) \quad (7.6)$$

が成立する．これを z で微分することで， Z の密度関数が判明する．

$$\begin{aligned} p_Z(z) &:= \frac{dF_Z(z)}{dz} = \frac{dF_X(a^{-1}z)}{dz} \\ &= \frac{da^{-1}z}{dz} \frac{dF_X(x)}{dx} \Big|_{x=a^{-1}z} = a^{-1} p_X(a^{-1}z) \end{aligned} \quad (7.7)$$

7.2 ガウス確率変数の加法とスカラ倍

X, Y をガウス確率変数とする．

7.2.1 加法 $Z = X + Y$ の分布

X, Y の 2 変数ガウス分布が既知とする．

$$\begin{aligned} p_{XY}(x, y) &:= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} \right. \\ &\quad \left. \times \left(\frac{(x-m_X)^2}{\sigma_X^2} - 2\rho \frac{(x-m_X)(y-m_Y)}{\sigma_X\sigma_Y} + \frac{(y-m_Y)^2}{\sigma_Y^2} \right) \right] \end{aligned} \quad (7.8)$$

(7.3) に代入すると，

$$\begin{aligned} p_Z(z) &= \int_{-\infty}^{\infty} p_{XY}(z-y, y) dy \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} \right. \\ &\quad \left. \times \left(\frac{(z-y-m_X)^2}{\sigma_X^2} - 2\rho \frac{(z-y-m_X)(y-m_Y)}{\sigma_X\sigma_Y} + \frac{(y-m_Y)^2}{\sigma_Y^2} \right) \right] dy \end{aligned} \quad (7.9)$$

積分を実行すると，

$$\begin{aligned} p_Z(z) &= \frac{1}{\sqrt{\sigma_X^2 + 2\rho\sigma_X\sigma_Y + \sigma_Y^2}\sqrt{2\pi}} \exp \left[-\frac{(z-m_X-m_Y)^2}{2(\sigma_X^2 + 2\rho\sigma_X\sigma_Y + \sigma_Y^2)} \right] \\ &= \mathcal{N}(z; m_X + m_Y, \sigma_X^2 + 2C_{XY} + \sigma_Y^2) \end{aligned} \quad (7.10)$$

を得る．すなわち， $Z = X + Y$ は，平均 $m_Z = m_X + m_Y = E[X + Y]$ ，分散 $\sigma_Z^2 = \sigma_X^2 + 2C_{XY} + \sigma_Y^2 = \langle (X - m_X + Y - m_Y)^2 \rangle = V[X + Y]$ のガウス分布に従う．

7.2.2 スカラ倍 $Z = aX$ の分布

X のガウス分布 $\mathcal{N}(x; m_x, \sigma_x^2)$ を既知とする．(7.7) より，

$$\begin{aligned} p_Z(z) &= a^{-1} p_X(a^{-1}z) = \frac{1}{a\sigma_x\sqrt{2\pi}} \exp\left\{-\frac{(a^{-1}z - m_x)^2}{2\sigma_x^2}\right\} \\ &= \frac{1}{a\sigma_x\sqrt{2\pi}} \exp\left\{-\frac{(z - am_x)^2}{2a^2\sigma_x^2}\right\} = \mathcal{N}(z; am_x, a^2\sigma_x^2) \quad (7.11) \end{aligned}$$

という公式を得る．すなわち， $Z = aX$ は，平均 $m_z = am_x = E[aX]$ ，分散 $\sigma_z^2 = a^2\sigma_x^2 = V[aX]$ のガウス確率変数となる．

7.2.3 線形結合 $Z = a_1X_1 + a_2X_2$ の分布

$m_i = E[X_i]$ ， $\sigma_i^2 := V[X_i]$ ， $C_{12} := \text{Cov}[X_1, X_2]$ とする．まず，(7.11) より， $Z_i = a_iX_i$ は平均 $m'_i = E[a_iX_i] = a_im_i$ ，分散 $\sigma'^2_i = V[a_iX_i] = a_i^2\sigma_i^2$ のガウス確率変数となる．さらに，(7.10) より， $Z = Z_1 + Z_2$ は，平均が

$$m_Z = E[Z_1 + Z_2] = E[a_1X_1 + a_2X_2] = a_1m_1 + a_2m_2 \quad (7.12)$$

で，分散が

$$\begin{aligned} \sigma_Z^2 &= V[Z_1 + Z_2] = V[a_1X_1 + a_2X_2] \\ &= a_1^2\sigma_1^2 + 2a_1a_2C_{12} + a_2^2\sigma_2^2 = \begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & C_{12} \\ C_{12} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad (7.13) \end{aligned}$$

のガウス確率変数となる．

以上を帰納的に用いると，次の結果が得られる．

算法 7.1 (ガウス確率変数の線形結合) X_1, X_2, \dots, X_n を既知のガウス確率変数とする．これらの線形結合

$$Z = a_1X_1 + a_2X_2 + \dots + a_nX_n \quad (7.14)$$

は，平均が

$$m_Z = E[Z] = a_1m_1 + a_2m_2 + \dots + a_nm_n \quad (7.15)$$

分散が

$$\begin{aligned} \sigma_Z^2 &= V[Z] \\ &= \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix} \begin{pmatrix} \sigma_1^2 & C_{12} & \dots & C_{1n} \\ C_{12} & \sigma_2^2 & \dots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \dots & \sigma_n^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad (7.16) \end{aligned}$$

のガウス確率変数となる。

7.3 共分散行列と多変数ガウス分布

ガウス確率変数 X_1, \dots, X_n を成分とする確率ベクトル $\mathbf{X} = (X_1, \dots, X_n)^T$ を、ガウス確率ベクトルという。

7.3.1 共分散行列

算法 7.1 に現れた行列を、分散共分散行列または略して共分散行列と呼び、次のように表記する。

$$\text{Cov}[\mathbf{X}] \quad \text{または} \quad \Sigma_{xx} := \begin{pmatrix} \sigma_1^2 & C_{12} & \cdots & C_{1n} \\ C_{12} & \sigma_2^2 & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & \sigma_n^2 \end{pmatrix} \quad (7.17)$$

共分散行列は、確率ベクトルの全成分の全ての組み合わせの相関を表している。対角成分には分散 $\sigma_i^2 = \text{V}[X_i]$ が並び、それ以外には共分散 $C_{ij} := \text{Cov}[X_i, X_j]$ が並び、

ここで、ベクトルから行列を作るテンソル積と呼ばれる算法

$$\begin{aligned} \mathbf{X}\mathbf{X}^T &= \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} (X_1, X_2, \dots, X_n) \\ &:= \begin{pmatrix} X_1X_1 & X_1X_2 & \cdots & X_1X_n \\ X_2X_1 & X_2X_2 & \cdots & X_2X_n \\ \vdots & \vdots & \ddots & \vdots \\ X_nX_1 & X_nX_2 & \cdots & X_nX_n \end{pmatrix} \end{aligned} \quad (7.18)$$

を導入すると、共分散行列は、

$$\text{Cov}[\mathbf{X}] = \langle (\mathbf{X} - \bar{\mathbf{x}})(\mathbf{X} - \bar{\mathbf{x}})^T \rangle, \quad \bar{\mathbf{x}} := \langle \mathbf{X} \rangle \quad (7.19)$$

と表記できる。共分散の対称性 $C_{ij} = C_{ji}$ より、 $\text{Cov}[\mathbf{X}]$ は対称行列となる。すなわち、 $\text{Cov}[\mathbf{X}] = \text{Cov}[\mathbf{X}]^T$ 。

その他、2つの確率ベクトル \mathbf{X}, \mathbf{Y} の相関を定量化するときには、

$$\text{Cov}[\mathbf{X}, \mathbf{Y}] \quad \text{または} \quad \Sigma_{xy} := \langle (\mathbf{X} - \bar{\mathbf{x}})(\mathbf{Y} - \bar{\mathbf{y}})^T \rangle,$$

$$\bar{x} := \langle X \rangle, \bar{y} := \langle y \rangle \quad (7.20)$$

という行列も用いる．これを，相互共分散行列という． $\text{Cov}[X, Y]$ は，一般には，対称行列とはならない．なお，以上の表記法においては， $\text{Cov}[X, X] = \text{Cov}[X]$ となることに注意せよ．

例題 7.1 $n = 3$ のときの $\text{Cov}[X]$ の成分を書き下せ．

演習 7.1 同じく， $\text{Cov}[X, Y]$ の成分を書き下せ．

7.3.2 多変数ガウス分布

ガウス確率ベクトル X の共分散行列を Σ_{xx} とする．このとき， X の分布は，

$$\begin{aligned} p(x) &= \mathcal{N}(x; \bar{x}, \Sigma_{xx}) \\ &:= \frac{1}{(\sqrt{2\pi})^n \sqrt{|\Sigma_{xx}|}} \exp \left\{ -\frac{1}{2} (x - \bar{x})^T \Sigma_{xx}^{-1} (x - \bar{x}) \right\} \end{aligned} \quad (7.21)$$

に従う．これを n 変数ガウス分布という．

算法 7.2 (ガウス確率ベクトルの構成要件)

1. ガウス確率ベクトル X の共分散行列 Σ_{xx} は，半正定値行列になる．
2. 逆に，(7.21) を勝手な行列 $\Sigma_{xx} = A$ で与えたとき， A が半正定値行列ならば，対応する X は必ずガウス確率ベクトルとなる（全成分がガウス分布となる）．

▶▶ (半正定値行列) 任意のベクトル x に対して， $x^T A x \geq 0$ となる行列 A を，半正定値行列という．この条件を破ると対角成分が負になる状況が起り，分散が非負となる確率論の計算を逸脱する．半正定値行列 A には， $A = C^T C$ となる行列 C が存在する．

7.4 ガウス確率ベクトルの線形変換

X をガウス確率ベクトルとする．定数行列 A による X の線形変換 $Y = AX$ を考える． $\langle X \rangle = \bar{x}$ と Σ_{xx} を既知とする．このとき，両辺の期待値をとると，

$$\bar{y} = A\bar{x} \quad (7.22)$$

が得られる．また， Y の分散共分散行列は，

$$\Sigma_{yy} := \langle (Y - \bar{y})(Y - \bar{y})^T \rangle = \langle (AX - A\bar{x})(AX - A\bar{x})^T \rangle$$

$$= A \langle (X - \bar{x})(X - \bar{x})^T \rangle A^T = A \Sigma_{xx} A^T \quad (7.23)$$

となる． Σ_{xx} の半正定値性から， Σ_{yy} も半正定値となる．

▶▶ ガウス確率ベクトルの共分散行列 Σ_{xx} は半正定値だから， $\Sigma_{xx} = B^T B$ となる行列 B が存在する．このとき， $\Sigma_{yy} = AB^T B A^T = (BA^T)^T (BA^T)$ となる行列 $C = BA^T$ が存在するので， Σ_{yy} も半正定値である．

ゆえに， Y は Σ_{yy} を共分散行列とするガウス確率ベクトルであり，その分布は，

$$\begin{aligned} p_Y(y) &= \mathcal{N}(y; \bar{y}, \Sigma_{yy}) \\ &= \frac{1}{(\sqrt{2\pi})^n \sqrt{|\Sigma_{yy}|}} \exp \left\{ -\frac{1}{2} (y - \bar{y})^T \Sigma_{yy}^{-1} (y - \bar{y}) \right\} \end{aligned} \quad (7.24)$$

となる．

算法 7.3 (ガウス確率ベクトルの線形変換) ガウス確率ベクトル X の線形変換 $Y = AX$ は，ガウス確率ベクトルとなる．その平均は $\bar{y} = A\bar{x}$ ，共分散行列は $\Sigma_{yy} = A\Sigma_{xx}A^T$ となる．

7.4.1 共分散行列の対角化による「無相関」化

さて，線形代数によれば， Σ_{xx} の固有ベクトルを列ベクトルとする対角化行列 $S = (u_1, u_2, u_3)$ を使うと，

$$\Sigma'_{xx} := S^{-1} \Sigma_{xx} S = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \quad (7.25)$$

となって， Σ_{xx} を対角化した Σ'_{xx} が得られる． λ_i は固有値である．

また，共分散行列 Σ_{xx} は対称行列であるから，その固有ベクトル u_1, u_2, u_3 は互いに直交する．ゆえに，対角化行列 $S = (u_1, u_2, u_3)$ は，直交変換の性質を示す．

$$S^T = S^{-1} \quad (7.26)$$

ただし，そうなるために u_i の長さを $\|u_i\| = 1$ ($i = 1, 2, 3$) に調整しておく．

以上の対角化行列 S を使って，確率ベクトル X を Y に変換すると，

$$Y := S^{-1} X = S^T X \quad (7.27)$$

算法 7.3 で $A = S^{-1}$ とおくと， Y の共分散行列は， $S^{-1} = S^T$ より，

$$\Sigma_{yy} = A \Sigma_{xx} A^T = S^T \Sigma_{xx} (S^T)^T = S^{-1} \Sigma_{xx} S = \Sigma'_{xx} \quad (7.28)$$

となって対角化される．すなわち， Σ_{yy} の非対角成分は全て 0 となる．

その効果として，変換後の Y の成分 Y_1, Y_2, Y_3 は互いに無相関

$$\text{Cov}[Y_1, Y_2] = \text{Cov}[Y_2, Y_3] = \text{Cov}[Y_1, Y_3] = 0 \quad (7.29)$$

となる．

このように，無相関という性質は，確率変数 X_1, X_2, X_3 の取り方に依存している．すなわち，元の $X = (X_1, X_2, X_3)^T$ に相関があっても，対角化行列 S による変換 $Y = S^{-1}X$ によって，無相関な組 Y_1, Y_2, Y_3 を作ることができる．

▶▶ (作為的な相関?) 逆にいうと，等高線が楕円になるような密度関数の分布であれば，確率変数の座標変換 (回転) により，相関無しに見せたり，相関有りに見せたりできる？

7.4.2 2 変数ガウス分布の例

図 5.2 の 2 変数ガウス分布 $p(x_1, x_2)$ の共分散行列は，

$$\Sigma_{xx} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix} \quad (7.30)$$

である．固有値は 0.2, 1.8 となり，規格化した固有ベクトルより，

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \quad (7.31)$$

が得られる．このとき，新しい確率ベクトル

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = S^T \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad (7.32)$$

の Y_1 と Y_2 は無相関となる．なぜなら，

$$\Sigma_{yy} = S^T \Sigma_{xx} S = \begin{pmatrix} 0.2 & 0 \\ 0 & 1.8 \end{pmatrix} \quad (7.33)$$

より $\text{Cov}[Y_1, Y_2] = 0$ である．

元の $p(x_1, x_2) = \mathcal{N}(x; \bar{x}, \Sigma_{xx})$ と，対角化後の $p'(y_1, y_2) = \mathcal{N}(y; \bar{y}, \Sigma_{yy})$ の比較を 図 7.1 に示す．元の (a) には x_1, x_2 の比例関係が見てとれる．これに対して，対角化後の (b) では， y_1, y_2 の比例関係が消失するよう密度関数が変換されている．

以上の対角化を応用したものに，統計学の因子分析や主成分分析などがある．

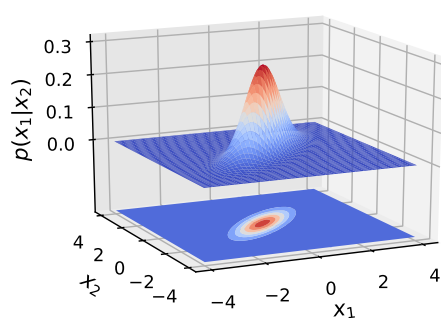
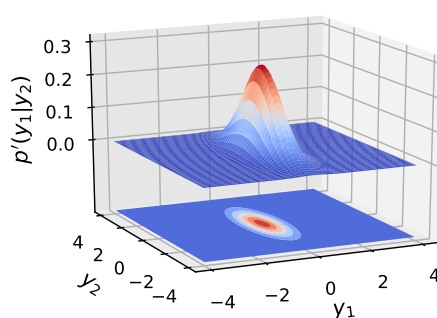
(a) 結合密度関数 $p(x_1, x_2)$ (b) 対角化後のもの $p'(y_1, y_2)$

図 7.1 2 変数ガウス分布と対角化後のガウス分布 ($m_1 = m_2 = 0, \sigma_1 = \sigma_2 = 1, \rho = 0.8$)

例題の解答例

例題 7.1 p57

$$\begin{aligned}
\text{Cov}[\mathbf{X}] &:= \left\langle (\mathbf{X} - \bar{\mathbf{x}})(\mathbf{X} - \bar{\mathbf{x}})^T \right\rangle \\
&= \left\langle \begin{pmatrix} X_1 - \bar{x}_1 \\ X_2 - \bar{x}_2 \\ X_3 - \bar{x}_3 \end{pmatrix} (X_1 - \bar{x}_1, X_2 - \bar{x}_2, X_3 - \bar{x}_3) \right\rangle \\
&= \left\langle \begin{pmatrix} (X_1 - \bar{x}_1)^2 & (X_1 - \bar{x}_1)(X_2 - \bar{x}_2) & (X_1 - \bar{x}_1)(X_3 - \bar{x}_3) \\ (X_2 - \bar{x}_2)(X_1 - \bar{x}_1) & (X_2 - \bar{x}_2)^2 & (X_2 - \bar{x}_2)(X_3 - \bar{x}_3) \\ (X_3 - \bar{x}_3)(X_1 - \bar{x}_1) & (X_3 - \bar{x}_3)(X_2 - \bar{x}_2) & (X_3 - \bar{x}_3)^2 \end{pmatrix} \right\rangle \\
&= \begin{pmatrix} \langle (X_1 - \bar{x}_1)^2 \rangle & \langle (X_1 - \bar{x}_1)(X_2 - \bar{x}_2) \rangle & \langle (X_1 - \bar{x}_1)(X_3 - \bar{x}_3) \rangle \\ \langle (X_2 - \bar{x}_2)(X_1 - \bar{x}_1) \rangle & \langle (X_2 - \bar{x}_2)^2 \rangle & \langle (X_2 - \bar{x}_2)(X_3 - \bar{x}_3) \rangle \\ \langle (X_3 - \bar{x}_3)(X_1 - \bar{x}_1) \rangle & \langle (X_3 - \bar{x}_3)(X_2 - \bar{x}_2) \rangle & \langle (X_3 - \bar{x}_3)^2 \rangle \end{pmatrix}
\end{aligned}$$

8.1 最小二乗法

確率ベクトルの $X = (X, Y)^T$ の標本 Π_n が、図 8.1 のように得られたとする。

$$\Pi_n = \left\{ \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \dots, \begin{pmatrix} x_n \\ y_n \end{pmatrix} \right\} \quad (8.1)$$

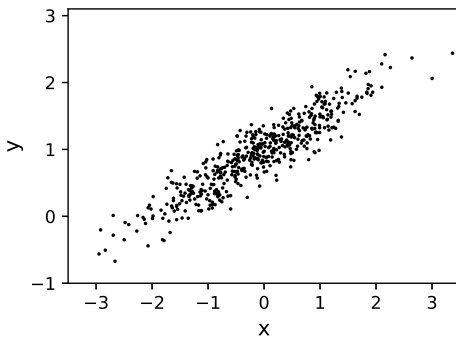


図 8.1 2次元の試行列

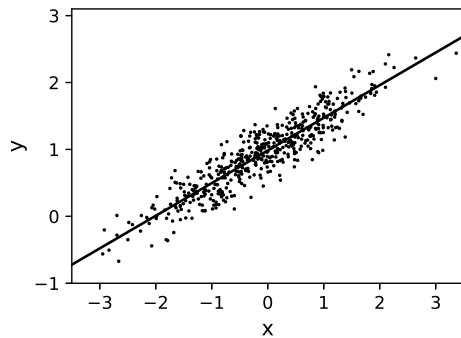


図 8.2 回帰直線

8.1.1 推定問題

この試行列の傾向を学習することで、新たに x_i が観測されたときの y_i を推定する問題を考える。

そのために、 X と Y の実測値 x_i, y_i の関係について、次のようなモデルを導入する。

$$\hat{y}_i := f(x_i) \quad (8.2)$$

$f(x)$ をモデル関数、 \hat{y}_i を推定値という。また、 Y の測定値を

$$y_i := f(x_i) + e_i \quad (8.3)$$

でモデル化する． e_i は誤差である．

図 8.1 の試行列には，明らかな比例関係が見てとれるので，

$$f(x) = a_0 + a_1 x \quad (8.4)$$

を採用する．これを線形回帰モデルという．そのグラフを回帰直線という．

8.1.2 最小二乗法

誤差 e_i の 2 乗の和をとる．

$$J = \sum_{i=1}^n (e_i)^2 = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (8.5)$$

これを残差二乗和という．二乗和を平方和ともいう．

J を最小にするパラメータ a_0, a_1 を見付けることを，最小二乗法という． J の停留値を求めるため， a_0, a_1 による J の偏微分を 0 におくと，

$$\begin{cases} \frac{\partial J}{\partial a_0} = -2 \sum_{i=1}^n \frac{\partial f}{\partial a_0} (y_i - f(x_i)) = -2 \sum_{i=1}^n (y_i - a_0 - a_1 x_i) = 0 \\ \frac{\partial J}{\partial a_1} = -2 \sum_{i=1}^n \frac{\partial f}{\partial a_1} (y_i - f(x_i)) = -2 \sum_{i=1}^n x_i (y_i - a_0 - a_1 x_i) = 0 \end{cases} \quad (8.6)$$

両辺を $-2n$ で割ると， a_0, a_1 に関する連立方程式が得られる．

$$\begin{cases} 0 = \frac{1}{n} \sum_{i=1}^n (y_i - a_0 - a_1 x_i) = \langle Y \rangle - a_0 - a_1 \langle X \rangle \\ 0 = \frac{1}{n} \sum_{i=1}^n x_i (y_i - a_0 - a_1 x_i) = \langle XY \rangle - a_0 \langle X \rangle - a_1 \langle X^2 \rangle \end{cases} \quad (8.7)$$

$\langle \cdot \rangle$ はここでは標本平均を表す．これらを解くと，

$$\begin{aligned} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} &= \begin{pmatrix} 1 & m_X \\ m_X & \langle X^2 \rangle \end{pmatrix}^{-1} \begin{pmatrix} m_Y \\ \langle XY \rangle \end{pmatrix} \\ &= \frac{1}{\langle X^2 \rangle - m_X^2} \begin{pmatrix} \langle X^2 \rangle & -m_X \\ -m_X & 1 \end{pmatrix} \begin{pmatrix} m_Y \\ \langle XY \rangle \end{pmatrix} \\ &= \frac{1}{\sigma_X^2} \begin{pmatrix} \sigma_X^2 m_Y - C_{XY} m_X \\ C_{XY} \end{pmatrix} = \begin{pmatrix} m_Y - \rho_{XY} \frac{\sigma_Y}{\sigma_X} m_X \\ \rho_{XY} \frac{\sigma_Y}{\sigma_X} \end{pmatrix} \end{aligned} \quad (8.8)$$

となる．表 6.1 の記号を用いた．ゆえに，推定値は，

$$\hat{y}_i = f(x_i) = \rho_{XY} \frac{\sigma_Y}{\sigma_X} (x_i - m_X) + m_Y \quad (8.9)$$

となる[†]．この \hat{y}_i を y_i の最小二乗推定値という．図 8.2 は Π_n 上の標本平均 $m_X, m_Y, \sigma_X, \sigma_Y, \rho_{XY}$ から求めた回帰直線である．

[†] 久米⁵⁾ p87 など．

8.2 確率変数の線形代数

8.2.1 結合モーメントと内積

数ベクトル $x = (x_1, \dots, x_n)^T$, $y = (y_1, \dots, y_n)^T$ の内積は,

$$\langle x, y \rangle = x_1 y_1 + \dots + x_n y_n = x^T y \quad (8.10)$$

のように計算した. 一方で, 内積は,

$$\langle x, y \rangle = \|x\| \|y\| \cos \theta \quad (8.11)$$

でも計算できた. $\|\cdot\|$ はノルムでベクトルの長さを表す. θ は x, y のなす角である. それでは, どちらが正真正銘の内積なのだろうか?

現代数学の発想は少し違う. 計算の理論であるから, 紙と鉛筆でどう計算できるかしが気にしない. それが現実の何を表そうとも, 紙の上の計算が同じなら現代数学にとっては同じものである. この観点で, 内積に特有の計算ルールが抽出された.

定義 8.1 (内積の公理) 次の法則で式変形できる積 $\langle x, y \rangle$ を内積という.

- (0) 実数値: $\langle x, y \rangle$ は常に実数値^{†2}.
- (1) 対称性: $\langle x, y \rangle = \langle y, x \rangle$
- (2) 分配則: $\langle ax, y \rangle = a \langle x, y \rangle$, $\langle x, ay \rangle = a \langle x, y \rangle$, $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$,
 $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$.
- (3) 正定値性: 自分自身との内積は常に $\langle x, x \rangle \geq 0$. 特に $\langle x, x \rangle = 0$ となるのは $x = 0$ のときだけ.

このルール (公理) に従う限り, どのように式変形しても内積 $\langle x, y \rangle$ は内積であり続ける. 例えば, 力学的仕事は内積 $\langle x, F \rangle$ で表せるが, このルールを逸脱しない限り, どのように式変形しても $\langle x, F \rangle$ は力学的仕事であり続ける. 逆に, ルールにない式変形を行ってしまうと, 計算結果にある表式 $\langle x, F \rangle$ が力学的仕事を表す保証はない.

ここで, 確率変数 X, Y の 2 次結合モーメント

$$\langle XY \rangle := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy p(x, y) dx dy \quad (8.12)$$

の式変形は, 実は, 定義 8.1 の内積の公理に該当してしまう. 実際,

- (0) 実数値: $\langle XY \rangle$ は常に実数値.
- (1) 対称性: $\langle XY \rangle = \langle YX \rangle$

^{†2} ここでは実数値バージョンの内積を示している. 複素数値バージョンでは若干ルールが変わる.

(2) 分配則: $\langle (aX)Y \rangle = a \langle XY \rangle$, $\langle X(aY) \rangle = a \langle XY \rangle$, $\langle (X+Y)Z \rangle = \langle XZ \rangle + \langle YZ \rangle$,
 $\langle X(Y+Z) \rangle = \langle XY \rangle + \langle XZ \rangle$.

(3) 正定値性: 自分自身との結合モーメントは常に $\langle X^2 \rangle \geq 0$.

というところまで普通に該当してしまう。ただし、最後の

- 特に $\langle X^2 \rangle = 0$ となるのは $X = 0$ のときだけ。

のところは、ゼロ 0 が何だか分らない。これは次節で述べるように、

- 任意の見本 ω について常に $0(\omega) = 0$ となる確率変数をゼロ 0 とする。
 と定義すると辻褄が合う。以上、 $\langle XY \rangle$ の計算は、内積 $\langle X, Y \rangle$ と同じである。

8.2.2 確率変数のベクトル化

内積と同様に、現代数学では「ベクトル」を計算ルールとして捉える。

定義 8.2 (ベクトルの公理) 次のように式変形できる x, y, \dots をベクトルという。

(L1) $x + y = y + x$.

(L2) $(x + y) + z = x + (y + z)$.

(L3) どんな x に足しても、 $x + \mathbf{0} = x$ となる $\mathbf{0}$ が使える。(ゼロベクトル)

(L4) どんな x にも、相方 $(-x)$: $x + (-x) = \mathbf{0}$ がいる。(逆ベクトル)

(L5) $\lambda(\mu x) = (\lambda\mu)x$.

(L6) スカラ 1 の作用は、 $1x = x$.

(L7) $(\lambda + \mu)x = \lambda x + \mu x$.

(L8) $\lambda(x + y) = \lambda x + \lambda y$.

ベクトルの全体集合 $V := \{x, y, \dots\}$ をベクトル空間という。

▶▶ 定義 8.2 は、ベクトルを太字のまま筆算するための式変形ルールである。この公式で計算している限り、成分によるチェックはする必要がない。

この計算ルールを満足するものは、物理的実体が何であろうとも、紙の上の計算では区別できない(記号の違いは無視するとして)。それをベクトルと呼ぶわけだ。

具体的対象をベクトル化するには、5つの操作項目を定めればよい。表 8.1 は確率変数 X をベクトル化するための操作項目である。測定器の比喻で述べている。確率変数 X とは、見本 ω を入力すると測定値 $x = X(\omega)$ を返す関数

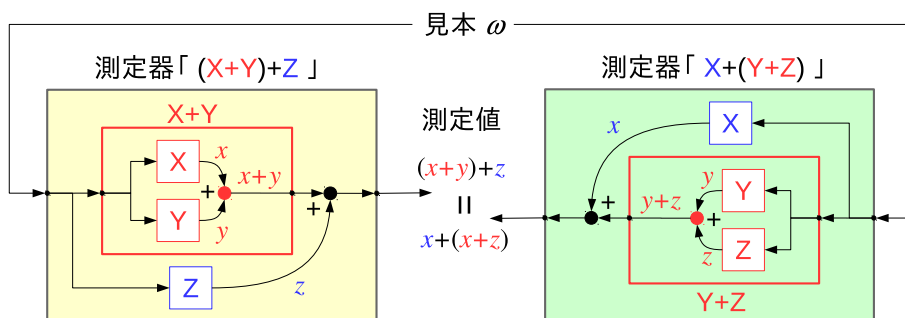
$$\text{見本 } \omega \longrightarrow \boxed{X} \longrightarrow x \text{ 測定値}$$

だった。これに該当する物理的存在とは測定器であろう。

詳細は省くが、表 8.1 の操作は、定義 8.2 の 8 つのルールを全て満足する。一例としてルー

表 8.1 確率変数 $X(\cdot)$ のベクトル化 (測定器で比喻)

操作項目	文字表記	対応する測定器 $x = X(\omega), y = Y(\omega)$
(a) 等号	$X = Y$	どんな見本 ω を代入しても測定値が一致 $X(\omega) = x = y = Y(\omega)$ すること
(b) 加法	$X + Y$	$x + y$ を返す新たな測定器 “ $X + Y$ ”(.)
(c) スカラ倍	λX	λx を返す新たな測定器 “ λX ”(.)
(d) 零ベクトル	O	常に 0 を返す測定器 “ O ”(.)
(e) 逆ベクトル	$-X$	$-x$ を返す新たな測定器 “ $-X$ ”(.)

図 8.3 確率変数における (L2) $(X + Y) + Z = X + (Y + Z)$

ル (L2) の成立を図 8.3 に示す．測定器 “ $(X + Y) + Z$ ” と測定器 “ $X + (Y + Z)$ ” は，どんな見本 ω についても測定値が一致する．

演習 8.1 定義 8.2 の他のルールについても図式化せよ．

以上，表 8.1 の操作を付与された確率変数 X, Y, \dots はベクトルと化す．

8.2.3 確率変数の線形代数

というわけで，通常のベクトル x, y, \dots とのその内積 $\langle x, y \rangle$ において成立することは，確率変数 X, Y, \dots とその 2 次結合モーメント $\langle XY \rangle$ の世界でもそのまま成立する (逆は一般には成立しない)．その結果，次の手順が可能になる．

- まず，ベクトル x, y, \dots とその内積 $\langle x, y \rangle$ の世界で，矢印の作図などを通してアイデアを練る．
- 得られた結果の x, y, \dots を X, Y, \dots に置き換え， $\langle x, y \rangle$ を $\langle XY \rangle$ に置き換えると，それは確率変数の結果になっている．

言い方を変えると，確率変数 X, Y, \dots は，矢印ベクトルだと思って作図して考えてよい．そうしても，得られた結果に間違いは起らない．

とはいえ， $\langle X, Y \rangle$ や $\langle X^2 \rangle$ が発散するような確率変数 X, Y をとると，議論が破綻してしまう．これを防ぐには，制約条件 $\langle X^2 \rangle < \infty$ を課せばよい[†]．このような制約条件が課され

[†] この条件を，2 乗可積分という．

たベクトルの集合を，ヒルベルト空間という．

以上の内容をまとめると，次のような設定が得られる．

定義 8.3 (確率変数の内積) $X, Y \in \mathcal{H}$ の 2 次結合モーメントを，必要に応じて，

$$\langle X, Y \rangle := \langle XY \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy p(x, y) dx dy \quad (8.13)$$

と書く． $\langle X, Y \rangle$ は内積となり，定義 8.1 で計算できる．

定義 8.4 (ヒルベルト空間) 2 次モーメントが有限な確率変数の集合

$$\left\{ X(\cdot) \mid \langle X^2 \rangle < \infty \right\} \quad (8.14)$$

に表 8.1 の 5 項目と，定義 8.3 の内積を付与したものを \mathcal{H} と書き，確率変数のヒルベルト空間と呼ぶ．その要素 $X, Y, \dots \in \mathcal{H}$ は，定義 8.2 のベクトル演算と，定義 8.1 に従う内積 $\langle X, Y \rangle$ で計算できる．

定義 8.5 (確率変数の直交性) 無相関 $\langle X, Y \rangle = \langle XY \rangle = 0$ であることを， $X \perp Y$ と書き， X と Y は直交するという．

定義 8.6 (確率変数のノルムと距離) 2 次モーメントの正の平方根

$$\|X\| := \sqrt{\langle X, X \rangle} = \sqrt{\langle X^2 \rangle} \quad (8.15)$$

を X のノルムという．特に $\langle X \rangle = 0$ ときは $\|X\| = \sigma_X$ (標準偏差) となる．特に $d(X, Y) := \|X - Y\|$ を確率変数 X, Y の距離と呼び，確率変数どうしの近さを表す．

矢印ベクトル x のノルム $\|x\|$ は矢印の長さを表すが，確率変数のノルム $\|X\|$ は X のばらつきの大きさを表している．したがって， X, Y の距離 $\|X - Y\|$ は， X と Y の違いを， $X - Y$ のばらつきの大きさを測ったものである．

8.3 ベクトルの近似

近似とは，大雑把にいうと，不完全なモデル (模型) で本物を真似することである．

簡単のため $V = \mathbb{R}^3$ の場合を考える．まず，近似対象のベクトル $x = (1, 2, 1)^T \in V$ を 1 つ取って固定する (本物)．そのモデルを m としよう． V のなかで最良のモデルを探すと本物 x が該当してしまうので， m が動く空間 M は，一般に V の部分集合 $M \subset V$ とされる．さらに， M 内のベクトル演算は M からはみ出さないとする．

- $m_1, m_2 \in M \implies m_1 + m_2 \in M$
- $m \in M \implies ax \in M$

こうした M を, V の部分空間という. 以下, V の部分空間 M をモデル空間と呼ぶことにする. ここでは, 例えば, 次のようにしてみよう.

$$M := \left\{ a_0 \mathbf{u}_1 + a_1 \mathbf{u}_2 \mid a_0, a_1 \in \mathbb{R}, \mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \in \mathbb{R}^3 \right\} \quad (8.16)$$

ここで, 誤差ベクトル

$$\mathbf{e} := \mathbf{x} - \mathbf{m} \quad (8.17)$$

を導入すると, 誤差の大きさは $\|\mathbf{e}\| = \|\mathbf{x} - \mathbf{m}\|$ で評価できる.

以上の設定において, 誤差 $\|\mathbf{x} - \mathbf{m}\|$ を最小化するモデル $\mathbf{m} = \mathbf{m}_0$ を求めたい. \mathbf{m}_0 を最小化ベクトルという.

\mathbf{x} や \mathbf{m} は 3 次元ベクトルなので, 図 8.4 のような作図で考えることができる. 制約のない

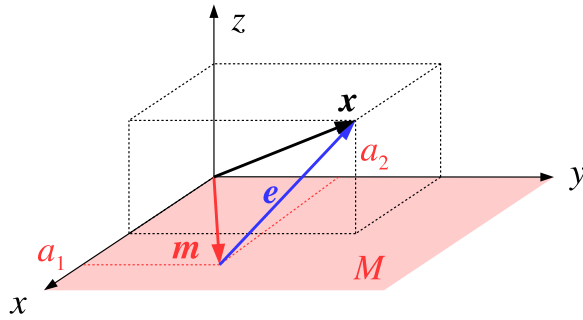


図 8.4 ベクトルの近似

\mathbf{x} に対して, \mathbf{m} は xy 平面 (モデル空間 M) 内に制限されている. \mathbf{m} から \mathbf{x} に向うベクトルの長さ $\|\mathbf{e}\|$ が誤差の大きさを表す.

図より, 誤差 $\|\mathbf{e}\|$ を最小化するベクトル $\mathbf{m} = \mathbf{m}_0$ は, M の法線方向 (z 軸) から照射された平行光が M に落す, \mathbf{x} の影に一致することが分かる. これを

$$\mathbf{m}_0 = \text{proj}(\mathbf{x} | M) \quad (8.18)$$

と書き, 右辺を, \mathbf{x} の M 上への直交射影と呼ぶ. その作り方から, $(\mathbf{x} - \mathbf{m}_0) \perp M$ が成立する. ただし,

$$(\mathbf{x} - \mathbf{m}_0) \perp M \stackrel{\text{定義}}{\iff} \langle (\mathbf{x} - \mathbf{m}_0), \mathbf{m} \rangle = 0 \quad \text{for } \forall \mathbf{m} \in M \quad (8.19)$$

である[†]．すなわち， $x - m_0$ は M 上の任意方向のベクトルと直交する．条件 (8.19) の成否を確かめるには，無限個の $\forall m \in M$ についてチェックする必要がある．それは無理なので，次の算法が編み出された．有限回のチェックで (8.19) の成否が判定できる．

算法 8.1 u_1, \dots, u_n によって張られる部分空間 M に対して，

$$(x - m_0) \perp M \xLeftrightarrow[\text{必要}]{\text{十分}} \langle (x - m_0), u_i \rangle = 0, i = 1, \dots, n \quad (8.20)$$

► 証明 まず， \Rightarrow を示すため (8.19) を仮定すると， $u_i \in M$ より $\langle (x - m_0), u_i \rangle = 0$ が成立する．次に， \Leftarrow を示すため $\langle (x - m_0), u_i \rangle = 0$ を仮定する．任意の m は適当な係数 c_i で， $m = c_1 u_1 + \dots + c_n u_n$ と書ける．このとき，

$$\langle (x - m_0), m \rangle = c_1 \langle (x - m_0), u_1 \rangle + \dots + c_n \langle (x - m_0), u_n \rangle = 0$$

が c_i によらず成立するが，これは (8.19) の成立を意味する．

モデル空間 (8.16) を張るベクトル u_1, u_2 の線形結合で，最小化ベクトルの候補を，

$$m_0 = a_0 u_1 + a_1 u_2 \quad (8.21)$$

と表す．算法 8.1 を適用すると，

$$0 = \langle x - m_0, u_1 \rangle = \left\langle \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} - \begin{pmatrix} a_0 \\ a_1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle = 1 - a_0 \quad (8.22)$$

$$0 = \langle x - m_0, u_2 \rangle = \left\langle \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} - \begin{pmatrix} a_0 \\ a_1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle = 2 - a_1 \quad (8.23)$$

より $a_0 = 1, a_1 = 2$ が判明し，次の最小化ベクトルが得られる．

$$m_0 = a_0 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + a_1 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

8.4 確率変数の推定

前節の算法を，そっくりそのまま確率変数に適用すると，確率変数の推定問題が定式化できる．8.1 節の最小二乗法では標本 Π_n 上で議論を展開したが，ここでは同じことを確率空間 (Ω, \mathcal{F}, P) 上で行う．

[†] “ \forall ” の使い方は，A.1.6 節を見よ．

8.4.1 推定問題

あるランダム現象の測定量を確率変数 X, Y で表す．これらをヒルベルト空間上のベクトル $X, Y \in \mathcal{H}$ とみなす．

\mathcal{H} は一般に無限次元なので (独立なベクトルを無限個とれる)，その要素は有限個のパラメータでは表せない．そこで，有限次元の部分空間 $\mathcal{M} \subset \mathcal{H}$ をとり， $Y \in \mathcal{H}$ を $\hat{Y} \in \mathcal{M}$ で近似することを考える．この $\hat{Y} \in \mathcal{M}$ を Y の推定値という． \mathcal{M} は有限次元だから，その要素である \hat{Y} は有限個のパラメータで表せる．

推定誤差を評価するため，誤差ベクトル $e := Y - \hat{Y}$ のノルム $\|e\|$ を誤差の大きさとする．これを最小化する \hat{Y} について，8.3 節と全く同じ算法が使える．記号を取り替え，若干集約して再掲すると，

算法 8.2 (確率変数の推定) \mathcal{M} を確率変数 U_1, \dots, U_n で張られた部分空間とする．誤差 $\|Y - \hat{Y}\|$ を最小化する推定値 $\hat{Y} \in \mathcal{M}$ は， $Y \in \mathcal{H}$ の \mathcal{M} 上への直交射影

$$\hat{Y} = \text{proj}(Y|\mathcal{M}) \quad (8.24)$$

となる．そのための必要十分条件は， $(Y - \hat{Y}) \perp \mathcal{M}$ ，すなわち

$$\langle (Y - \hat{Y}), U_i \rangle = 0 \quad (i = 1, \dots, n) \quad (8.25)$$

で与えられる．

8.4.2 線形回帰の例題

線形回帰型のモデル空間

$$\mathcal{M} := \left\{ a_0 + a_1 X \mid a_0, a_1 \in \mathbb{R}; X \in \mathcal{H} \right\} \quad (8.26)$$

をとる．定数項を，定値関数 (どんな見本 ω の測定値も定数 a_0 となる確率変数)

$$a_0(\omega) := a_0 \quad \forall \omega \in \Omega \quad (8.27)$$

とみなすと， \mathcal{M} は確率変数 $a_0, X \in \mathcal{H}$ で張られた部分空間になっている．

算法 8.2 より，本物 $Y \in \mathcal{H}$ との誤差 $\|Y - \hat{Y}\|$ を最小にする推定値 $\hat{Y} \in \mathcal{M}$ は，

$$\hat{Y} = \text{proj}(Y|\mathcal{M}) \subset \mathcal{M} \quad (8.28)$$

であり，そのための条件は，同じく算法 8.2 より，

$$\begin{cases} 0 = \langle (Y - \hat{Y}), a_0 \rangle = \langle (Y - \hat{Y})a_0 \rangle = a_0 \langle Y - \hat{Y} \rangle \\ \quad \therefore 0 = \langle Y - \hat{Y} \rangle \\ 0 = \langle (Y - \hat{Y}), X \rangle = \langle (Y - \hat{Y})X \rangle \end{cases} \quad (8.29)$$

である． $\hat{Y} \in \mathcal{M}$ の候補を， $\hat{Y} = a_0 + a_1 X$ において代入すると，

$$\begin{cases} 0 = \langle Y - \hat{Y} \rangle = \langle Y - a_0 - a_1 X \rangle = \langle Y \rangle - a_0 - a_1 \langle X \rangle \\ 0 = \langle (Y - \hat{Y})X \rangle = \langle (Y - a_0 - a_1 X)X \rangle \\ \quad = \langle XY \rangle - a_0 \langle X \rangle - a_1 \langle X^2 \rangle \end{cases} \quad (8.30)$$

となり，(8.7) と同じ式が得られる．ゆえに，(8.8) と同じパラメータを得る．

$$\begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} m_Y - \rho_{XY} \frac{\sigma_Y}{\sigma_X} m_X \\ \rho_{XY} \frac{\sigma_Y}{\sigma_X} \end{pmatrix} \quad (8.31)$$

以上より，推定値は，

$$\hat{Y} = \hat{Y}(X) := a_0 + a_1 X = \rho_{XY} \frac{\sigma_Y}{\sigma_X} (X - m_X) + m_Y \quad (8.32)$$

のような X の関数となる．

8.4.3 条件付き期待値との関係

(8.32) の関数形は，算法 6.10 p50 で示した正規型確率変数の条件付き期待値

$$E[X_1|x_2] = \rho \frac{\sigma_1}{\sigma_2} (x_2 - m_2) + m_1 \quad (8.33)$$

と一致している．すなわち，(8.26) の線形回帰モデル \mathcal{M} 上の推定値は，次の性質を示す．

算法 8.3 $X, Y \in \mathcal{H}$ がガウス型の場合，

$$\mathcal{M} := \left\{ a_0 + a_1 X \mid a_0, a_1 \in \mathbb{R}; X \in \mathcal{H} \right\}$$

をモデルとする推定値 $\hat{Y} = \text{proj}(Y|\mathcal{M})$ は，次の条件付き期待値に一致する．

$$\hat{Y} = \text{proj}(Y|\mathcal{M}) = E[Y|X] = \rho_{XY} \frac{\sigma_Y}{\sigma_X} (X - m_X) + m_Y \quad (8.34)$$

演習 8.2 ガウス型確率変数 $X, Y \in \mathcal{H}$ について， X の平均を 0，分散を 1 とし， Y の平均を 1，分散を 4 とする．両者の相関係数を 3 とするとき，条件付き期待値 $E[Y|X]$ を求めよ．(X の関数となる)

これに対して, X, Y がガウス型でない場合は, $E[Y|X]$ の関数形が変わるので, 一般には $\text{proj}(Y|\mathcal{M}) \neq E[Y|X]$ である. ただし, 未知の確率変数の分布は, ガウス分布で近似されることも多い. その場合は, 算法 8.3 の解釈が使われることになる.

8.5 数 値 例

- Ch08.ipynb

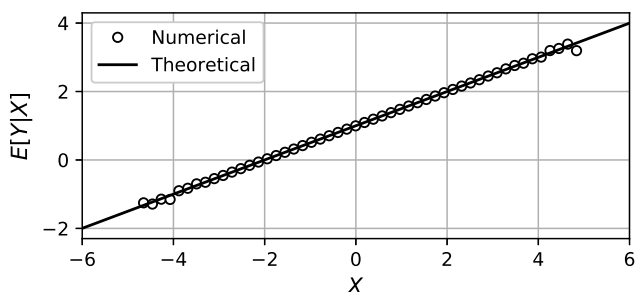


図 8.5 条件付き期待値との関係. 実線が算法 8.3 の理論式, 印は条件付き期待値 $E[Y|X]$ の数値計算値. の端部は標本点が疎らなため精度が落ちている.

9

確率ベクトルの推定

8.4 節の確率変数の推定問題を，確率ベクトルに拡張する．

9.1 条件付き期待値

2 個のガウス確率ベクトル $\mathbf{Y} := (Y_1, \dots, Y_m)^T$, $\mathbf{X} := (X_1, \dots, X_n)^T$ を考える．

9.1.1 確率ベクトルの条件付き期待値

このとき， $m + n$ 個の確率変数が共存するので，その法則は $m + n$ 変数結合密度関数

$$p(\mathbf{y}, \mathbf{x}) := p(y_1, \dots, y_m, x_1, \dots, x_n) \quad (9.1)$$

で記述される． \mathbf{X} を条件とする \mathbf{Y} の条件付き密度関数

$$p(\mathbf{y}|\mathbf{x}) = \frac{p(\mathbf{y}, \mathbf{x})}{p(\mathbf{x})} := \frac{p(y_1, \dots, y_m, x_1, \dots, x_n)}{p(x_1, \dots, x_n)} \quad (9.2)$$

を導入すると， \mathbf{x} を条件とする \mathbf{Y} の条件付き期待値の関数形は，

$$\langle \mathbf{Y} | \mathbf{x} \rangle = \int_{-\infty}^{\infty} \mathbf{y} p(\mathbf{y} | \mathbf{x}) d\mathbf{y} \quad (9.3)$$

で得られる．積分は y_1, \dots, y_m に関する m 重積分を意味するものとする．

9.1.2 多変数ガウス分布

$m + n$ 次元ベクトル $(\mathbf{X}^T, \mathbf{Y}^T)^T$ の共分散行列を次のように導入する．

$$\begin{aligned} \Sigma &:= \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix} \\ &:= \begin{pmatrix} \langle (\mathbf{Y} - \bar{\mathbf{y}})(\mathbf{Y} - \bar{\mathbf{y}})^T \rangle & \langle (\mathbf{Y} - \bar{\mathbf{y}})(\mathbf{X} - \bar{\mathbf{x}})^T \rangle \\ \langle (\mathbf{X} - \bar{\mathbf{x}})(\mathbf{Y} - \bar{\mathbf{y}})^T \rangle & \langle (\mathbf{X} - \bar{\mathbf{x}})(\mathbf{X} - \bar{\mathbf{x}})^T \rangle \end{pmatrix} \end{aligned} \quad (9.4)$$

\mathbf{X} の期待値を $\bar{\mathbf{x}}$ と書いた．このとき，結合密度関数は， $m + n$ 変数ガウス分布

$$p(\mathbf{y}, \mathbf{x}) = \frac{1}{(\sqrt{2\pi})^{m+n} \sqrt{|\Sigma|}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} \mathbf{y} - \bar{\mathbf{y}} \\ \mathbf{x} - \bar{\mathbf{x}} \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} \mathbf{y} - \bar{\mathbf{y}} \\ \mathbf{x} - \bar{\mathbf{x}} \end{pmatrix} \right\} \quad (9.5)$$

として与えられる．

9.1.3 ガウス確率ベクトルの条件付き期待値

算法 9.1 $m \times n$ 行列 A_0 を

$$A_0 \Sigma_{xx} = \Sigma_{yx} \quad (9.6)$$

の解とすると、 m 次元確率ベクトル (後に誤差ベクトルという)

$$E := Y - \bar{y} - A_0(X - \bar{x}) \quad (9.7)$$

は、 X と独立となる．また、 E の期待値と分散共分散行列は、

$$\langle E \rangle = 0, \quad \Sigma_{ee} = \Sigma_{yy} - A_0 \Sigma_{xy} = \Sigma_{yy} - A_0 \Sigma_{xx} A_0^T \quad (9.8)$$

例題 9.1 示せ．

算法 9.2 X を条件とする Y の条件付き期待値は、

$$\langle Y | X \rangle = \bar{y} + A_0(X - \bar{x}) \quad (9.9)$$

となる．ただし、 A_0 は、 $A_0 \Sigma_{xx} = \Sigma_{yx}$ を満足する $m \times n$ 行列である．

例題 9.2 示せ．

演習 9.1 例題 6.5 p51 を復習せよ．

9.2 確率ベクトルの推定

9.2.1 ヒルベルト空間

確率ベクトル X は成分が $X_i \in \mathcal{H}$ なので、次のような空間を動く．

$$\mathcal{H}^n := \{ X = (X_1, X_2, \dots, X_n)^T \mid X_i \in \mathcal{H} \} \quad (9.10)$$

$X, Z \in \mathcal{H}^n$ の内積を、

$$\langle X, Z \rangle := \langle X^T Z \rangle = \langle X_1 Z_1 + \dots + X_n Z_n \rangle = \text{tr} \langle X Z^T \rangle \quad (9.11)$$

のように定めると、これは定義 8.1 を満足する．ようにするに、確率ベクトルの内積とは、成分を普通に内積した後に期待値をとったものである．下線部はよく用いられる等価表現である．

▶▶ 下線部の tr は、行列のトレース (対角成分の和) を表す。例えば、

$$\begin{aligned}\text{tr} \langle \mathbf{x} \mathbf{z}^T \rangle &= \text{tr} \left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} (z_1, z_2) \right\rangle = \text{tr} \left\langle \begin{pmatrix} x_1 z_1 & x_1 z_2 \\ x_2 z_1 & x_2 z_2 \end{pmatrix} \right\rangle \quad (6.33) \text{ p51} \\ &= \text{tr} \left(\begin{pmatrix} \langle x_1 z_1 \rangle & \langle x_1 z_2 \rangle \\ \langle x_2 z_1 \rangle & \langle x_2 z_2 \rangle \end{pmatrix} \right) = \langle x_1 z_1 \rangle + \langle x_2 z_2 \rangle = \langle x_1 z_1 + x_2 z_2 \rangle \\ &= \langle \mathbf{x}^T \mathbf{z} \rangle \text{ (内積)}\end{aligned}$$

▶▶ 平均操作は、 $2n$ 変数密度関数 $p(x_1, \dots, x_n, z_1, \dots, z_n)$ に対してとる。

$$\begin{aligned}\langle f(X_1, \dots, X_n, Z_1, \dots, Z_n) \rangle \\ = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, z_n) p(x_1, \dots, z_n) dx_1 \dots dz_n\end{aligned} \quad (9.12)$$

ノルムは、これまでと同様に、自分自身との内積の平方根

$$\| \mathbf{X} \| := \sqrt{\langle \mathbf{X}, \mathbf{X} \rangle} = \sqrt{\langle \mathbf{X}^T \mathbf{X} \rangle} \quad (9.13)$$

とすればよい。ここで、 $X_i \in \mathcal{H}$ より $\langle X_i^2 \rangle < \infty$ なので、

$$\| \mathbf{X} \|^2 = \langle X_1^2 + \dots + X_n^2 \rangle = \langle X_1^2 \rangle + \dots + \langle X_n^2 \rangle < \infty \quad (9.14)$$

という制約条件が自動的に課される。ゆえに、 \mathcal{H}^n もまたヒルベルト空間である。

9.2.2 推定問題

ヒルベルト空間上のベクトル $\mathbf{X} \in \mathcal{H}^n$ は、計算方法が普通のベクトルと同じである。これは確率変数 $X \in \mathcal{H}$ も同様だった。というわけで、確率変数に対する算法 8.2 p70 と全く同じ算法が通用する。若干集約して示すと、

算法 9.3 (確率ベクトルの推定) \mathcal{M} を確率ベクトル U_1, \dots, U_n で張られた部分空間とする。誤差 $\| \mathbf{Y} - \hat{\mathbf{Y}} \|$ を最小化する推定値 $\hat{\mathbf{Y}} \in \mathcal{M}$ は、 $\mathbf{Y} \in \mathcal{H}^m$ の \mathcal{M} 上への直交射影 $\hat{\mathbf{Y}} = \text{proj}(\mathbf{Y} | \mathcal{M})$ となる ($\iff (\mathbf{Y} - \hat{\mathbf{Y}}) \perp \mathcal{M}$)。そのための必要十分条件は、

$$\langle (\mathbf{Y} - \hat{\mathbf{Y}}), U_i \rangle = 0 \quad (i = 1, \dots, n) \quad (9.15)$$

で与えられる。

9.3 確率ベクトルの線形回帰

9.3.1 確率ベクトル \mathbf{X} からの線形回帰

$\mathbf{Y} \in \mathcal{H}^m$ のモデルとして、次のような線形回帰モデルを考える。

$$\mathcal{M} := \{ \mathbf{a}_0 + A\mathbf{X} \mid \mathbf{a}_0 \in \mathcal{H}^m, \mathbf{X} \in \mathcal{H}^n, A \in \mathbb{R}^{m \times n} \} \quad (9.16)$$

\mathbf{a}_0 は定数ベクトル, A は定数行列である. 定数項を, 定ベクトル値関数

$$\mathbf{a}_0(\omega) = \mathbf{a}_0 \quad \forall \omega \in \Omega \quad (9.17)$$

と見なすと[†], \mathcal{M} は確率ベクトル $\mathbf{a}_0, A\mathbf{X} \in \mathcal{H}^m$ で張られた, \mathcal{H}^m の部分空間となる.

算法 9.3 を $U_1 = \mathbf{a}_0, U_2 = A\mathbf{x}$ として使うと, 推定値 $\hat{\mathbf{Y}} = \text{proj}(\mathbf{Y}|\mathcal{M})$ について,

$$\begin{cases} 0 = \langle \mathbf{Y} - \hat{\mathbf{Y}}, \mathbf{a}_0 \rangle := \langle (\mathbf{Y} - \hat{\mathbf{Y}})^T \mathbf{a}_0 \rangle = \langle (\mathbf{Y} - \hat{\mathbf{Y}})^T \rangle \mathbf{a}_0 \\ 0 = \langle \mathbf{Y} - \hat{\mathbf{Y}}, A\mathbf{X} \rangle := \text{tr} \langle (\mathbf{Y} - \hat{\mathbf{Y}})(A\mathbf{X})^T \rangle \\ \quad = \text{tr} \left(\langle (\mathbf{Y} - \hat{\mathbf{Y}})\mathbf{X}^T \rangle A^T \right) \end{cases} \quad (9.18)$$

が成立するが, 特殊な場合 $\mathbf{a}_0 = \mathbf{0}, A = O$ を除くと, 上式は,

$$\begin{cases} 0 = \langle \mathbf{Y} - \hat{\mathbf{Y}} \rangle \\ O = \langle (\mathbf{Y} - \hat{\mathbf{Y}})\mathbf{X}^T \rangle \quad (\text{ゼロ行列}) \end{cases} \quad (9.19)$$

と等価である. これに, 推定値の候補 $\hat{\mathbf{Y}} = \mathbf{a}_0 + A\mathbf{X}$ を代入すると,

$$\begin{cases} 0 = \langle \mathbf{Y} - \mathbf{a}_0 - A\mathbf{X} \rangle = \bar{\mathbf{y}} - \mathbf{a}_0 - A\bar{\mathbf{x}} \\ O = \langle (\mathbf{Y} - \mathbf{a}_0 - A\mathbf{X})\mathbf{X}^T \rangle = \langle \mathbf{Y}\mathbf{X}^T \rangle - \mathbf{a}_0\bar{\mathbf{x}}^T - A\langle \mathbf{X}\mathbf{X}^T \rangle \end{cases} \quad (9.20)$$

\mathbf{X} の期待値を $\bar{\mathbf{x}}$ と書いた. 第 1 式 $\mathbf{a}_0 = \bar{\mathbf{y}} - A\bar{\mathbf{x}}$ を第 2 式に代入すると,

$$O = \langle \mathbf{Y}\mathbf{X}^T \rangle - \bar{\mathbf{y}}\bar{\mathbf{x}}^T + A\bar{\mathbf{x}}\bar{\mathbf{x}}^T - A\langle \mathbf{X}\mathbf{X}^T \rangle \quad (9.21)$$

となる. この式は, 前節の記号 $\langle \mathbf{Y}\mathbf{X}^T \rangle - \bar{\mathbf{y}}\bar{\mathbf{x}}^T = \Sigma_{yx}, \langle \mathbf{X}\mathbf{X}^T \rangle - \bar{\mathbf{x}}\bar{\mathbf{x}}^T = \Sigma_{xx}$ を使うと, $\Sigma_{yx} = A\Sigma_{xx}$ と書ける. これを満足する A を用いて, 推定値は,

$$\hat{\mathbf{Y}} = \mathbf{a}_0 + A\mathbf{X} = \bar{\mathbf{y}} + A(\mathbf{X} - \bar{\mathbf{x}}) \quad (9.22)$$

となる. これは, 算法 9.2 の条件付き期待値 $\langle \mathbf{Y}|\mathbf{X} \rangle$ と, 関数形が同じである.

以上の計算をまとめると, 線形回帰に関する次のような結果が得られる.

算法 9.4 (確率ベクトルからの線形回帰) 定数ベクトル \mathbf{a}_0 と定数行列 A をパラメータとする線形回帰モデル

$$\mathcal{M} := \{ \mathbf{a}_0 + A\mathbf{X} \mid \mathbf{a}_0 \in \mathcal{H}^m, \mathbf{X} \in \mathcal{H}^n, A \in \mathbb{R}^{m \times n} \} \subset \mathcal{H}^m \quad (9.23)$$

[†] “ \forall ” の使い方は, A.1.6 節を見よ.

において，誤差 $\|Y - \hat{Y}\|$ を最小化する推定値は，直交射影 $\hat{Y} = \text{proj}(Y|\mathcal{M}) \in \mathcal{M}$ となる．そうなるための必要十分条件は，次式となる．

$$\langle Y - \hat{Y} \rangle = 0, \quad \langle (Y - \hat{Y})X^T \rangle = O \quad (\text{ゼロ行列}) \quad (9.24)$$

これより推定値 \hat{Y} は，

$$\Sigma_{yx} = A\Sigma_{xx} \quad (9.25)$$

を満足する行列 A を用いて，次のように書ける．

$$\hat{Y} = a_0 + AX = \bar{y} + A(X - \bar{x}) \quad (9.26)$$

特に Y, X がガウス型の場合，推定値 \hat{Y} は，次の条件付き期待値に一致する．

$$\hat{Y} = \text{proj}(Y|\mathcal{M}) = \langle Y|X \rangle \quad (9.27)$$

9.3.2 確率ベクトル列 X_1, \dots, X_n からの線形回帰

算法 9.4 は，独立変数が複数の場合に拡張できる．

算法 9.5 (確率ベクトル列からの線形回帰) 定数ベクトル a_0 と定数行列 A_1, \dots, A_k をパラメータとする線形回帰モデル

$$\mathcal{M} := \left\{ a_0 + A_1 X_1 + \dots + A_k X_k \mid a_0 \in \mathcal{H}^m, X_i \in \mathcal{H}^n, A_i \in \mathbb{R}^{m \times n} \right\} \subset \mathcal{H}^m \quad (9.28)$$

において，誤差 $\|Y - \hat{Y}\|$ を最小化する推定値は，直交射影 $\hat{Y} = \text{proj}(Y|\mathcal{M}) \in \mathcal{M}$ となる．そうなるための必要十分条件は，次式となる．

$$\langle Y - \hat{Y} \rangle = 0, \quad \langle (Y - \hat{Y})X_i^T \rangle = O \quad (i = 1, \dots, k) \quad (9.29)$$

これより推定値 \hat{Y} は，

$$\Sigma_{yx_i} = \sum_{j=1}^k A_j \Sigma_{x_j x_i} \quad (i = 1, \dots, k) \quad (9.30)$$

を満足する行列 A_j を用いて，次のように書ける．

$$\hat{Y} = a_0 + \sum_{j=1}^k A_j X_j = \bar{y} + \sum_{j=1}^k A_j (X_j - \bar{x}_j) \quad (9.31)$$

特に Y, X_1, \dots, X_k がガウス型の場合，推定値 \hat{Y} は，次の条件付き期待値に一致する．

$$\hat{Y} = \text{proj}(Y|\mathcal{M}) = \langle Y | X_1, \dots, X_k \rangle \quad (9.32)$$

右辺は、 X_1, \dots, X_k を条件とする Y の条件付き期待値である。

例題 9.3 示せ。

例題の解答例

例題 9.1 p74

まず , $\langle E \rangle = \langle Y \rangle - \bar{y} - A_0(\langle X \rangle - \bar{x}) = 0$. また ,

$$\begin{aligned}\langle EX^T \rangle &= \langle (Y - \bar{y})X^T \rangle - A_0 \langle (X - \bar{x})X^T \rangle \\ &= \langle (Y - \bar{y})(X - \bar{x})^T \rangle - A_0 \langle (X - \bar{x})(X - \bar{x})^T \rangle \\ &\quad + \langle Y - \bar{y} \rangle \bar{x}^T - A_0 \langle X - \bar{x} \rangle \bar{x}^T \\ &= \Sigma_{yx} - A_0 \Sigma_{xx} = 0\end{aligned}$$

より , E, X は無相関であり , ガウス性より独立でもある . さらに ,

$$\begin{aligned}\Sigma_{ee} &:= \left\langle \left(Y - \bar{y} - A_0(X - \bar{x}) \right) \left((Y - \bar{y})^T - (X - \bar{x})^T A_0^T \right) \right\rangle \\ &= \langle (Y - \bar{y})(Y - \bar{y})^T \rangle + A_0 \langle (X - \bar{x})(X - \bar{x})^T \rangle A_0^T \\ &\quad - A_0 \langle (X - \bar{x})(Y - \bar{y})^T \rangle - \langle (Y - \bar{y})(X - \bar{x})^T \rangle A_0^T \\ &= \Sigma_{yy} + \underline{A_0 \Sigma_{xx} A_0^T - A_0 \Sigma_{xy} - \Sigma_{yx} A_0^T} = \Sigma_{yy} - A_0 \Sigma_{xx} A_0^T\end{aligned}$$

$\therefore \Sigma_{xy} = \Sigma_{yx}^T = \Sigma_{xx} A_0^T$ より , 下線部 = 0 .

例題 9.2 p74

E と X は独立だから , $\langle E|X \rangle = \langle E \rangle$ となるが , 算法 9.1 より $\langle E \rangle = 0$ である . これと
算法 6.11 p50 より ,

$$\begin{aligned}0 &= \langle E|x \rangle = \langle Y - \bar{y} - A_0(X - \bar{x})|x \rangle = \langle Y|x \rangle - \bar{y} - A_0(\langle X|x \rangle - \bar{x}) \\ &= \langle Y|x \rangle - \bar{y} - A_0(x - \bar{x}) \quad \therefore \langle Y|x \rangle = \bar{y} + A_0(x - \bar{x})\end{aligned}$$

となって , $\langle Y|X \rangle$ の関数形を得る .

例題 9.3 p78

\mathcal{M} は $a_0, A_1 X_1, \dots, A_k X_k$ によって張られるので , 算法 9.3 より ,

$$\begin{cases} 0 = \langle (Y - \hat{Y}), a_0 \rangle = \langle (Y - \hat{Y})^T \rangle a_0 \\ 0 = \langle (Y - \hat{Y}), A_i X_i \rangle = \text{tr} \left(\langle (Y - \hat{Y}), X_i^T \rangle A_i^T \right) \quad (i = 1, \dots, k) \end{cases}$$

となるから，特別な a_0, A_i を除いて，(9.29) が成立する．(9.29) に推定値の候補 $\hat{Y} := a_0 + \sum_{j=1}^k A_j X_j$ を代入すると，第 1 式より $a_0 = \bar{y} - \sum_{j=1}^k A_j \bar{x}_j$ を得る．これを第 2 式に代入すると，

$$\begin{aligned} O &= \langle Y X_i^T \rangle - a_0 \bar{x}_i^T - \sum_{j=1}^k A_j \langle X_j X_i^T \rangle \\ &= \langle Y X_i^T \rangle - \bar{y} \bar{x}_i^T - \sum_{j=1}^k A_j (\langle X_j X_i^T \rangle - \bar{x}_j \bar{x}_i^T) = \Sigma_{yx_i} - \sum_{j=1}^k A_j \Sigma_{x_j x_i} \end{aligned}$$

となって (9.30) を得る．(9.31) は a_0 を代入すれば明らか．最後の証明は略す．

10 || 確 率 過 程

10.1 確 率 過 程

確率変数 X とは、見本 $\omega \in \Omega$ の関数 $x = X(\omega)$ であった。その関数形が時間 t とともに変化するものを $X(t, \omega)$ または $X_t(\omega)$ と書き、確率過程という。

特に、時間 t が実数のものを連続時間確率過程と呼び、整数のものを離散時間確率過程と呼ぶ。以下では、離散時間確率過程を扱う。

10.1.1 確率過程の記述

離散時間確率過程 (以下、確率過程) を考える。

$$X_0, X_1, \dots, X_n, \dots \quad (10.1)$$

時刻 $t = 0$ では、1 個の確率変数 X_0 しかないので、その性質は密度関数

$$p(x_0) \quad (10.2)$$

で表せる。時刻 $t = 1$ では、2 個の確率変数 X_0, X_1 が共存するので、その確率的な性質は 2 変数結合密度関数

$$p(x_1, x_0) \quad (10.3)$$

で表される。このあとの議論の都合上、未来を左に書いた。

さらに時刻を進めると、順次 $p(x_2, x_1, x_0), p(x_3, x_2, x_1, x_0), \dots$ が必要になり、時刻 $t = n$ までの性質を表すには、 $n + 1$ 変数結合密度関数

$$p(x_n, x_{n-1}, \dots, x_1, x_0) \quad (10.4)$$

が必要になる。

応用の場面において、 n は 10 や 20 では済まない。そこで理論上の枠組みとして、 $n \rightarrow \infty$ を想定すると、 ∞ 変数結合密度関数が必要になってしまう。これを具体的に扱うことは難しい。

10.1.2 白 色 雑 音

そこで、最も単純な確率過程として、 X_t ($t = 0, 1, \dots, n$) は互いに独立だと仮定する。このとき、結合密度関数は、

$$p(x_n, x_{n-1}, \dots, x_1, x_0) = p(x_n)p(x_{n-1}) \cdots p(x_1)p(x_0) \quad (10.5)$$

のように、 $p(x_t)$ ($t = 0, 1, \dots, n$) の積になる。こうなる確率過程を白色 (はくしょく) 雑音とよび、これを本書では W_t と書く。

白色雑音の性質は、各時刻の密度関数 $p(x_t)$ で完全に記述される。また、独立なら無相関であるから、異なる時刻の W_t, W_s は無相関である。

算法 10.1 (白色雑音の性質) 白色雑音 W_t は次の性質をもつ。

$$\text{Cov}[W_t, W_s] = \langle (W_t - \bar{w}_t)(W_s - \bar{w}_s) \rangle = V[W_t] \delta_{ts} \quad (10.6)$$

$V[W_t]$ は W_t の分散であり、 δ_{ts} はクロネッカーのデルタである。

$$\delta_{ts} := \begin{cases} 1 & (t = s) \\ 0 & (t \neq s) \end{cases} \quad (10.7)$$

この性質を、白色雑音のデルタ相関という。

10.1.3 マルコフ過程

2 番目に単純な確率過程として、 X_t は X_{t-1} にだけ依存すると仮定する。この性質は、5.4.4 節の条件付密度関数を用いると、次のように表わせる。

$$p(x_n | x_{n-1}, \dots, x_0) = p(x_n | x_{n-1}) \quad (10.8)$$

順次、適用すると、

$$\begin{aligned} p(x_n, x_{n-1}, \dots, x_0) &= p(x_n | x_{n-1}, \dots, x_0) p(x_{n-1}, x_{n-2}, \dots, x_0) \\ &= p(x_n | x_{n-1}) p(x_{n-1}, x_{n-2}, \dots, x_0) \\ &= p(x_n | x_{n-1}) p(x_{n-1} | x_{n-2}) p(x_{n-2}, x_{n-3}, \dots, x_0) \\ &= p(x_n | x_{n-1}) p(x_{n-1} | x_{n-2}) \cdots p(x_1 | x_0) p(x_0) \end{aligned} \quad (10.9)$$

となる。こうなる確率過程をマルコフ過程という。各項の条件付き密度関数

$$p(x_t | x_{t-1}) \quad (t = 1, \dots, n) \quad (10.10)$$

は、状態を単位時間だけ進める役割を演じることから、推移密度関数と呼ばれる。

マルコフ過程の性質は、初期密度関数 $p(x_0)$ と、推移密度関数 $p(x_t|x_{t-1})$ によって完全に記述される。

【算法 10.2 (マルコフ過程の性質)】 マルコフ過程 X_t の推移密度関数について、次の公式が成立する。

$$p(x_t|x_s) = \int_{-\infty}^{\infty} p(x_t|x_m) p(x_m|x_s) dx_m \quad (10.11)$$

これをチャップマン–コルモゴロフ方程式という。

10.1.4 ガウス過程

各項 X_0, X_1, \dots が全てガウス型であるような確率過程 X_t を、ガウス過程という。特に、ガウス型の白色雑音をガウス白色雑音と呼び、ガウス型のマルコフ過程をガウス–マルコフ過程という。

10.1.5 ベクトル値の確率過程

時間とともに変化する確率ベクトル X_t を、ベクトル値過程という。成分の性質に応じて、白色型、マルコフ型、ガウス型などの接頭語がつく。

【算法 10.3 (ガウス型ベクトル値過程の性質)】 X_t をガウス型ベクトル値過程とする。各項の線形変換の和

$$Z = A_0 X_0 + A_1 X_1 + \dots + A_t X_t \quad (10.12)$$

は、ガウス確率ベクトルとなる。

▶ 証明 算法 7.3 p58 より、各項 $Y_i = A_i X_i$ はガウス確率ベクトルである。ゆえに、その総和 Z の各成分はガウス確率変数の和となるので、(7.10) p54 より、 Z の成分は全てガウス確率変数となる。そのようなベクトル Z をガウス確率ベクトルと呼んだ。

10.2 確率的な状態方程式

10.2.1 決定論的な状態方程式

システムの状態をベクトル $x(t)$ で表し、その時間発展を差分方程式

$$x(t+1) = F_t x(t), \quad x(0) = x_0, \quad (t = 0, 1, 2, \dots) \quad (10.13)$$

でモデル化することがよく行われる． F_t はモデルの構造を表す行列で，ここでは時変（時間 t に依存する）とした． x_0 は初期値で，時刻 $t = 0$ の状態を表す．問題に即した F_t の成分をとることで，現実の様々な状態量の時間発展を模倣できる．

(10.13) の解は，逐次代入 $x(1) = F(0)x_0$, $x(2) = F(1)x(1)$ によって求まる．

$$x(t) = \Phi(t, 0)x_0, \quad \Phi(t, 0) := F(t-1)F(t-2) \cdots F(1)F(0) \quad (10.14)$$

このように，決定論的な状態方程式 (10.13) の状態 $x(t)$ は，逐次代入によって直ちに定まる．なお，行列 $\Phi(t, 0) := F(t-1)F(t-2) \cdots F(1)F(0)$ を，推移行列という．

10.2.2 ランダムな外乱を受ける場合

次に，(10.13) がランダムな外乱 $w(t)$ を受ける場合を考える（ w はダブリューの太字であり，見本を表すオメガ ω とは異なる）．

$$x(t+1) = F_t x(t) + G_t w(t), \quad x(0) = x_0, \quad (t = 0, 1, 2, \dots) \quad (10.15)$$

G_t は外乱の入り方を表す行列である．さて，この場合，ランダムな外乱 $w(t)$ は試行ごとに变化する．したがって，状態量 $x(t)$ も試行ごとに变化する．

10.2.3 確率的な状態方程式

そこで，状態 x_t と外乱 w_t を確率過程としてモデル化する．なお，今後は行列とベクトルを区別するため，確率ベクトルを小文字で表す．

$$x_{t+1} = F_t x_t + G_t w_t, \quad (t = 0, 1, 2, \dots) \quad (10.16)$$

初期値 x_0 は，所与の確率ベクトルとする．以下，これを単に状態方程式と呼ぶことにする．各ベクトルと行列の次元と名称を下表に示す．

x_t	状態量	n 次元確率ベクトル
w_t	システム雑音	r 次元確率ベクトル
F_t	状態推移行列	$n \times n$ 定数行列
G_t	駆動行列	$n \times r$ 定数行列

このシステムについて次の仮定をおく．

(1) 初期値に対する仮定 初期状態 x_0 はガウス確率ベクトルで，

$$\langle x_0 \rangle := \bar{x}_0, \quad \text{Cov}[x_0] = \langle (x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T \rangle := \Sigma_0 \quad (10.17)$$

とする．

(2) 雑音に対する仮定 w_t は r 次元ガウス白色雑音で、全ての時刻 t で、

$$\langle w_t \rangle = \mathbf{0} \quad (10.18)$$

とする。白色雑音なので、異なる時刻の w_t, w_s は無相関となる。

$$\langle w_t w_s^T \rangle = Q_t \delta_{ts} \quad (10.19)$$

Q_t は時刻 t の w_t の共分散行列とする (同時刻の成分間の相関は否定しない)。 δ_{ts} はクロネッカーのデルタである。

さらに、どの時刻 t の雑音 w_t も、初期状態 x_0 と独立とする。このとき、さきの仮定 $\langle w_t \rangle = \mathbf{0}$ より

$$\langle x_0 w_t^T \rangle \stackrel{\text{独立}}{=} \langle x_0 \rangle \langle w_t^T \rangle = \bar{x}_0 \mathbf{0}^T = \mathbf{0} \quad \forall t \geq 0 \quad (10.20)$$

となる[†]。すなわち、 $x_0 \perp w_t$ が成立する。

10.3 期待値と共分散行列の推移則

10.3.1 確率ベクトルの推移

状態方程式 (10.16) の時刻 $s < t$ からの推移を追うと、

$$\begin{aligned} x_{s+1} &= F_s x_s + G_s w_s \\ x_{s+2} &= F_{s+1} x_{s+1} + G_{s+1} w_{s+1} \\ &= F_{s+1} F_s x_s + F_{s+1} G_s w_s + G_{s+1} w_{s+1} \\ x_{s+3} &= F_{s+2} x_{s+2} + G_{s+2} w_{s+2} \\ &= F_{s+2} F_{s+1} F_s x_s + F_{s+2} F_{s+1} G_s w_s + F_{s+2} G_{s+1} w_{s+1} + G_{s+2} w_{s+2} \end{aligned}$$

となる。ここで、 F_t の作用の累積を表す推移行列

$$\Phi(t, s) := \begin{cases} F_{t-1} F_{t-2} \cdots F_s & (s < t) \\ I & (s = t) \end{cases} \quad \text{単位行列} \quad (10.21)$$

を導入すると、解は次のように表示できる。

$$x_t = \Phi(t, s) x_s + \sum_{k=s}^{t-1} \Phi(t, k+1) G_k w_k \quad (10.22)$$

[†] “ \forall ” の使い方は、A.1.6 節を見よ。

10.3.2 状態と雑音の独立性

ベクトル v_1, \dots, v_n の線形結合の全体を $\text{Span}\{v_1, \dots, v_n\}$ と書く (v_1, \dots, v_n で張られる空間のこと) . まず, (10.22) に $s = 0$ を代入すると,

$$x_t = \Phi(t, 0)x_0 + \sum_{k=0}^{t-1} \Phi(t, k+1)G_k w_k \quad (10.23)$$

となるので, 解は,

$$x_t \in \text{Span}\{x_0, G_0 w_0, \dots, G_{t-1} w_{t-1}\} \quad (10.24)$$

となることが分かる. 10.2.3(2) 節より, x_0, w_0, \dots, w_{t-1} は, w_t と独立だから, それらの線形変換 $x_0, G_0 w_0, \dots, G_{t-1} w_{t-1}$ も w_t と独立である. ゆえに, それらの線形結合である x_t は, w_t と独立となる. すなわち, x_t は同時刻以降の w_s ($t \leq s$) と独立である. このとき,

$$\langle x_t w_s^T \rangle \stackrel{\text{独立}}{=} \langle x_t \rangle \langle w_s^T \rangle = \langle x_t \rangle \mathbf{0}^T = \mathbf{0} \quad (t \leq s) \quad (10.25)$$

すなわち, $x_t \perp w_s$ ($t \leq s$) が成立する.

同様に, x_0, w_0, \dots, w_{t-1} はガウス確率ベクトルと仮定したから, その線形変換の和 x_t はガウス確率ベクトルである. すなわち, x_t はガウス過程である.

最後に, 状態方程式,

$$x_t = F_{t-1}x_{t-1} + G_{t-1}w_{t-1} \in \text{Span}\{x_{t-1}, w_{t-1}\}$$

より, x_t は 1 時刻前の x_{t-1}, w_{t-1} のみに従属するので, x_t はマルコフ過程である.

以上をまとめると, 次の結果が得られる.

算法 10.4 x_t は, w_s ($t \leq s$) と独立なガウス-マルコフ過程である.

10.3.3 期待値と共分散行列の推移

x_t はガウス過程であることが分かったので, 各時刻の x_t の性質は, 次の期待値と共分散行列によって完全に記述される.

$$\langle x_t \rangle = \bar{x}_t, \quad \Sigma_t = \langle (x_t - \bar{x}_t)(x_t - \bar{x}_t)^T \rangle \quad (10.26)$$

さらに, x_t はマルコフ過程なので, \bar{x}_t や Σ_t は 1 時刻前の情報だけに依存するはずである.

まず期待値の推移だが, 状態方程式 (10.16) の両辺の期待値をとると,

$$\langle x_{t+1} \rangle = F_t \langle x_t \rangle + G_t \langle w_t \rangle = F_t \langle x_t \rangle + G_t \mathbf{0} \quad (10.27)$$

となるので、次の結果が得られる。

$$\bar{x}_{t+1} = F_t \bar{x}_t \quad (10.28)$$

続いて、 Σ_{t+1} を Σ_t で書き下す方向で変形を進める。まず、

$$\Sigma_{t+1} := \langle (x_{t+1} - \bar{x}_{t+1})(x_{t+1} - \bar{x}_{t+1})^T \rangle = \langle x_{t+1} x_{t+1}^T \rangle - \bar{x}_{t+1} \bar{x}_{t+1}^T \quad (10.29)$$

となるが、右辺第2項は、(10.28) より、

$$\bar{x}_{t+1} \bar{x}_{t+1}^T = F_t \bar{x}_t \bar{x}_t^T F_t^T \quad (10.30)$$

となる。右辺第1項は、状態方程式 (10.16) を代入すると、

$$\begin{aligned} \langle x_{t+1} x_{t+1}^T \rangle &= \langle (F_t x_t + G_t w_t)(F_t x_t + G_t w_t)^T \rangle = F_t \langle x_t x_t^T \rangle F_t^T \\ &\quad + F_t \langle x_t w_t^T \rangle G_t^T + G_t \langle w_t x_t^T \rangle F_t^T + G_t \langle w_t w_t^T \rangle G_t^T \end{aligned} \quad (10.31)$$

となる。 x_t と w_t の共分散行列は、状態方程式の解 (10.22) を使って、 x_t を x_0 まで遡及すると消去できる。実際、(10.20) と、 w_t のデルタ相関 (10.19) より、

$$\begin{aligned} \langle x_t w_t^T \rangle &= \left\langle \left(\Phi(t, 0) x_0 + \sum_{k=0}^{t-1} \Phi(t, k+1) G_k w_k \right) w_t^T \right\rangle \\ &= \Phi(t, 0) \underbrace{\langle x_0 w_t^T \rangle}_{=O} + \sum_{k=0}^{t-1} \Phi(t, k+1) G_k \underbrace{\langle w_k w_t^T \rangle}_{=O} = O \end{aligned} \quad (10.32)$$

を得る。これを転置すると $\langle w_t x_t^T \rangle = O$ も示される。以上を (10.29) に代入すると、

$$\begin{aligned} \Sigma_{t+1} &= F_t \langle x_t x_t^T \rangle F_t^T - F_t \bar{x}_t \bar{x}_t^T F_t^T + G_t \langle w_t w_t^T \rangle G_t^T \\ &= F_t \langle x_t x_t^T - \bar{x}_t \bar{x}_t^T \rangle F_t^T + G_t \langle w_t w_t^T \rangle G_t^T \\ &= F_t \Sigma_t F_t^T + G_t Q_t G_t^T \end{aligned}$$

が得られる。

以上、目論み通り、 \bar{x}_{t+1} 、 Σ_{t+1} を1時刻前の \bar{x}_t 、 Σ_t で表せた。すなわち、

算法 10.5 (期待値と共分散行列の推移則) (10.16) の状態量 x_t の期待値と共分散行列

$$\bar{x}_t := \langle x_t \rangle, \quad \Sigma_t := \langle (x_t - \bar{x}_t)(x_t - \bar{x}_t)^T \rangle \quad (10.33)$$

は、次の推移則で時間発展する。

$$\bar{x}_{t+1} = F_t \bar{x}_t, \quad (10.34)$$

$$\Sigma_{t+1} = F_t \Sigma_t F_t^T + G_t Q_t G_t^T \quad (t = 0, 1, 2, \dots) \quad (10.35)$$

演習 10.1 恒等的に $Q_t = O$ のとき, Σ_t は恒等的にどうなるか, 物理的に考察せよ. ヒントは (10.16) と (10.19) .

10.4 モンテカルロ法による数値例

$n = 2$ の数値例を示す. 行列は全て定行列とし, その成分と初期値を次のようにとる.

$$F = \begin{pmatrix} 0 & 0.5 \\ -0.8 & 0 \end{pmatrix}, G = \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix}, Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, x_0 = \begin{pmatrix} 0.1 \\ 0 \end{pmatrix} \quad (10.36)$$

10.4.1 標本路の数値計算

確率過程 $X(t, \omega)$ において, 見本 ω を特定の $\bar{\omega}$ に固定すると, 対応する時間の関数 $X(t, \bar{\omega})$ がピックアップされる. これを, 確率過程の標本路 (sample path) という.

標本路を数値的に求めるには, 状態方程式への逐次代入

$$x_1 = Fx_0 + Gw_0, \quad x_2 = Fx_1 + Gw_1, \quad x_3 = Fx_2 + Gw_2, \quad \dots$$

により, 初期値 x_0 からの時間発展 x_t を求めればよい. その際, ガウス白色雑音

$$w_t = \begin{pmatrix} w_t^1 \\ w_t^2 \end{pmatrix} \quad (10.37)$$

の各成分 w_t^1, w_t^2 に, 順次, 正規乱数列 N_0, N_1, N_2, \dots を代入する. 正規乱数列の部分列もまた正規乱数列なので, 代入の順序は任意でかまわない. ようするに,

$$\begin{aligned} \omega = 1 : & \begin{cases} w_0^1 = N_0, & w_1^1 = N_1, & w_2^1 = N_2, & \dots, & w_{15}^1 = N_{15} \\ w_0^2 = N_{16}, & w_1^2 = N_{17}, & w_2^2 = N_{18}, & \dots, & w_{15}^2 = N_{31} \end{cases} \\ \omega = 2 : & \begin{cases} w_0^1 = N_{32}, & w_1^1 = N_{33}, & w_2^1 = N_{34}, & \dots, & w_{15}^1 = N_{47} \\ w_0^2 = N_{48}, & w_1^2 = N_{49}, & w_2^2 = N_{50}, & \dots, & w_{15}^2 = N_{53} \end{cases} \\ & \vdots \end{aligned}$$

としてもよいし,

$$\begin{aligned} \omega = 1 : & \begin{cases} w_0^1 = N_0, & w_1^1 = N_2, & w_2^1 = N_4, & \dots, & w_{15}^1 = N_{30} \\ w_0^2 = N_1, & w_1^2 = N_3, & w_2^2 = N_5, & \dots, & w_{15}^2 = N_{31} \end{cases} \\ \omega = 2 : & \begin{cases} w_0^1 = N_{32}, & w_1^1 = N_{34}, & w_2^1 = N_{36}, & \dots, & w_{15}^1 = N_{52} \\ w_0^2 = N_{33}, & w_1^2 = N_{35}, & w_2^2 = N_{37}, & \dots, & w_{15}^2 = N_{53} \end{cases} \\ & \vdots \end{aligned}$$

としてもよい．他の代入順でも，代入ごとに乱数が更新されていれば問題ない．

もちろん，代入順を代えと，ガウス白色雑音の成分 w_t^1, w_t^2 の時間波形は変化するので（分布は同じ），得られる標本路の波形は変化する．

このようにして求めた標本路の数値例を図 10.1 に示す．標本数は 50 とした．

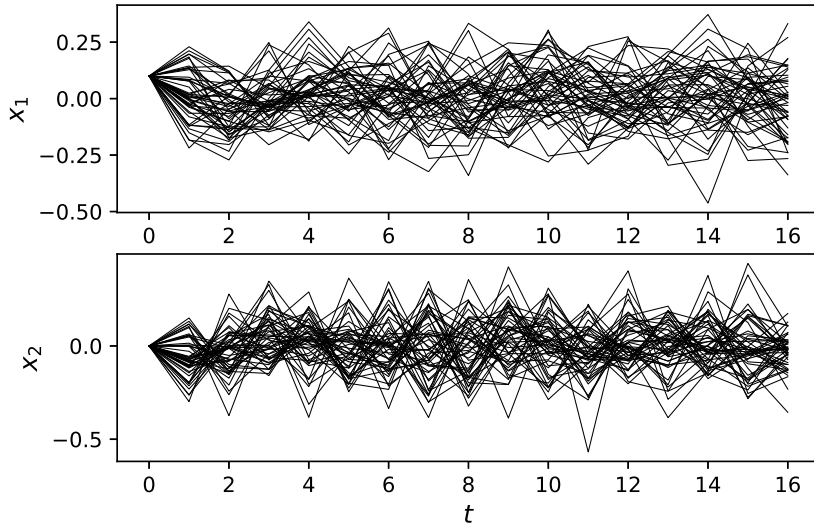


図 10.1 標本路の時間発展

10.4.2 平均値と共分散行列の数値計算 — モンテカルロ法

図 10.1 のように生成した複数の標本路の値を，各時刻 t において平均したものが，図 10.2 の最上段の○印である．同様に求めた分散行列の成分を 2 段目以降に示す．

このような，数値計算で生成した複数回の試行結果から，統計量を求める方法をモンテカルロ法 (Monte-Carlo method) という．

モンテカルロ法で求める平均値や共分散は標本平均であるから，乱数列 N_i の与え方によって計算結果は変化する．この変化を小さくするには，算法 6.3 p45 の中心極限定理で学んだように，標本数を十分大きくとる必要がある．この数値例では，標本数を 5×10^4 としており，相応の計算時間を要した．

10.4.3 平均値と共分散行列 — 算法 10.5

これに対して，算法 10.5 の差分方程式の数値解を，図 10.2 の実線で示す．この結果は理論値であるため，中心極限定理のような計算結果の不確定性は生じない．

このように，確率解析による算法 10.5 を用いると，モンテカルロ法より高精度な結果を，極めて少数回の逐次代入計算 (図 10.2 では 16 回) で得ることができる．

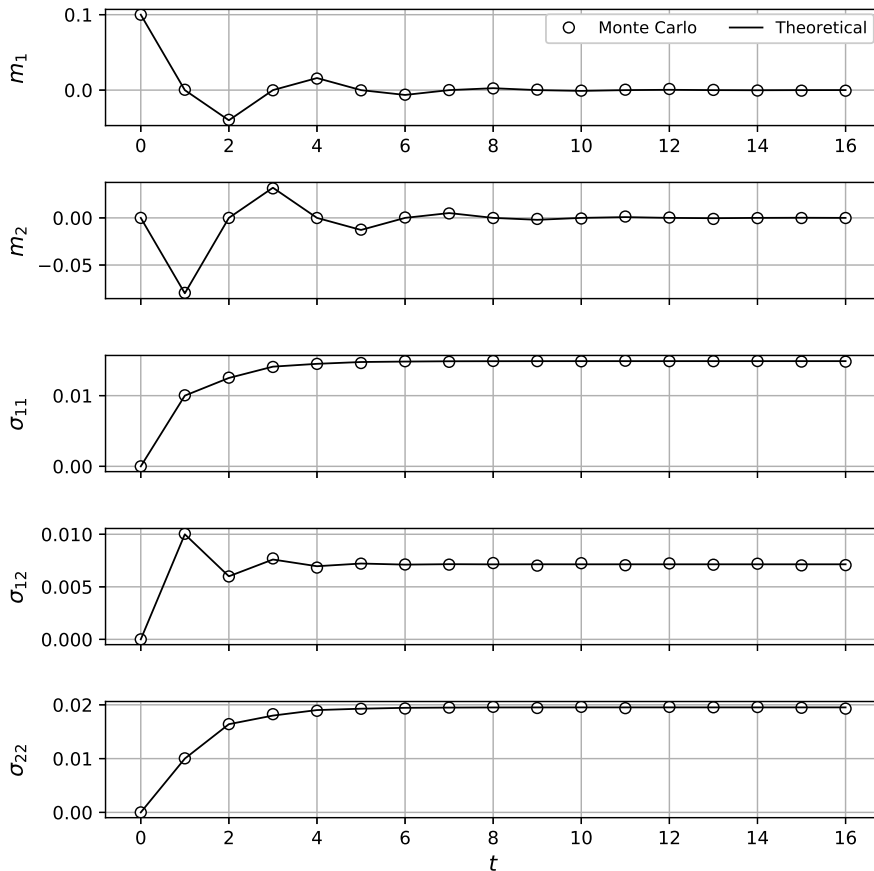


図 10.2 平均値と共分散行列の時間発展

具体的に両者の計算時間を比較すると，ある試行において，モンテカルロ法は 3.37 s，算法 10.5 による理論計算は 3.80×10^{-4} s であった．すなわち，同程度の結果を得るために，モンテカルロ法は 8887 倍の計算時間を要したことになる．もちろん，標本数を減らせば計算は速くなるが，理論値との乖離は大きくなる．

11 | カルマンフィルタ

11.1 線形確率システム

状態方程式 (10.16) p84 に，観測方程式を連立した

$$\begin{cases} \mathbf{x}_{t+1} = F_t \mathbf{x}_t + G_t \mathbf{w}_t & (\text{状態方程式}) \\ \mathbf{y}_t = H_t \mathbf{x}_t + \mathbf{v}_t & (\text{観測方程式}) \end{cases} \quad (t = 0, 1, 2, \dots) \quad (11.1)$$

を線形確率システムという．

\mathbf{x}_t	状態量	n 次元確率ベクトル
\mathbf{y}_t	観測量	p 次元確率ベクトル
\mathbf{w}_t	システム雑音	r 次元確率ベクトル
\mathbf{v}_t	観測雑音	p 次元確率ベクトル
F_t	状態推移行列	$n \times n$ 定数行列
G_t	駆動行列	$n \times r$ 定数行列
H_t	観測行列	$p \times n$ 定数行列

(1) 観測雑音に対する仮定 10.2.3 節の仮定に加えて，次の仮定をおく．

\mathbf{v}_t は p 次元ガウス白色雑音で，全ての時刻 t で，

$$\langle \mathbf{v}_t \rangle = \mathbf{0} \quad (11.2)$$

とする．白色雑音なので，デルタ相関 (異なる時刻の $\mathbf{v}_t, \mathbf{v}_s$ は無相関) を示す．

$$\langle \mathbf{v}_t \mathbf{v}_s^T \rangle = R_t \delta_{ts} \quad (11.3)$$

R_t は時刻 t の \mathbf{v}_t の共分散行列である (デルタ相関においても，同時刻の成分間の相関は否定されない) ．

また，どの時刻 t の雑音 \mathbf{v}_t も初期状態 \mathbf{x}_0 と独立とする．このとき $\langle \mathbf{v}_t \rangle = \mathbf{0}$ より

$$\langle \mathbf{x}_0 \mathbf{v}_t^T \rangle \stackrel{\text{独立}}{=} \langle \mathbf{x}_0 \rangle \langle \mathbf{v}_t^T \rangle = \bar{\mathbf{x}}_0 \mathbf{0}^T = \mathbf{0} \quad \forall t \geq 0 \quad (11.4)$$

となる[†]．すなわち， $x_0 \perp v_t$ が成立する．これと，観測方程式の形から， x_t は同時刻以降の v_s と独立であることが分かる．これより，

$$\langle x_t v_s^T \rangle = O \quad (t \leq s) \quad (11.5)$$

すなわち， $x_t \perp v_s$ ($t \leq s$) が成立する．

(2) 観測量と雑音の独立性 以上の結果，以下が帰結される．観測方程式 $y_t = H_t x_t + v_t$ より，観測量 y_t は未来の観測雑音 v_s ($t < s$) と独立である．このとき，

$$\langle y_t v_s^T \rangle \stackrel{\text{独立}}{=} \langle y_t \rangle \langle v_s^T \rangle = \langle y_t \rangle 0^T = O \quad (t < s) \quad (11.6)$$

すなわち， $y_t \perp v_s$ ($t < s$) が成立する．さらに，(10.25) p86 の $x_t \perp w_s$ ($t \leq s$) より，

$$\langle y_t w_s^T \rangle = O \quad (t \leq s) \quad (11.7)$$

すなわち， $y_t \perp w_s$ ($t \leq s$) が成立する．

11.2 状態推定問題

状態量 x_s は測定できず，時刻 t までの観測量 y_0, \dots, y_t のみが測定可能とする．この状況で，測定可能な y_0, \dots, y_t から状態の推定値 \hat{x}_s を得る問題を考える．

そのために，状態量 $x_s \in \mathcal{H}^n$ のモデル空間を，時刻 t までの観測量 $y_0, y_1, \dots, y_t \in \mathcal{H}^p$ で生成する．一般に $p \leq n$ とする．

$$Y^t := \left\{ a + \sum_{i=0}^t A_i y_i \mid a \in \mathbb{R}^n, A_i \in \mathbb{R}^{n \times p} \right\} \quad (11.8)$$

これをデータ空間と呼ぼう．

推定誤差 $\|x_s - \hat{x}_s\|$ を最小化する推定値 $\hat{x}_s \in Y^t$ は， x_s の Y^t 上への直交射影

$$\hat{x}_s = \text{proj}(x_s | Y^t) \stackrel{\text{定義}}{\iff} x_s - \hat{x}_s \perp Y^t \quad (11.9)$$

で与えられた．また， x_s, y_t はともにガウス型と仮定したから，推定値は， y_0, \dots, y_t を条件とする x_s の条件付き期待値に一致した．

$$\hat{x}_s = \text{proj}(x_s | Y^t) = \langle x_s | y_0, \dots, y_t \rangle \quad (11.10)$$

以下， $\langle x_s | y_0, \dots, y_t \rangle$ を略して $\langle x_s | \mathcal{Y}^t \rangle$ と書く．

このような \hat{x}_s の推定問題は， s, t の大小に応じて，次のように呼ばれる．

[†] “ \forall ” の使い方は，A.1.6 節を見よ．

$s < t$	$t = s$	$t < s$
平滑化 (smoothing)	濾波 (filtering)	予測 (prediction)

これらを区別するため，次のような表記を導入する．

- 予測推定値

$$\hat{\mathbf{x}}_{t+1/t} := \text{proj}(\mathbf{x}_{t+1}|Y^t) = \langle \mathbf{x}_{t+1} | \mathbf{y}_0, \dots, \mathbf{y}_t \rangle \quad (11.11)$$

$$\hat{\mathbf{x}}_{t/t-1} := \text{proj}(\mathbf{x}_t|Y^{t-1}) = \langle \mathbf{x}_t | \mathbf{y}_0, \dots, \mathbf{y}_{t-1} \rangle \quad (11.12)$$

- 濾波推定値

$$\hat{\mathbf{x}}_{t/t} := \text{proj}(\mathbf{x}_t|Y^t) = \langle \mathbf{x}_t | \mathbf{y}_0, \dots, \mathbf{y}_t \rangle \quad (11.13)$$

$$\hat{\mathbf{x}}_{t-1/t-1} := \text{proj}(\mathbf{x}_{t-1}|Y^{t-1}) = \langle \mathbf{x}_{t-1} | \mathbf{y}_0, \dots, \mathbf{y}_{t-1} \rangle \quad (11.14)$$

11.3 カルマンフィルタの導出

11.3.1 予測推定値 $\hat{\mathbf{x}}_{t+1/t}$

(11.1) の状態方程式の両辺について， \mathcal{Y}^t に関する条件付き期待値をとると，

$$\langle \mathbf{x}_{t+1} | \mathcal{Y}^t \rangle = F_t \langle \mathbf{x}_t | \mathcal{Y}^t \rangle + G_t \langle \mathbf{w}_t | \mathcal{Y}^t \rangle \quad (11.15)$$

となるが，(11.7) の $\mathbf{y}_t \perp \mathbf{w}_s$ ($t \leq s$) より，

$$\langle \mathbf{w}_t | \mathcal{Y}^t \rangle := \langle \mathbf{w}_t | \mathbf{y}_0, \dots, \mathbf{y}_t \rangle$$

となるが， \mathbf{w}_t と $\mathbf{y}_0, \dots, \mathbf{y}_t$ は独立なので，

$$= \langle \mathbf{w}_t \rangle = \mathbf{0} \quad (11.16)$$

となる．ゆえに，まず次の結果が得られる．

$$\hat{\mathbf{x}}_{t+1/t} = F_t \hat{\mathbf{x}}_{t/t} \quad (11.17)$$

続いて，条件付き共分散行列をとると，

$$\Sigma_{t+1/t} := \left\langle \mathbf{x}_{t+1} \mathbf{x}_{t+1}^T \middle| \mathbf{y}_0, \dots, \mathbf{y}_t \right\rangle - \hat{\mathbf{x}}_{t+1/t} \hat{\mathbf{x}}_{t+1/t}^T \quad (11.18)$$

となるが， $\mathbf{y}_0, \dots, \mathbf{y}_t$ に対して，これより未来の \mathbf{x}_{t+1} は独立なので，条件が外れて，

$$\Sigma_{t+1/t} := \langle \mathbf{x}_{t+1} \mathbf{x}_{t+1}^T \rangle - \hat{\mathbf{x}}_{t+1/t} \hat{\mathbf{x}}_{t+1/t}^T \quad (11.19)$$

となる．ゆえに，(10.31) p87 以降と同じ計算により，

$$\Sigma_{t+1/t} = F_t \Sigma_{t/t} F_t^T + G_t Q_t G_t^T \quad (11.20)$$

が得られる．以上をまとめると，

算法 11.1 (予測推定値) 予測推定値 $\hat{x}_{t+1/t}$ とその共分散行列 $\Sigma_{t+1/t}$ は，濾波推定値 $\hat{x}_{t/t}$ とその共分散行列 $\Sigma_{t/t}$ で次のように表せる．

$$\begin{cases} \hat{x}_{t+1/t} = F_t \hat{x}_{t/t} \\ \Sigma_{t+1/t} = F_t \Sigma_{t/t} F_t^T + G_t Q_t G_t^T \end{cases} \quad (t = 0, 1, 2, \dots) \quad (11.21)$$

11.3.2 観測誤差ベクトル

観測量 y_0, \dots, y_{t-1} を条件とする y_t の条件付き期待値

$$\hat{y}_{t/t-1} = \langle y_t | \mathcal{Y}^{t-1} \rangle := \langle y_t | y_0, \dots, y_{t-1} \rangle \quad (11.22)$$

と，その誤差ベクトル

$$\vec{y}_{t/t-1} := y_t - \hat{y}_{t/t-1} = y_t - \langle y_t | \mathcal{Y}^{t-1} \rangle \quad (11.23)$$

をとる．誤差ベクトルに (11.1) の観測方程式を代入すると，

$$\begin{aligned} \vec{y}_{t/t-1} &= H_t x_t + v_t - \langle H_t x_t + v_t | \mathcal{Y}^{t-1} \rangle \\ &= H_t x_t + v_t - H_t \langle x_t | \mathcal{Y}^{t-1} \rangle + \langle v_t | \mathcal{Y}^{t-1} \rangle \\ &= H_t (x_t - \hat{x}_{t/t-1}) + v_t \quad \because \langle v_t | \mathcal{Y}^{t-1} \rangle = \mathbf{0} \end{aligned} \quad (11.24)$$

$$= H_t \vec{x}_{t/t-1} + v_t, \quad \vec{x}_{t/t-1} := x_t - \hat{x}_{t/t-1} \quad (11.25)$$

となる．これより，以下の結果が導かれる．

算法 11.2 観測誤差ベクトル $\vec{y}_{t/t-1}$ は，ガウス白色雑音であり，次を満足する．

1. $\langle \vec{y}_{t/t-1} \rangle = \mathbf{0}$, $\langle \vec{y}_{t/t-1} \vec{y}_{s/s-1}^T \rangle = (R_t + H_t \Sigma_{t/t-1} H_t^T) \delta_{ts}$
2. $\vec{y}_{t/t-1} \perp Y^{t-1}$

► 証明 (1) は計算すれば明らか．(2) について， $x_t - \hat{x}_{t/t-1}$ は直交射影 $\hat{x}_{t/t-1} = \text{proj}(x_t | Y^{t-1})$ の誤差ベクトルなので， $x_t - \hat{x}_{t/t-1} \perp Y^{t-1}$ であり，すなわち，

$$\langle (x_t - \hat{x}_{t/t-1}) y^T \rangle = 0 \quad \forall y \in Y^{t-1}$$

となる．ゆえに，

$$\langle H_t(\mathbf{x}_t - \hat{\mathbf{x}}_{t/t-1})\mathbf{y}^T \rangle = H_t \langle (\mathbf{x}_t - \hat{\mathbf{x}}_{t/t-1})\mathbf{y}^T \rangle = O \quad \forall \mathbf{y} \in Y^{t-1}$$

となるので， $H_t(\mathbf{x}_t - \hat{\mathbf{x}}_{t/t-1}) \perp Y^{t-1}$ となる．また， $\mathbf{v}_t \perp Y^{t-1}$ は (11.6) p92 の $\mathbf{y}_t \perp \mathbf{v}_s$ ($t < s$) より明らか．

11.3.3 直和の構成

必要な集合演算を補充する．あるベクトル空間から，互いに素な部分空間 A, B :

$$A \cap B = \{\mathbf{0}\} \quad (11.26)$$

をとる (共通部分はゼロベクトルだけ)．それらの要素の和

$$\mathbf{z} = \mathbf{a} + \mathbf{b}, \quad \mathbf{a} \in A, \quad \mathbf{b} \in B \quad (11.27)$$

を，(ベクトルの) 直和という．その全体集合を

$$A \oplus B := \{\mathbf{a} + \mathbf{b} \mid \mathbf{a} \in A, \mathbf{b} \in B\} \quad (11.28)$$

と書き，(ベクトル空間の) 直和という．直和 $\mathbf{a} + \mathbf{b} \in A \oplus B$ が，別のベクトル \mathbf{c} と直交する条件は， $\mathbf{a} \perp \mathbf{c}$ かつ $\mathbf{b} \perp \mathbf{c}$ である (一方だけ直交させても和は直交するとは限らない)．したがって，部分空間の直和 $A \oplus B$ にベクトル \mathbf{c} が直交する条件は，

$$\mathbf{c} \perp A \oplus B \quad \begin{array}{c} \text{必要} \\ \hline \text{十分} \end{array} \quad \mathbf{c} \perp A \text{ かつ } \mathbf{c} \perp B \quad (11.29)$$

のように 2 つに分離する．その結果，直交射影も 2 つに分離して，

$$\text{proj}(\mathbf{x} | A \oplus B) = \text{proj}(\mathbf{x} | A) + \text{proj}(\mathbf{x} | B) \quad (11.30)$$

という公式が成立する．

もう 1 つ，集合 P, Q が等しいことを次のように定義する．

$$P = Q \quad \begin{array}{c} \text{定義} \\ \hline \end{array} \quad \text{任意の要素 } x \text{ について } x \in P \iff x \in Q \quad (11.31)$$

例えば， $P = \{a, b, c\}$ と $Q = \{b, a, c\}$ は等しい集合だが，確かに「 $x \in P \Rightarrow x \in Q$ 」と「 $x \in P \Leftarrow x \in Q$ 」が双方とも成立する．

本題に戻る．以上を応用すると，次の算法が得られる．

算法 11.3 (11.23) の観測誤差ベクトル $\vec{\mathbf{y}}_{t/t-1} := \mathbf{y}_t - \hat{\mathbf{y}}_{t/t-1}$ が生成するデータ空間

$$\vec{Y}^t = \left\{ B\vec{\mathbf{y}}_{t/t-1} \mid B \in \mathbb{R}^{n \times p} \right\} \quad (11.32)$$

について，次が成立する．

1. $Y^{t-1} \cap \vec{Y}^t = \{\emptyset\}$.
2. $Y^t = Y^{t-1} \oplus \vec{Y}^t$

▶▶ (集合のイコール) (2) の等号は集合の等号である．一般に，集合 A, B が $A = B$ であるとは， A の任意の要素 x が B に含まれ，なおかつ B の任意の要素 y が A に含まれることである．すなわち，

$$A = B \stackrel{\text{定義}}{\iff} (x \in A \implies x \in B) \text{ かつ } (y \in B \implies y \in A)$$

である．これが示せれば， A, B は同じ集合である．

▶ 証明 (1) の成立は， $B\vec{y}_{t/t-1}$ を張る y_t が， Y^{t-1} に含まれないので明らか．(2) の「 $z \in Y^t \implies z \in Y^{t-1} \oplus \vec{Y}^t$ 」を示す．任意の $z \in Y^t$ をとると， Y^t の定義 (11.8) より，

$$z = a + \sum_{i=1}^t A_i y_i = a + \sum_{i=1}^{t-1} A_i y_i + A_t y_t$$

と書ける．これに，(11.23) を移項した $y_t = \vec{y}_{t/t-1} + \hat{y}_{t/t-1}$ を代入する．条件付き期待値 $\hat{y}_{t/t-1}$ は y_0, \dots, y_{t-1} の線形関数なので，新たなパラメータ a', A'_i, B' によって，

$$z = a + \sum_{i=1}^t A_i y_i = \underbrace{a' + \sum_{i=1}^{t-1} A'_i y_i}_{\in Y^{t-1}} + \underbrace{B' \vec{y}_{t/t-1}}_{\in \vec{Y}^t}$$

の形式に整理できる．ゆえに，この z は直和 $Y^{t-1} \oplus \vec{Y}^t$ の要素である．次に「 $z \in Y^{t-1} \oplus \vec{Y}^t \implies z \in Y^t$ 」を示す．任意の $z \in Y^{t-1} \oplus \vec{Y}^t$ をとる．これは，

$$z = a + \sum_{i=1}^{t-1} A_i y_i + B \vec{y}_{t/t-1}$$

の形式で書ける．これに， $\vec{y}_{t/t-1} := y_t - \hat{y}_{t/t-1}$ を代入すると，新たな a'', A''_i により，

$$z = a'' + \sum_{i=1}^t A''_i y_i$$

の形式で書けるので， z は Y^t の要素である．以上， $z \in Y^t \iff z \in Y^{t-1} \oplus \vec{Y}^t$ の成立が示された．したがって， Y^t と $Y^{t-1} \oplus \vec{Y}^t$ は同じ集合である．

11.3.4 濾波推定値 $\hat{x}_{t/t}$

時刻 t における濾波推定値は，算法 11.3 (2) より，直交射影で，

$$\hat{x}_{t/t} = \text{proj}(x_t | Y^t) = \text{proj}(x_t | Y^{t-1} \oplus \vec{Y}^t) \quad (11.33)$$

と書ける．ゆえに，直交射影の公式 (11.30) により，

$$\hat{x}_{t/t} = \text{proj}(x_t | Y^{t-1}) + \text{proj}(x_t | \vec{Y}^t) = \hat{x}_{t/t-1} + \langle x_t | \vec{y}_{t/t-1} \rangle \quad (11.34)$$

という結果が得られる．条件付き期待値 $\langle x_t | \vec{y}_{t/t-1} \rangle$ は， $\vec{y}_{t/t-1}$ の線形関数なので，適当な行列 B によって，形式的に次のように表示できる．

$$\hat{x}_{t/t} = \hat{x}_{t/t-1} + B \vec{y}_{t/t-1} \quad (11.35)$$

問題は B の定め方だが，次節の方法で定めた $B = K_t$ を，カルマンゲインという．

11.3.5 カルマンゲイン K_t の導出

関係式 (11.35) を $B = K_t$ で改めて，

$$\hat{x}_{t/t} = \hat{x}_{t/t-1} + K_t \vec{y}_{t/t-1} \quad (11.36)$$

と書き，濾波推定値の誤差ベクトル

$$\vec{x}_{t/t} := x_t - \hat{x}_{t/t} \quad (11.37)$$

をとる．以下，誤差 $\|\vec{x}_{t/t}\|$ を最小化する K_t を求める．

そのための条件は $\vec{x}_{t/t} \perp Y^t$ であった．ただし，算法 11.3 (2) より $Y^t = Y^{t-1} \oplus \vec{Y}^t$ なので，これは (11.29) より，次の条件と等価である．

$$\vec{x}_{t/t} \perp Y^{t-1} \quad \text{かつ} \quad \vec{x}_{t/t} \perp \vec{Y}^t \quad (11.38)$$

まず，直ちに $\vec{x}_{t/t} \perp Y^{t-1}$ が示される．実際，(11.37) に (11.36) を代入すると，

$$\vec{x}_{t/t} := \vec{x}_{t/t-1} - K_t \vec{y}_{t/t-1}, \quad \vec{x}_{t/t-1} := x_t - \hat{x}_{t/t-1} \quad (11.39)$$

となる．予測推定誤差 $\vec{x}_{t/t-1}$ は直交射影 $\hat{x}_{t/t-1} = \text{proj}(x_t | Y^{t-1})$ の誤差ベクトルなので， $\vec{x}_{t/t-1} \perp Y^{t-1}$ が成立する．また，算法 11.2 より $\vec{y}_{t/t-1} \perp Y^{t-1}$ であるから，

$$\vec{x}_{t/t} := \vec{x}_{t/t-1} - K_t \vec{y}_{t/t-1} \perp Y^{t-1} \quad (11.40)$$

が示された．

次に，残りの条件 $\vec{x}_{t/t} \perp \vec{Y}^t$ ，すなわち $\langle \vec{x}_{t/t} \vec{y}_{t/t-1}^T \rangle = 0$ を変形していく．(11.40) を代入すると， $0 = \langle \vec{x}_{t/t} \vec{y}_{t/t-1}^T \rangle = \langle (\vec{x}_{t/t-1} - K_t \vec{y}_{t/t-1}) \vec{y}_{t/t-1}^T \rangle = \langle \vec{x}_{t/t-1} \vec{y}_{t/t-1}^T \rangle - K_t \langle \vec{y}_{t/t-1} \vec{y}_{t/t-1}^T \rangle$ となるので， K_t は，予測推定誤差 $\vec{x}_{t/t-1}$ と出力誤差 $\vec{y}_{t/t-1}$ の分散・共分散で書き下せることが分かる．

$$K_t = \langle \vec{x}_{t/t-1} \vec{y}_{t/t-1}^T \rangle \langle \vec{y}_{t/t-1} \vec{y}_{t/t-1}^T \rangle^{-1} \quad (11.41)$$

この右辺をさらに整理する．まず， $\langle \vec{x}_{t/t-1} \vec{y}_{t/t-1}^T \rangle$ は，(11.25) p94 を代入すると，

$$\left\langle \vec{x}_{t/t-1} \vec{y}_{t/t-1}^T \right\rangle = \left\langle \vec{x}_{t/t-1} \vec{x}_{t/t-1}^T \right\rangle H_t^T + \underbrace{\left\langle \vec{x}_{t/t-1} v_t^T \right\rangle}_{=0: (11.5)} = \Sigma_{t/t-1} H_t^T \quad (11.42)$$

となる．残りの $\left\langle \vec{y}_{t/t-1} \vec{y}_{t/t-1}^T \right\rangle$ は算法 11.2 より，

$$\left\langle \vec{y}_{t/t-1} \vec{y}_{t/t-1}^T \right\rangle = R_t + H_t \Sigma_{t/t-1} H_t^T \quad (11.43)$$

である．以上， K_t は次のように書き下せることが分かった．

$$\begin{aligned} K_t &= \left\langle \vec{x}_{t/t-1} \vec{y}_{t/t-1}^T \right\rangle \left\langle \vec{y}_{t/t-1} \vec{y}_{t/t-1}^T \right\rangle^{-1} \\ &= \Sigma_{t/t-1} H_t^T (R_t + H_t \Sigma_{t/t-1} H_t^T)^{-1} \end{aligned} \quad (11.44)$$

この K_t をカルマンゲインという．

最後に，共分散の推移則を整理する．(11.39) を移項した

$$\vec{x}_{t/t-1} = \vec{x}_{t/t} + K_t \vec{y}_{t/t-1} \quad (11.45)$$

の両辺の共分散をとると， $\vec{x}_{t/t} \perp \vec{y}_{t/t-1}$ ，すなわち $\left\langle \vec{x}_{t/t} \vec{y}_{t/t-1}^T \right\rangle = O$ に注意して，

$$\begin{aligned} \Sigma_{t/t-1} &:= \left\langle \vec{x}_{t/t} \vec{x}_{t/t}^T \right\rangle \\ &= \left\langle (\vec{x}_{t/t} + K_t \vec{y}_{t/t-1})(\vec{x}_{t/t} + K_t \vec{y}_{t/t-1})^T \right\rangle \\ &= \left\langle \vec{x}_{t/t} \vec{x}_{t/t}^T \right\rangle + K_t \left\langle \vec{y}_{t/t-1} \vec{y}_{t/t-1}^T \right\rangle K_t^T \\ &\quad + \underbrace{K_t \left\langle \vec{y}_{t/t-1} \vec{x}_{t/t}^T \right\rangle + \left\langle \vec{x}_{t/t} \vec{y}_{t/t-1}^T \right\rangle K_t^T}_{=O} \\ &= \Sigma_{t/t} + K_t \left\langle \vec{y}_{t/t-1} \vec{y}_{t/t-1}^T \right\rangle K_t^T \\ &= \Sigma_{t/t} + K_t (R_t + H_t \Sigma_{t/t-1} H_t^T) K_t^T \quad \because \text{算法 11.2 p94 (1)} \\ &= \Sigma_{t/t} + K_t (R_t + H_t \Sigma_{t/t-1} H_t^T) K_t^T \end{aligned} \quad (11.46)$$

となるが，共分散行列の対称性 $R_t^T = R_t$ ， $\Sigma^T = \Sigma$ より，

$$\begin{aligned} K_t^T &= \left\{ (R_t + H_t \Sigma_{t/t-1} H_t^T)^{-1} \right\}^T H_t \Sigma_{t/t-1}^T \\ &= (R_t^T + H_t \Sigma_{t/t-1}^T H_t^T)^{-1} H_t \Sigma_{t/t-1}^T \\ &= (R_t + H_t \Sigma_{t/t-1} H_t^T)^{-1} H_t \Sigma_{t/t-1} \end{aligned}$$

となるので，

$$\Sigma_{t/t-1} = \Sigma_{t/t} + K_t H_t \Sigma_{t/t-1} \quad (11.47)$$

が得られる．以上をまとめると，

算法 11.4 (濾波推定値) 線形確率システムの状態量 x_t の濾波推定値 $\hat{x}_{t/t}$ は，

$$\begin{aligned} \hat{x}_{t/t} &= \hat{x}_{t/t-1} + K_t \vec{y}_{t/t-1} \\ &= \hat{x}_{t/t-1} + K_t (y_t - H_t \hat{x}_{t/t-1}) && \text{【濾波推定値】} \\ K_t &= \Sigma_{t/t-1} H_t^T (H_t \Sigma_{t/t-1} H_t^T + R_t)^{-1} && \text{【カルマンゲイン】} \\ \Sigma_{t/t} &= (I - K_t H_t) \Sigma_{t/t-1} && \text{【推定誤差共分散】} \end{aligned}$$

で得られる．ただし I は単位行列である．

11.4 カルマンフィルタ

11.4.1 基本アルゴリズム

算法 11.1 と算法 11.4 を組み合わせると，図 11.1 のような逐次推定が実現できる．

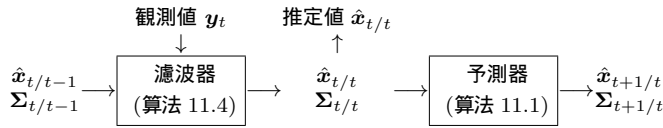


図 11.1 カルマンフィルタによる逐次推定

すなわち，前時刻の理論予測値 $\hat{x}_{t/t-1}$ を，実際の観測値 y_t で修正することで，現時刻の濾波推定値 $\hat{x}_{t/t}$ を得る．これを初期値として次時刻の理論予測値 $\hat{x}_{t+1/t}$ を計算する．以上を繰り返す．このような逐次推定アルゴリズムを，カルマンフィルタという．

具体的な手順は次のように整理できる．

算法 11.5 (カルマンフィルタ 1) 線形確率システム：

$$\begin{cases} x_{t+1} = F_t x_t + G_t w_t, & x_t \in \mathbb{R}^n, w_t \in \mathbb{R}^r \\ y_t = H_t x_t + v_t, & y_t, v_t \in \mathbb{R}^p, t = 0, 1, 2, \dots \end{cases}$$

の最小分散濾波推定値 $\hat{x}_{t/t}$ は，次のように逐次計算できる^{†2}．

(0) 初期条件：

$$\hat{\mathbf{x}}_{0/-1} = \bar{\mathbf{x}}_0, \quad \Sigma_{0/-1} = \Sigma_0 \quad (11.48)$$

(1) 濾波推定： $\mathbf{y}_t, \hat{\mathbf{x}}_{t/t-1}, \Sigma_{t/t-1}$ を既知として，

$$K_t = \Sigma_{t/t-1} H_t^T (H_t \Sigma_{t/t-1} H_t^T + R_t)^{-1} \text{【カルマンゲイン】} \quad (11.49)$$

$$\hat{\mathbf{x}}_{t/t} = \hat{\mathbf{x}}_{t/t-1} + K_t (\mathbf{y}_t - H_t \hat{\mathbf{x}}_{t/t-1}) \quad \text{【濾波推定値】} \quad (11.50)$$

$$\Sigma_{t/t} = (I - K_t H_t) \Sigma_{t/t-1} \quad \text{【濾波推定誤差】} \quad (11.51)$$

ただし，逆行列が存在しない場合は，疑似逆行列を用いる．

(2) 予測推定：

$$\hat{\mathbf{x}}_{t+1/t} = F_t \hat{\mathbf{x}}_{t/t} \quad \text{【予測推定値】} \quad (11.52)$$

$$\Sigma_{t+1/t} = F_t \Sigma_{t/t} F_t^T + G_t Q_t G_t^T \quad \text{【予測推定誤差】} \quad (11.53)$$

11.4.2 離散時間リカッチ方程式

算法 11.5 の式を集約して整理することを考える．まず，濾波推定値 (11.50) を予測推定値 (11.52) に代入すると，

$$\begin{aligned} \hat{\mathbf{x}}_{t+1/t} &= F_t \hat{\mathbf{x}}_{t/t} = F_t \hat{\mathbf{x}}_{t/t-1} + F_t K_t (\mathbf{y}_t - H_t \hat{\mathbf{x}}_{t/t-1}) \\ &= F_t \hat{\mathbf{x}}_{t/t-1} + F_t \Sigma_{t/t-1} H_t^T (H_t \Sigma_{t/t-1} H_t^T + R_t)^{-1} (\mathbf{y}_t - H_t \hat{\mathbf{x}}_{t/t-1}) \end{aligned} \quad (11.54)$$

となるが，予測推定値 $\hat{\mathbf{x}}_{t/t-1}$ とその誤差共分散 $\Sigma_{t/t-1}$ を改めて，

$$\begin{cases} \hat{\mathbf{x}}_t := \hat{\mathbf{x}}_{t/t-1} \\ \Sigma_t := \Sigma_{t/t-1} \end{cases} \quad (11.55)$$

と表記すると，

$$\hat{\mathbf{x}}_{t+1} = F_t \hat{\mathbf{x}}_t + \underbrace{F_t \Sigma_t H_t^T (H_t \Sigma_t H_t^T + R_t)^{-1}}_{K_t} (\mathbf{y}_t - H_t \hat{\mathbf{x}}_t) \quad (11.56)$$

という予測推定式が得られる．同様に，濾波推定誤差 (11.51) を予測推定誤差 (11.53) に代入すると，

^{†2} 片山¹⁾ p90 など．

$$\begin{aligned}
\Sigma_{t+1} &= F_t \left\{ (I - K_t H_t) \Sigma_t \right\} F_t^T + G_t Q_t G_t^T \\
&= F_t \Sigma_t F_t^T + G_t Q_t G_t^T - F_t K_t H_t \Sigma_t F_t^T
\end{aligned} \tag{11.57}$$

を得る．この式を，離散時間リカッチ方程式[†]という．以上をまとめると，

算法 11.6 (カルマンフィルタ 2) 線形確率システム：

$$\begin{cases} x_{t+1} = F_t x_t + G_t w_t, & x_t \in \mathbb{R}^n, w_t \in \mathbb{R}^r \\ y_t = H_t x_t + v_t, & y_t, v_t \in \mathbb{R}^p, t = 0, 1, 2, \dots \end{cases} \tag{11.58}$$

の最小分散濾波推定値 $\hat{x}_{t/t}$ は，次のように逐次計算できる．

(0) 初期条件：

$$\hat{x}_0 = \bar{x}_0, \quad \Sigma_0 = \Sigma_0 \tag{11.59}$$

(1) 濾波推定： y_t, \hat{x}_t, Σ_t を既知として，

$$K_t = \Sigma_t H_t^T (H_t \Sigma_t H_t^T + R_t)^{-1} \text{【カルマンゲイン】} \tag{11.60}$$

$$\hat{x}_{t/t} = \hat{x}_t + K_t (y_t - H_t \hat{x}_t) \quad \text{【濾波推定値】} \tag{11.61}$$

ただし，逆行列が存在しない場合は，疑似逆行列を用いる．

(2) 予測推定：

$$\hat{x}_{t+1} = F_t \hat{x}_{t/t} = F_t \hat{x}_t + F_t K_t (y_t - H_t \hat{x}_t) \quad \text{【予測推定値】} \tag{11.62}$$

$$\Sigma_{t+1} = F_t \Sigma_t F_t^T + G_t Q_t G_t^T - F_t K_t H_t \Sigma_t F_t^T \quad \text{【予測推定誤差】} \tag{11.63}$$

11.5 定常カルマンフィルタ

カルマンフィルタの係数行列 F_t, G_t, H_t, Q_t, R_t が全て，定行列 F, G, H, Q, R である場合を考える．このとき， Σ_t の時間発展式 (11.63) は，定係数の漸化式となる．それが安定なら，定常値 $\Sigma = \lim_{t \rightarrow \infty} \Sigma_t$ が存在し，カルマンゲインも定常値 $K = \lim_{t \rightarrow \infty} K_t$ をもつ．

11.5.1 基本アルゴリズム

この条件下では，幸運なことが起きる．定常値 Σ, K を (11.63) と (11.60) に代入すると，

[†] 片山¹⁾p92 など．

$$\Sigma = F\Sigma F^T + GQG^T - FKH\Sigma F^T, \quad K = \Sigma H^T (H\Sigma H^T + R)^{-1} \quad (11.64)$$

という連立代数方程式を得る．第 1 式を離散時間代数リカッチ方程式という．(11.64) の未知数は K, Σ だけなので，なんと，カルマンゲイン K は (11.64) だけで求まる．この K を使えば，濾波推定値 $\hat{x}_{t/t}$ と予測推定値 \hat{x}_{t+1} は，次式で更新できる．

$$\hat{x}_{t/t} = \hat{x}_t + K(y_t - H\hat{x}_t), \quad \hat{x}_{t+1} = F\hat{x}_{t/t} \quad (11.65)$$

以上，カルマンフィルタを定係数としたところ，算法 11.6 が，(11.64)，(11.65) に集約されてしまった．これを，定常カルマンフィルタという[†]．次のように整理できる．

算法 11.7 (定常カルマンフィルタ) 定係数の線形確率システム：

$$\begin{cases} x_{t+1} = Fx_t + Gw_t, & x_t \in \mathbb{R}^n, w_t \in \mathbb{R}^r \\ y_t = Hx_t + v_t, & y_t, v_t \in \mathbb{R}^p, t = 0, 1, 2, \dots \end{cases} \quad (11.66)$$

の最小分散濾波推定値 $\hat{x}_{t/t}$ は，次のように逐次計算できる．

(0) 前処理と初期条件：次の代数方程式から Σ と K を事前に求めておく．

$$\Sigma = F\Sigma F^T + GQG^T - FKH\Sigma F^T \quad (\text{離散時間代数リカッチ方程式}) \quad (11.67)$$

$$K = \Sigma H^T (H\Sigma H^T + R)^{-1} \quad (\text{定常カルマンゲイン}) \quad (11.68)$$

また，予測推定値の初期値 \hat{x}_0 を定めておく．

(1) 濾波推定値：

$$\hat{x}_{t/t} = \hat{x}_t + K(y_t - H\hat{x}_t) \quad (11.69)$$

(2) 予測推定値：

$$\hat{x}_{t+1} = F\hat{x}_{t/t} \quad (11.70)$$

定常カルマンフィルタの大きな利点として，カルマンゲイン K が前処理で決まってしまうので，状態推定器には，(11.69) と (11.70) だけを実装すればよい．あとは，代数リカッチ方程式 (11.67) を具体的にどう解くかが問題になるが，これは簡単である．ほとんどの数値解析言語は，代数リカッチ方程式の解法を備えているので，それを利用すればよい．ただし，カルマンフィルタ用のインターフェースになっていないことが多いので，適宜，行列の読み替えが必要になるかもしれない．

[†] 片山¹⁾ p108 など．

▶▶ (読み替えのヒント) 例えば, ある言語の離散時間代数リカッチ方程式ソルバが,

$$X = \bar{A}^T X \bar{A} - (\bar{A}^T X \bar{B})(\bar{R} + \bar{B}^T X \bar{B})^{-1}(\bar{B}^T X \bar{A}) + Q$$

の X を求める仕様だった場合, (11.67) を解くには $\bar{A} \mapsto F^T, \bar{B} \mapsto H^T, \bar{Q} \mapsto G^T Q G, \bar{R} \mapsto R$ という読み替えが必要になる.

11.5.2 定常カルマンフィルタの安定性

定常カルマンフィルタの推定値 \hat{x}_t の安定性を調べるため, 誤差ベクトル $\vec{x}_t := \hat{x}_t - x_t$ を復活させると, $y_t - H\hat{x}_t = Hx_t + v_t - H\hat{x}_t = H(\hat{x}_t - \vec{x}_t) + v_t - H\hat{x}_t = v_t - H\vec{x}_t$ より, 予測推定値 (11.62) は,

$$\hat{x}_{t+1} = F\hat{x}_t + FK(y_t - H\hat{x}_t) = F\hat{x}_t - FKH\vec{x}_t + FKv_t \quad (11.71)$$

となる. これと (11.58) の第 1 式より,

$$\begin{aligned} \vec{x}_{t+1} &= \hat{x}_{t+1} - x_{t+1} = F\hat{x}_t - FK_t H_t \vec{x}_t + FKv_t - Fx_t - Gw_t \\ &= F\vec{x}_t - FKH\vec{x}_t + FKv_t - Gw_t \\ \therefore \vec{x}_{t+1} &= F(I - KH)\vec{x}_t + S_t, \quad S_t := FKv_t - Gw_t \end{aligned} \quad (11.72)$$

が得られる. I は単位行列である. この漸化式 (離散力学系) が安定ならば, 推定誤差 \vec{x}_t は安定となる. すなわち, 次の安定条件が得られる[†].

算法 11.8 (定常カルマンフィルタの安定条件) 行列 $F(I - KH)$ の固有値の絶対値 (複素数の意味での) が, 全て 1 未満のとき, 推定誤差 \vec{x}_t は漸近安定となる.

当然のことながら, これを満たさない条件では, 推定誤差が収束しないのでカルマンフィルタは使用できない.

11.6 数 値 例

1 自由度振動系の状態推定の例を示す. ランダムな線形 1 自由度振動系とその観測問題を考える.

$$\frac{d}{dt} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -k & -c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} w(t) \quad (11.73)$$

[†] 片山¹⁾ p110 など.

$$y = (0, 1) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + v(t) = x_2 + v(t) \quad (11.74)$$

$w(t)$, $v(t)$ はそれぞれ $N(0, Q)$, $N(0, R)$ のガウス白色雑音とする．すなわち，この系は不規則外乱 $w(t)$ を受けており，雑音 $v(t)$ に乱された速度 x_2 しか観測できない．

以上で導出した離散時間カルマンフィルタを用いるため，確率システム用のオイラー法（算法 12.1）で離散化する．雑音項の時間増分に平方根がつくので注意せよ．

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{n+1} - \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_n = \Delta t \begin{pmatrix} 0 & 1 \\ -k & -c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_n + \sqrt{\Delta t} \begin{pmatrix} 0 \\ 1 \end{pmatrix} w_n \quad (11.75)$$

これを整理して，線形離散時間確率システムの一般形で表示すると，次のようになる．

$$\begin{aligned} \mathbf{x}_{n+1} &= \underbrace{\begin{pmatrix} 1 & \Delta t \\ -k\Delta t & 1 - c\Delta t \end{pmatrix}}_F \mathbf{x}_n + \underbrace{\begin{pmatrix} 0 \\ \sqrt{\Delta t} \end{pmatrix}}_G w_n \\ y_n &= \underbrace{(0, 1)}_H \mathbf{x}_n + v_n = (x_2)_n + v_n \end{aligned} \quad (11.76)$$

図 11.2 に数値例を示す．数値例の作成には，本書レポジトリ (iv 頁) の Ch11.ipynb 用いた．

上段の図は，観測値（黒色）と真値（灰色）の時間波形を表す．下段の左は，カルマンフィルタによる推定結果（黒色）と真値（灰色）を表す．下段の右は，同じく，観測値の数値積分による推定結果（黒色）と真値（灰色）を表す．

▶▶（数値積分による推定？） 一般に，測定値が速度のとき，これを数値的に積分して変位の推定値とすることがある．

数値積分による推定結果（下段の右）では，積分により y_n の雑音成分が累積するため，推定値（黒色）が真値（灰色）から大きくかけ離れている．このように，速度を積分して変位を推定する方法は，雑音に弱い．

▶▶（ジャイロセンサのドリフト） 以上の問題は，ジャイロセンサが検出する角速度から角変位を求めるときにも生じる．例えば，角変位が 0 のときから計測を始めて，数値積分で角変位を推定していくと，角変位の真値が 0 に戻っても，推定値は 0 に戻らない．この不具合をドリフトという．その解消のために，カルマンフィルタが実用されている．

これに対して，下段の左がカルマンフィルタによる推定結果（黒線）である．真値（灰色）に追従した推定値が得られている．

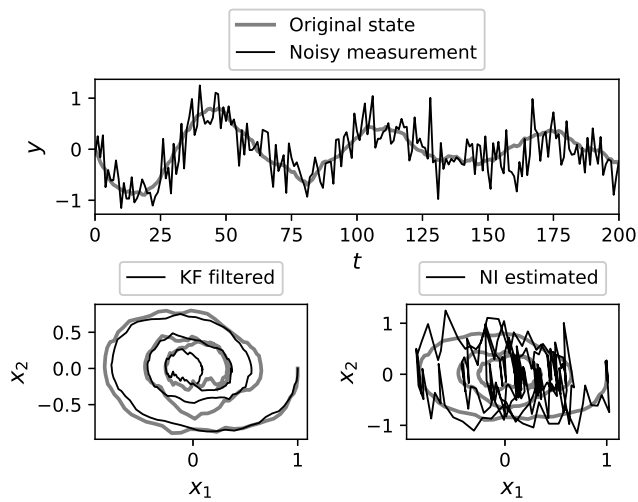


図 11.2 カルマンフィルタによる状態推定 (下段左) と数値積分による変位推定 (下段右) との比較 ($c = 0.3$, $k = 1$, $\Delta t = 0.1$, $Q = 0.01$, $R = 0.1$) . 上段の黒線は元にした観測量 y_n を表す .

12

連続時間カルマンフィルタ

カルマンフィルタを、フィードバック制御に応用したい (13 章) . そのための準備として、前章で導出した離散時間の結果を、連続時間に拡張する . 紙面も尽きてきたので、厳密性は犠牲にして、少々乱暴に導出する .

12.1 連続時間系の離散時間近似

12.1.1 連続時間の状態方程式

まず、時間 t を実数とする次の状態方程式を考える .

$$\dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) + D(t)\mathbf{w}(t), \quad \mathbf{x}(t) \in \mathbb{R}^n, \mathbf{w}(t) \in \mathbb{R}^r \quad (12.1)$$

$\mathbf{w}(t)$ は連続時間のガウス白色雑音で、次の性質を示す (ものとして定義される) .

$$\langle \mathbf{w}(t) \rangle = \mathbf{0}, \quad \langle \mathbf{w}(t)\mathbf{w}(s)^T \rangle = Q(t)\delta(t-s) \quad (s \geq t) \quad (12.2)$$

離散時間の場合と同様に、この性質をデルタ相関という . $\delta(t)$ はディラックのデルタ関数である . $\delta(t)$ は、面積 1 のインパルスで、面積を変えずに幅 $\rightarrow 0$ としたもので、数学的には次のように定義される .

$$\delta(t) = 0 \quad (t \neq 0), \quad \int_{-\infty}^{\infty} f(t)\delta(t-a)dt = f(a) \quad (12.3)$$

デルタ相関の仮定により、白色雑音の成分 $w_i(t)$ は、無限小時間だけ過去の自分 $w_i(t-dt)$ や未来の自分 $w_i(t+dt)$ と無相関になる . そうなる $w_i(t)$ は、いたるところ不連続な時間関数になる . すなわち、 $w_i(t)$ のグラフは極度にギザギザしていて、時間方向にいくら拡大して見ても、ギザギザは消えない .

12.1.2 離散時間近似

時間 t を、サンプリング時間 Δt で離散化する .

$$t_k = t_0 + k\Delta t \quad (k = 0, \dots, n) \quad (12.4)$$

各時刻 t_k の状態量と雑音を,

$$\mathbf{x}_k := \mathbf{x}(t_k), \quad \mathbf{w}_k := \mathbf{w}(t_k) \quad (12.5)$$

と書く．このとき， Δt が十分に小さければ，微分の定義より，

$$\dot{\mathbf{x}}(t_k) \approx \frac{\mathbf{x}(t_k + \Delta t) - \mathbf{x}(t_k)}{\Delta t} = \frac{\mathbf{x}_{k+1} - \mathbf{x}_k}{\Delta t} \quad (12.6)$$

という近似が成立する．状態方程式 (12.1) に代入して分母を払うと，

$$\mathbf{x}_{k+1} = \mathbf{x}_k + A(t_k)\mathbf{x}_k\Delta t + D(t_k)\mathbf{w}_k\Delta t \quad (12.7)$$

が得られる． I は単位行列である．仮に， $\mathbf{w}(t)$ が連続な時間関数であれば，これが状態方程式の最も単純な離散化を与える（オイラー法）．

しかしながら， $\mathbf{w}(t)$ が白色雑音の場合，雑音項 $D(t_k)\mathbf{w}_k\Delta t$ の強度が Δt に依存してしまう．なぜなら，例えば Δt を大きくすると， $\mathbf{w}(t)$ の細かいギザギザ（高周波成分）が (12.7) に伝わらなくなる．ゆえに，離散化した白色雑音 $\mathbf{w}_k := \mathbf{w}(t_k)$ のパワーは，もとの $\mathbf{w}(t)$ より小さくなる．

両者の関係を導く．簡単のため， $w(t)$ がスカラ値の場合を考える．前節で述べたように，連続時間の白色雑音 $w(t)$ の強度は，

$$\langle w(t)w(s) \rangle = Q(t)\delta(t-s) \quad (12.8)$$

で表せた ($Q(t)$ もスカラとする)．このグラフの面積 S は， $\delta(t-s)$ の面積が 1 であることから， $S = Q(t)$ である．一方で，離散化した $w_k := w(t_k)$ の強度を，

$$\langle w_k w_l \rangle = \bar{Q}_k \delta_{kl} \quad (12.9)$$

とする．このとき， $w(t_k)$ のグラフの面積 \bar{S} は，サンプリング時間 Δt に対して， $\bar{S} = \Delta t \bar{Q}_k$ となる．したがって， w_k に $w(t)$ と同等の強度を持たせるには，

$$Q(t) = S = \bar{S} = \Delta t \bar{Q}_k \quad (12.10)$$

となるような \bar{Q}_k を選べばよい．この \bar{Q}_k を用いると，連続時間系

$$\dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) + D(t)\mathbf{w}(t), \quad \langle \mathbf{w}(t)\mathbf{w}(s)^T \rangle = Q(t)\delta(t-s) \quad (12.11)$$

は，次の離散時間系で近似されることが分かる．

$$\mathbf{x}_{k+1} = \mathbf{x}_k + A(t_k)\mathbf{x}_k\Delta t + D(t_k)\mathbf{w}_k\Delta t, \quad \langle \mathbf{w}_k \mathbf{w}_l^T \rangle = \bar{Q}_k \delta_{kl} = \frac{Q(t_k)}{\Delta t} \delta_{kl} \quad (12.12)$$

このままでは見にくいので，先人にならって $\mathbf{w}'_k := \mathbf{w}_k \sqrt{\Delta t}$ を (12.12) に代入すると，

$$\begin{aligned}
\mathbf{x}_{k+1} &= \mathbf{x}_k + A(t_k)\mathbf{x}_k\Delta t + D(t_k)\mathbf{w}'_k\sqrt{\Delta t} \\
&= \{I + \Delta t A(t_k)\}\mathbf{x}_k + \{\sqrt{\Delta t}D(t_k)\}\mathbf{w}'_k, \\
\langle \mathbf{w}'_k \mathbf{w}'_l{}^T \rangle &= \Delta t \langle \mathbf{w}_k \mathbf{w}_l{}^T \rangle = Q(t_k)\delta_{kl}
\end{aligned} \tag{12.13}$$

を得る． \mathbf{w}' を改めて \mathbf{w} と書いて整理すると，確率論分野ではよく知られた[†]，次の算法 12.1 が得られる．(12.15) は，通常の常微分方程式のオイラー法によく似ているが，雑音項の時間ステップが $\sqrt{\Delta t}$ に書き換えられている．

算法 12.1 (離散時間近似) 雑音を受ける連続時間システムの

$$\dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) + D(t)\mathbf{w}(t), \quad \langle \mathbf{w}(t)\mathbf{w}(s)^T \rangle = Q(t)\delta(t-s) \tag{12.14}$$

の解過程は， $\mathbf{x}_k := \mathbf{x}(t_k)$ ， $\mathbf{w}_k := \mathbf{w}(t_k)$ ， $t_k := t_0 + k\Delta t$ ， $k = 0, 1, 2, \dots$ に対して，

$$\mathbf{x}_{k+1} = \mathbf{x}_k + A(t_k)\mathbf{x}_k\Delta t + D(t_k)\mathbf{w}_k\sqrt{\Delta t}, \quad \langle \mathbf{w}_k \mathbf{w}_l{}^T \rangle = Q(t_k)\delta_{kl} \tag{12.15}$$

もしくは，

$$\begin{cases} \mathbf{x}_{k+1} = F_k\mathbf{x}_k + G_k\mathbf{w}_k, & \langle \mathbf{w}_k \mathbf{w}_l{}^T \rangle = Q_k\delta_{kl}, \\ F_k = I + \Delta t A(t_k), & G_k = \sqrt{\Delta t}D(t_k), \quad Q_k = Q(t_k) \end{cases} \tag{12.16}$$

で近似できる．

例題 12.1 (期待値と共分散行列の推移則 (連続時間)) 離散時間の算法 10.5 に，(12.16) を代入し， $\Delta t \rightarrow \infty$ の極限をとると，連続時間の公式，

$$\dot{\bar{\mathbf{x}}} = A(t)\bar{\mathbf{x}}(t), \tag{12.17}$$

$$\dot{\Sigma}(t) = A(t)\Sigma(t) + \Sigma(t)A(t)^T + D(t)Q(t)D(t)^T \tag{12.18}$$

が得られる．これを示せ．

12.1.3 出力方程式の離散時間近似

連続時間の出力方程式，

$$\mathbf{y}(t) = C(t)\mathbf{x}(t) + \mathbf{v}(t), \quad \langle \mathbf{v}(t)\mathbf{v}(s)^T \rangle = R(t)\delta(t-s) \tag{12.19}$$

[†] 砂原⁽⁶⁾p89 など．

を考える．これを時間で離散化

$$\mathbf{y}_k := \mathbf{y}(t_k), \quad \mathbf{v}_k := \mathbf{v}(t_k), \quad t_k = t_0 + k\Delta t \quad (k = 0, \dots, n) \quad (12.20)$$

しても，単に代入されるだけで，

$$\mathbf{y}_k = \mathbf{y}(t_k) = C(t_k)\mathbf{x}(t_k) + \mathbf{v}(t_k) = C(t_k)\mathbf{x}_k + \mathbf{v}_k \quad (12.21)$$

特別なことは起こらない．時間微分の項がないので，離散化しても Δt が現れないためである．したがって，雑音強度の補正も不要であり，そのまま次式を得る．

$$\langle \mathbf{v}_k \mathbf{v}_k^T \rangle = R(t_k) \delta_{kl} \quad (12.22)$$

以上の結果を，算法 12.1 と合せてまとめると，次の算法が得られる．

算法 12.2 (確率システムの離散時間近似) 連続時間の確率システム

$$\begin{cases} \dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) + D(t)\mathbf{w}(t), & \mathbf{x}(t) \in \mathbb{R}^n, \mathbf{w}(t) \in \mathbb{R}^r, \\ \mathbf{y}(t) = C(t)\mathbf{x}(t) + \mathbf{v}(t), & \mathbf{y}(t), \mathbf{v}(t) \in \mathbb{R}^p, \\ \langle \mathbf{w}(t) \rangle = \mathbf{0}, & \langle \mathbf{w}(t)\mathbf{w}(s)^T \rangle = Q(t)\delta(t-s), \\ \langle \mathbf{v}(t) \rangle = \mathbf{0}, & \langle \mathbf{v}(t)\mathbf{v}(s)^T \rangle = R(t)\delta(t-s) \quad (s \geq t) \end{cases} \quad (12.23)$$

の解過程は，次の離散時間確率システムで近似できる．

$$\begin{cases} \mathbf{x}_{k+1} = F_k \mathbf{x}_k + G_k \mathbf{w}_k, & \mathbf{x}_k \in \mathbb{R}^n, \mathbf{w}_k \in \mathbb{R}^r \\ \mathbf{y}_k = H_k \mathbf{x}_k + \mathbf{v}_k, & \mathbf{y}_k, \mathbf{v}_k \in \mathbb{R}^p, k = 0, 1, 2, \dots \\ \langle \mathbf{w}_k \rangle = \mathbf{0}, & \langle \mathbf{w}_k \mathbf{w}_l^T \rangle = Q_k \delta_{kl}, \\ \langle \mathbf{v}_k \rangle = \mathbf{0}, & \langle \mathbf{v}_k \mathbf{v}_l^T \rangle = R_k \delta_{kl} \end{cases} \quad (12.24)$$

ただし，係数行列は次のようにとる．

$$\begin{cases} F_k = I + \Delta t A(t_k), & G_k = \sqrt{\Delta t} D(t_k), & H_k = C(t_k), \\ Q_k = Q(t_k), & R_k = R(t_k) \end{cases} \quad (12.25)$$

12.2 Kalman–Bucy フィルタ

算法 11.6 p101 のカルマンフィルタは差分方程式で表現されているが，これらに (12.25) を代入して，微分方程式による表現を導く．このような連続時間型のカルマンフィルタを Kalman–Bucy[†] フィルタという⁴⁾．

12.2.1 カルマンゲイン

(11.60) に，係数 (12.25) を代入すると，

$$\begin{aligned} K_k &= \Sigma_k C(t_k)^T \left(C(t_k) \Sigma_k C(t_k)^T + \frac{R(t_k)}{\Delta t} \right)^{-1} \\ &= \Delta t \Sigma_k C(t_k)^T (\Delta t C(t_k) \Sigma_k C(t_k)^T + R(t_k))^{-1} =: \Delta t K'_k \end{aligned} \quad (12.26)$$

が得られる． $\Delta t \rightarrow 0$ のとき， $K'_k \rightarrow \Sigma(t) C(t)^T R(t)^{-1}$ となる．

12.2.2 予測推定値

(11.62) に，係数 (12.25) を代入すると，

$$\begin{aligned} \hat{\mathbf{x}}_{k+1} &= \{I + \Delta t A(t_k)\} \hat{\mathbf{x}}_k + \{I + \Delta t A(t_k)\} \Delta t K'_k \{\mathbf{y}_k - C(t_k) \hat{\mathbf{x}}_k\} \\ &= \hat{\mathbf{x}}_k + \Delta t A(t_k) \hat{\mathbf{x}}_k + \Delta t K'_k \{\mathbf{y}_k - C(t_k) \hat{\mathbf{x}}_k\} + O(\Delta t^2) \end{aligned} \quad (12.27)$$

すなわち，

$$\frac{\hat{\mathbf{x}}_{k+1} - \hat{\mathbf{x}}_k}{\Delta t} = A(t_k) \hat{\mathbf{x}}_k + K'_k \{\mathbf{y}_k - C(t_k) \hat{\mathbf{x}}_k\} + O(\Delta t) \quad (12.28)$$

となり， $\Delta t \rightarrow 0$ の極限において，

$$\dot{\hat{\mathbf{x}}}(t) = A(t) \hat{\mathbf{x}}(t) + \Sigma(t) C(t)^T R(t)^{-1} \{\mathbf{y}(t) - C(t) \hat{\mathbf{x}}(t)\} \quad (12.29)$$

が得られる．また， $\Delta t \rightarrow 0$ の極限で $\hat{\mathbf{x}}_{t/t-\Delta t} \rightarrow \hat{\mathbf{x}}_{t/t}$ となるから，濾波推定値 $\hat{\mathbf{x}}(t)$ は (12.29) の解 $\hat{\mathbf{x}}(t)$ として得られることになる．

12.2.3 予測推定誤差

同じく (11.63) に代入すると，例題 12.1 を参考に，

$$\Sigma_{k+1} = F_k \Sigma_k F_k^T + G_k Q_k G_k^T - F_k K_k C(t_k) \Sigma_k F_k^T$$

[†] かるまん・ぶーしー

$$\begin{aligned}
&= \Sigma_k + \Delta t \left\{ A(t_k) \Sigma_k + \Sigma_k A(t_k)^T \right\} + \Delta t D(t_k) Q(t_k) D(t_k)^T \\
&\quad - \left\{ I + \Delta t A(t_k) \right\} \Delta t K'_k C(t_k) \Sigma_k \left\{ I + \Delta t A(t_k) \right\}^T + O(\Delta t^2) \\
&= \Sigma_k + \Delta t \left\{ A(t_k) \Sigma_k + \Sigma_k A(t_k)^T + D(t_k) Q(t_k) D(t_k)^T - K'_k C(t_k) \Sigma_k \right\} \\
&\quad + O(\Delta t^2) \tag{12.30}
\end{aligned}$$

より,

$$\begin{aligned}
\frac{\Sigma_{k+1} - \Sigma_k}{\Delta t} &= A(t_k) \Sigma_k + \Sigma_k A(t_k)^T + D(t_k) Q(t_k) D(t_k)^T - K'_k C(t_k) \Sigma_k \\
&\quad + O(\Delta t) \tag{12.31}
\end{aligned}$$

となる．ゆえに， $\Delta t \rightarrow 0$ の極限で，次の関係が得られる．

$$\dot{\Sigma} = A(t) \Sigma + \Sigma A(t)^T + D(t) Q(t) D(t)^T - \Sigma C(t)^T R(t)^{-1} C(t) \Sigma \tag{12.32}$$

以上をまとめると，連続時間システム用のフィルタが次のように得られる[†]．

算法 12.3 (Kalman–Bucy フィルタ) 連続時間線形確率システム：

$$\begin{cases} \dot{\mathbf{x}} = A(t) \mathbf{x} + D(t) \mathbf{w}(t), & \mathbf{x}(t) \in \mathbb{R}^n, \mathbf{w}(t) \in \mathbb{R}^r, \\ \mathbf{y} = C(t) \mathbf{x} + \mathbf{v}(t), & \mathbf{y}(t), \mathbf{v}(t) \in \mathbb{R}^p \end{cases} \tag{12.33}$$

$$\begin{cases} \langle \mathbf{w}(t) \rangle = \mathbf{0}, & \langle \mathbf{w}(t) \mathbf{w}(s)^T \rangle = Q(t) \delta(t-s), \\ \langle \mathbf{v}(t) \rangle = \mathbf{0}, & \langle \mathbf{v}(t) \mathbf{v}(s)^T \rangle = R(t) \delta(t-s) \quad (s \geq t) \end{cases} \tag{12.34}$$

の最小分散濾波推定値 $\hat{\mathbf{x}}(t)$ は，次の微分方程式系の解として得られる．すなわち，既知の観測量 $\mathbf{y}(t)$ と，適当な初期条件 $\hat{\mathbf{x}}(0) = \bar{\mathbf{x}}_0$ ， $\Sigma(0) = \Sigma_0$ に対して，

$$K(t) := \Sigma C(t)^T R(t)^{-1} \quad (\text{カルマンゲイン}) \tag{12.35a}$$

$$\dot{\hat{\mathbf{x}}} = A(t) \hat{\mathbf{x}} + K(t) \{ \mathbf{y} - C(t) \hat{\mathbf{x}} \} \quad (\text{濾波推定値}) \tag{12.35b}$$

$$\dot{\Sigma} = A(t) \Sigma + \Sigma A(t)^T + D(t) Q(t) D(t)^T - K(t) C(t) \Sigma \quad (\text{誤差}) \tag{12.35c}$$

算法 12.3 の具体的なプログラム例は，本書レポジトリ (iv 頁) の Ch12.ipynb で確認できる．

[†] 加藤⁷⁾ p137 など．

12.3 定常 Kalman–Bucy フィルタ

Kalman–Bucy フィルタについても, 11.5 節と同様な定常フィルタが構成できる. 同じく, 行列 $A(t)$, $C(t)$, $D(t)$, $Q(t)$, $R(t)$ が全て定行列 A , C , D , Q , R の場合を考える. このとき, 共分散行列 $\Sigma(t)$ の時間発展は定係数の常微分方程式に従うので, それが安定なら, 定常値 $\Sigma = \lim_{t \rightarrow \infty} \Sigma(t)$ が存在し, カルマンゲインも定常値 $K = \lim_{t \rightarrow \infty} K(t)$ をもつ.

12.3.1 基本アルゴリズム

定常値 Σ , K を, 定常条件 $\dot{\Sigma} = O$, $\dot{\hat{x}} = 0$ とともに, (12.35a), (12.35c) に代入すると, 次の連立代数方程式を得る.

$$O = A\Sigma + \Sigma A^T + DQD^T - KC\Sigma, \quad K = \Sigma C^T R^{-1} \quad (12.36)$$

第 1 式を, 連続時間代数リカッチ方程式という. (12.36) から直ちに Σ , K が求まる. 得られたカルマンゲイン K を使うと, 濾波推定値 \hat{x} が次式で更新できる.

$$\dot{\hat{x}} = A\hat{x} + K(y(t) - C\hat{x}) \quad (12.37)$$

以上, Kalman–Bucy フィルタを定係数としたところ, 算法 12.3 が, (12.36), (12.37) に集約されてしまった. これを, 定常 Kalman–Bucy フィルタという. 次のように整理できる.

算法 12.4 (定常 Kalman–Bucy フィルタ) 定係数の連続時間線形確率システム:

$$\begin{cases} \dot{x} = Ax + Dw(t), & x(t) \in \mathbb{R}^n, w(t) \in \mathbb{R}^r, \\ y = Cx + v(t), & y(t), v(t) \in \mathbb{R}^p \end{cases} \quad (12.38)$$

$$\begin{cases} \langle w(t) \rangle = 0, & \langle w(t)w(s)^T \rangle = Q\delta(t-s), \\ \langle v(t) \rangle = 0, & \langle v(t)v(s)^T \rangle = R\delta(t-s) \quad (s \geq t) \end{cases} \quad (12.39)$$

の最小分散濾波推定値 $\hat{x}(t)$ は, 次のようにして得られる.

(0) 前処理と初期条件: 次の代数方程式から Σ と K を事前に求めておく.

$$K := \Sigma C^T R^{-1} \quad (\text{定常カルマンゲイン}) \quad (12.40)$$

$$O = A\Sigma + \Sigma A^T + DQD^T - KC\Sigma \quad (\text{連続時間代数リカッチ方程式}) \quad (12.41)$$

また, 推定値の初期値 \hat{x}_0 を定めておく.

(1) 濾波推定値：

$$\dot{\hat{x}} = A\hat{x} + K\{y(t) - C\hat{x}\} \quad (\text{濾波推定値}) \quad (12.42)$$

すなわち，既知の観測量 $y(t)$ と，適当な初期条件 $\hat{x}(0) = \bar{x}_0$ ， $\Sigma(0) = \Sigma_0$ に対して，

離散時間のときと同様に，代数リカッチ方程式 (12.41) を事前に解いておけば，状態推定器には，(12.42) だけ実装すればよい．代数リカッチ方程式 (12.41) の解法が，利用可能な点も同様である．ただし，本章の連続時間代数リカッチ方程式 (12.41) は，前章の離散時間代数リカッチ方程式 (11.67) とは式の形が異なり，適用する解法も異なるので注意が必要である．

▶▶ (読み替えのヒント) 例えば，ある言語の連続時間代数リカッチ方程式ソルバが，

$$0 = X\bar{A} + \bar{A}^T X - X\bar{B}\bar{R}^{-1}\bar{B}^T X + \bar{Q}$$

の X を求める仕様だった場合，(12.41) を解くには $\bar{A} \mapsto A^T$ ， $\bar{B} \mapsto C^T$ ， $\bar{Q} \mapsto D^T Q D$ ， $\bar{R} \mapsto R$ という読み替えが必要になる．

12.3.2 定常 Kalman–Bucy フィルタの安定性

推定値 \hat{x} の安定性を調べるために，誤差ベクトル $\vec{x} := \hat{x} - x$ を導入すると， $y - C\hat{x} = Cx_t + v(t) - H\hat{x} = C(\hat{x} - \vec{x}) + v(t) - C\hat{x}_t = v(t) - C\vec{x}_t$ より，推定値 (12.35b) は，

$$\dot{\hat{x}} = A\hat{x} + K(y - C\hat{x}) = A\hat{x} + K(v(t) - C\vec{x}_t) = A\hat{x} - KC\vec{x} + Kv(t) \quad (12.43)$$

となる．これと (12.33) の第 1 式より，

$$\begin{aligned} \dot{\vec{x}} &= \dot{\hat{x}} - \dot{x} = A\hat{x} - KC\vec{x} + Kv(t) - Ax - Dw(t) \\ &= A\vec{x} - KC\vec{x} + Kv(t) - Dw(t) \\ \therefore \dot{\vec{x}} &= (A - KC)\vec{x}_t + S(t), \quad S(t) := Kv(t) - Dw(t) \end{aligned} \quad (12.44)$$

が得られる．この線形システムが安定ならば，推定誤差 \vec{x}_t は安定となる．すなわち，次の安定条件が得られる．

算法 12.5 (定常 Kalman–Bucy フィルタの安定条件) 行列 $(A - KC)$ の固有値実部が，全て負値のとき，推定誤差 \vec{x}_t は漸近安定となる．

当然のことながら，これを満たさない条件では，推定誤差が収束しないので Kalman–Bucy フィルタは使用できない．

12.4 数 値 例

11.6 節と同じ問題

$$\dot{\mathbf{x}} = \underbrace{\begin{pmatrix} 0 & 1 \\ -k & -c \end{pmatrix}}_A \mathbf{x} + \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}_D w(t) \quad (12.45)$$

$$y = \underbrace{(0, 1)}_C \mathbf{x} + v(t) = x_2 + v(t) \quad (12.46)$$

を，連続時間のまま算法 12.3 で解いた数値例を図 12.1 に示す．数値例の作成には，本書レポジトリ (iv 頁) の Ch12.ipynb を用いた．

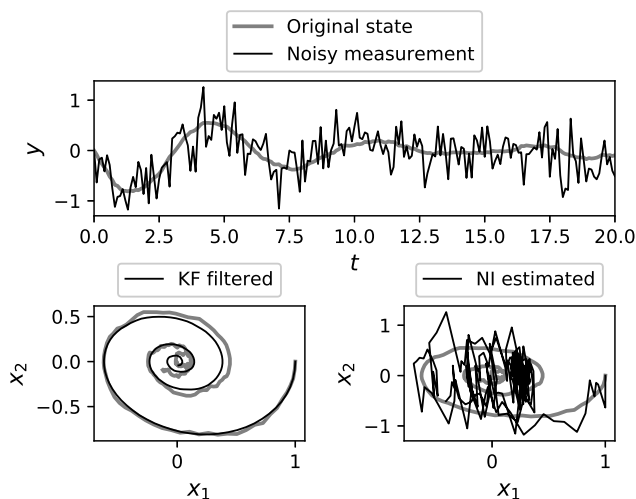


図 12.1 Kalman-Bucy フィルタによる状態推定 (下段左) と数値積分による変位推定 (下段右) の比較 ($c = 0.3$, $k = 1$, $\Delta t = 0.1$, $Q = 0.003$, $R = 0.1$) . 上段の黒線は元にした観測量 y_n を表す .

例題の解答例

例題 12.1 p108

算法 10.5 の期待値の推移則に (12.16) の F_k を代入すると ,

$$\bar{\mathbf{x}}_{k+1} = F_k \bar{\mathbf{x}}_k = (I + \Delta t A(t_k)) \bar{\mathbf{x}}_k = \bar{\mathbf{x}}_k + \Delta t A(t_k) \bar{\mathbf{x}}_k \quad (12.47)$$

より ,

$$\frac{\bar{\mathbf{x}}(t_k + \Delta t) - \bar{\mathbf{x}}(t_k)}{\Delta t} = A(t_k) \bar{\mathbf{x}}(t_k) \quad (12.48)$$

となる . $\Delta t \rightarrow 0$ の極限をとり , $t = t_k$ とすると , (12.1) の期待値の推移則

$$\dot{\bar{\mathbf{x}}}(t) = A(t) \bar{\mathbf{x}}(t) \quad (12.49)$$

を得る . 同様に , 共分散行列の推移則 (10.35) p87 に (12.16) を代入すると ,

$$\begin{aligned} \Sigma_{k+1} &= F_k \Sigma_k F_k^T + G_k Q_k G_k^T \\ &= \left(I + \Delta t A(t_k) \right) \Sigma_k \left(I + \Delta t A(t_k) \right)^T + \left(\sqrt{\Delta t} D(t_k) \right) Q(t_k) \left(\sqrt{\Delta t} D(t_k) \right)^T \\ &= \Sigma_k + \Delta t \left(A(t_k) \Sigma_k + \Sigma_k A(t_k)^T \right) + O(\Delta t^2) + \Delta t \left(D(t_k) Q(t_k) D(t_k)^T \right) \end{aligned} \quad (12.50)$$

となる .

▶▶ $O(\Delta t^2)$ で Δt の 2 乗以下の項を表す . 同様に , $O(\Delta t)$ は 1 乗以下の項である . こう表すときの O を (ランダウの) オミクロンと呼ぶ .

すなわち ,

$$\begin{aligned} \frac{\Sigma(t_k + \Delta t) - \Sigma(t_k)}{\Delta t} &= A(t_k) \Sigma(t_k) + \Sigma(t_k) A(t_k)^T + D(t_k) Q(t_k) D(t_k)^T + O(\Delta t) \end{aligned} \quad (12.51)$$

が得られる . $\Delta t \rightarrow 0$ の極限をとり , 改めて $t = t_k$ と置くと ,

$$\dot{\Sigma}(t) = A(t) \Sigma(t) + \Sigma(t) A(t)^T + D(t) Q(t) D(t)^T \quad (12.52)$$

という推移則が得られる .

13

LQG 制御への応用

Kalman–Bucy フィルタを制御に応用する .

13.1 制御入力を考慮したカルマンフィルタ

算法 12.3 p111 のシステムに , 制御入力 $u(t) \in \mathbb{R}^m$ を付加した問題を考える .

$$\dot{x} = A(t)x + B(t)u(t) + D(t)w(t), \quad x(t) \in \mathbb{R}^n, w(t) \in \mathbb{R}^r \quad (13.1)$$

$$y = C(t)x + v(t), \quad y(t), v(t) \in \mathbb{R}^p \quad (13.2)$$

ただし , $u(t)$ は , 時刻 0 から t までの観測量 y に依存すると仮定する . すなわち , 区間 $0 \leq s \leq t$ 上の $y(s)$ の時刻歴 (グラフ) を $\mathcal{Y}(t)$ と表記するとき , $u(t) := u(\mathcal{Y}(t))$ であると仮定する .

▶▶ 厳密にいうと , 確率変数 $y(s)$ ($0 \leq s \leq t$) が生成する σ 代数 $\mathcal{Y}(t)$ に関して , $u(t)$ は $\mathcal{Y}(t)$ -可測であると仮定する .

13.1.1 システムの分解

(13.1) は線形なので , 状態 $x = x(t)$ は次のように分解できる .

$$x = x_u + x_w : \begin{cases} \dot{x}_u = A(t)x_u + B(t)u(t) \\ \dot{x}_w = A(t)x_w + D(t)w(t) \end{cases} \quad (13.3)$$

確かに辺々足せば (13.1) に戻る . 同様に観測量 $y = y(t)$ も分解できる .

$$y = y_u + y_w : \begin{cases} y_u = C(t)x_u \\ y_w = C(t)x_w + v(t) \end{cases} \quad (13.4)$$

これも辺々足せば (13.2) に戻る . 以上をまとめると ,

$$x = x_u + x_w, \quad y = y_u + y_w :$$

$$\begin{cases} \dot{\mathbf{x}}_u = A(t)\mathbf{x}_u + B(t)\mathbf{u}(t) \\ \mathbf{y}_u = C(t)\mathbf{x}_u \end{cases} \quad (13.5a)$$

$$\begin{cases} \dot{\mathbf{x}}_w = A(t)\mathbf{x}_w + D(t)\mathbf{w}(t) \\ \mathbf{y}_w = C(t)\mathbf{x}_w + \mathbf{v}(t) \end{cases} \quad (13.5b)$$

という分解が成立する．

13.1.2 推定値の性質

$\mathbf{u}(t)$ は観測量の時刻歴の関数 $\mathbf{u}(t) = \mathbf{u}(\mathcal{Y}(t))$ だと仮定したので、これを入力とする (13.5a) の応答 $\mathbf{x}_u(t)$ も $\mathcal{Y}(t)$ の関数となる．このとき、 $\mathcal{Y}(t)$ をデータとする $\mathbf{x}_u(t)$ の推定値は、算法 6.11 p50 条件付き期待値の性質 (2) より、

$$\hat{\mathbf{x}}_u(t) := \langle \mathbf{x}_u(t) | \mathcal{Y}(t) \rangle = \langle \mathbf{x}_u(\mathcal{Y}(t)) | \mathcal{Y}(t) \rangle = \mathbf{x}_u(\mathcal{Y}(t)) = \mathbf{x}_u(t) \quad (13.6)$$

となる．これを (13.5a) に代入すると、 $\hat{\mathbf{x}}_u(t)$ の推移則

$$\dot{\hat{\mathbf{x}}}_u = \dot{\mathbf{x}}_u = A(t)\mathbf{x}_u + B(t)\mathbf{u}(t) = A(t)\hat{\mathbf{x}}_u + B(t)\mathbf{u}(t) \quad (13.7)$$

が得られる．他方、 $\hat{\mathbf{x}}_w(t)$ の推移則は、(13.5b) に算法 12.3 を適用して、

$$\dot{\hat{\mathbf{x}}}_w = A(t)\hat{\mathbf{x}}_w + \Sigma_w C(t)^T R(t)^{-1} (\mathbf{y}_w - C(t)\hat{\mathbf{x}}_w) \quad (13.8)$$

$$\dot{\Sigma}_w = A(t)\Sigma_w + \Sigma_w A(t)^T + D(t)Q(t)D(t)^T - \Sigma_w C(t)^T R(t)^{-1} C(t)\Sigma_w \quad (13.9)$$

となる．(13.7) と (13.8) を辺々足して、 $\mathbf{y}_u = C(t)\mathbf{x}_u = C(t)\hat{\mathbf{x}}_u$ を使うと、

$$\dot{\hat{\mathbf{x}}} = A(t)\hat{\mathbf{x}} + B(t)\mathbf{u}(t) + \Sigma_w C(t)^T R(t)^{-1} (\mathbf{y} - C(t)\hat{\mathbf{x}}) \quad (13.10)$$

が得られる．また、誤差 $\mathbf{x} - \hat{\mathbf{x}}$ の共分散は、

$$\begin{aligned} \Sigma &= \langle (\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T \rangle = \langle \mathbf{x}\mathbf{x}^T \rangle - \hat{\mathbf{x}}\hat{\mathbf{x}}^T \\ &= \langle (\mathbf{x}_u + \mathbf{x}_w)(\mathbf{x}_u + \mathbf{x}_w)^T \rangle - (\hat{\mathbf{x}}_u + \hat{\mathbf{x}}_w)(\hat{\mathbf{x}}_u + \hat{\mathbf{x}}_w)^T \end{aligned} \quad (13.11)$$

となるが、 $\mathbf{x}_u = \hat{\mathbf{x}}_u$ より、

$$\begin{aligned} &= \hat{\mathbf{x}}_u \hat{\mathbf{x}}_u^T + \hat{\mathbf{x}}_u \langle \mathbf{x}_w^T \rangle + \langle \mathbf{x}_w \rangle \hat{\mathbf{x}}_u^T + \langle \mathbf{x}_w \mathbf{x}_w^T \rangle - (\hat{\mathbf{x}}_u + \hat{\mathbf{x}}_w)(\hat{\mathbf{x}}_u + \hat{\mathbf{x}}_w)^T \\ &= \hat{\mathbf{x}}_u \hat{\mathbf{x}}_u^T + \hat{\mathbf{x}}_u \hat{\mathbf{x}}_w^T + \hat{\mathbf{x}}_w \hat{\mathbf{x}}_u^T + \langle \mathbf{x}_w \mathbf{x}_w^T \rangle - (\hat{\mathbf{x}}_u + \hat{\mathbf{x}}_w)(\hat{\mathbf{x}}_u + \hat{\mathbf{x}}_w)^T \\ &= \langle \mathbf{x}_w \mathbf{x}_w^T \rangle - \hat{\mathbf{x}}_w \hat{\mathbf{x}}_w^T = \langle (\mathbf{x}_w - \hat{\mathbf{x}}_w)(\mathbf{x}_w - \hat{\mathbf{x}}_w)^T \rangle =: \Sigma_w \end{aligned} \quad (13.12)$$

となり、 \mathbf{x}_w の誤差共分散 (13.9) に一致する．以上をまとめると、

算法 13.1 (制御入力を考慮した Kalman–Bucy フィルタ) 連続時間線形確率システム：

$$\begin{cases} \dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) + B(t)\mathbf{u}(t) + D(t)\mathbf{w}(t), & \mathbf{x}(t) \in \mathbb{R}^n, \mathbf{w}(t) \in \mathbb{R}^r, \\ \mathbf{y}(t) = C(t)\mathbf{x}(t) + \mathbf{v}(t), & \mathbf{y}(t), \mathbf{v}(t) \in \mathbb{R}^p, \mathbf{u}(t) \in \mathbb{R}^m \\ \langle \mathbf{w}(t) \rangle = \mathbf{0}, \quad \langle \mathbf{w}(t)\mathbf{w}(s)^T \rangle = Q(t)\delta(t-s), \\ \langle \mathbf{v}(t) \rangle = \mathbf{0}, \quad \langle \mathbf{v}(t)\mathbf{v}(s)^T \rangle = R(t)\delta(t-s) \quad (s \geq t) \end{cases}$$

の最小分散濾波推定値 $\hat{\mathbf{x}}(t)$ は、次の微分方程式系の解として得られる．すなわち、既知の観測量 $\mathbf{y}(t)$ と、適当な初期条件 $\hat{\mathbf{x}}(0) = \bar{\mathbf{x}}_0$, $\Sigma(0) = \Sigma_0$ に対して、

$$K(t) := \Sigma C(t)^T R(t)^{-1} \quad (\text{カルマンゲイン}) \quad (13.13a)$$

$$\dot{\hat{\mathbf{x}}} = A(t)\hat{\mathbf{x}} + B(t)\mathbf{u}(t) + K(t)(\mathbf{y}(t) - C(t)\hat{\mathbf{x}}) \quad (\text{推定値}) \quad (13.13b)$$

$$\dot{\Sigma} = A(t)\Sigma + \Sigma A(t)^T + D(t)Q(t)D(t)^T - K(t)C(t)\Sigma \quad (\text{誤差}) \quad (13.13c)$$

算法 12.3 と比較すると、推定値 (13.13b) に $B(t)\mathbf{u}(t)$ が追加されるが、誤差 (13.13c) は変化しない．

13.2 LQG レギュレータ

算法 12.3 p111 の制御入力を、通常の状態フィードバック $\mathbf{u}(t) := -F\mathbf{x}(t)$ ではなく、濾波推定値 $\hat{\mathbf{x}}(t)$ によるフィードバック

$$\mathbf{u}(t) := -F\hat{\mathbf{x}}(t) \quad (13.14)$$

とした制御方式を、**LQG 制御** (linear quadratic Gaussian control) という．LQG 制御は、状態量 $\mathbf{x}(t)$ や観測量 $\mathbf{y}(t) = C(t)\mathbf{x}(t)$ がノイズに乱されていても、カルマンフィルタの意味で最適なフィードバック制御が期待できる．

制御入力 $\mathbf{u}(t)$ に必要な $\hat{\mathbf{x}}(t)$ を決める (13.13) に、(13.14) を代入して整理すると、次のような算法が得られる．

算法 13.2 (LQG 制御) 連続時間線形確率システム：

$$\begin{cases} \dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) + B(t)\mathbf{u}(t) + D(t)\mathbf{w}(t), & \mathbf{x}(t) \in \mathbb{R}^n, \mathbf{w}(t) \in \mathbb{R}^r, \\ \mathbf{y}(t) = C(t)\mathbf{x}(t) + \mathbf{v}(t), & \mathbf{y}(t), \mathbf{v}(t) \in \mathbb{R}^p, \mathbf{u}(t) \in \mathbb{R}^m \end{cases}$$

$$\begin{cases} \langle \mathbf{w}(t) \rangle = \mathbf{0}, & \langle \mathbf{w}(t) \mathbf{w}(s)^T \rangle = Q(t) \delta(t-s), \\ \langle \mathbf{v}(t) \rangle = \mathbf{0}, & \langle \mathbf{v}(t) \mathbf{v}(s)^T \rangle = R(t) \delta(t-s) \quad (s \geq t) \end{cases}$$

の最小分散濾波推定値 $\hat{\mathbf{x}}(t)$ による状態フィードバック制御器 $\mathbf{u}(t) = \mathbf{u}^{\text{LQG}}(t)$ を LQG レギュレータ (linear quadratic Gaussian regulator) という。 $\mathbf{u}^{\text{LQG}}(t)$ は次のように得られる。

既知の観測量 $\mathbf{y}(t)$ と，適当な初期条件 $\hat{\mathbf{x}}(0) = \bar{\mathbf{x}}_0$ ， $\Sigma(0) = \Sigma_0$ に対して，

$$\mathbf{u}(t) = \mathbf{u}^{\text{LQG}}(t) := -F\hat{\mathbf{x}}(t) \quad (\text{LQG 制御器}) \quad (13.15a)$$

$$K(t) := \Sigma C(t)^T R(t)^{-1} \quad (\text{カルマンゲイン}) \quad (13.15b)$$

$$\dot{\hat{\mathbf{x}}} = (A(t) - B(t)F)\hat{\mathbf{x}} + K(t)(\mathbf{y}(t) - C(t)\hat{\mathbf{x}}) \quad (\text{推定値}) \quad (13.15c)$$

$$\dot{\Sigma} = A(t)\Sigma + \Sigma A(t)^T + D(t)Q(t)D(t)^T - K(t)C(t)\Sigma \quad (\text{誤差}) \quad (13.15d)$$

これを，LQG レギュレータ (linear quadratic Gaussian regulator) という。 算法 13.2 の プログラム例は，本書レポジトリ (iv 頁) の Ch13.ipynb で確認できる。

特に，行列 $A(t)$ ， $B(t)$ ， $C(t)$ ， $D(t)$ ， $Q(t)$ ， $R(t)$ が全て定行列の場合，

$$\begin{cases} \mathbf{u}(t) := -F\hat{\mathbf{x}}(t), \\ K := \Sigma C^T R^{-1}, \\ \dot{\hat{\mathbf{x}}} = (A - BF)\hat{\mathbf{x}} + K(\mathbf{y}(t) - C\hat{\mathbf{x}}), \\ O = A\Sigma + \Sigma A^T + DQD^T - KC\Sigma \quad (O \text{ はゼロ行列}) \end{cases} \quad (13.16)$$

を，定常 LQG レギュレータ (stationary LQG regulator) という。

13.3 分離定理

簡単のため，定常 LQG レギュレータ (13.16) について考える。カルマンゲイン K は， Σ に依存し，この Σ の支配方程式 (第 4 式) はフィードバックゲイン F を含まない。すなわち，カルマンゲイン K は，フィードバックゲイン F の選択とは無関係に定まる。最適制御理論[†]を参考に，以下，詳しく見てみる。

(13.16) の $\mathbf{u}(t) := -F\hat{\mathbf{x}}(t)$ を，(13.1) に代入すると，

$$\dot{\mathbf{x}} = A\mathbf{x} - BF\hat{\mathbf{x}}(t) + D\mathbf{w}(t) \quad (13.17)$$

[†] 加藤⁷⁾p145 など。

となる．誤差ベクトル $\vec{x} := \hat{x} - x$ をとって， \hat{x} を消去すると，

$$\dot{x} = (A - BF)x - BF\vec{x} + Dw(t) \quad (13.18)$$

となる．また， $y = Cx + v(t)$ より，

$$y - C\hat{x} = Cx + v(t) - C(\vec{x} + x) = -C\vec{x} + v(t) \quad (13.19)$$

となるから，(13.16) の第 3 式は，

$$\dot{\hat{x}} = (A - BF)\hat{x} - KC\vec{x} + Kv(t) \quad (13.20)$$

と書ける．(13.20) から (13.18) を辺々引くと，誤差ベクトルの法則を得る．

$$\begin{aligned} \dot{\vec{x}} &= (A - BF)\vec{x} + (BF - KC)\vec{x} - Dw(t) + Kv(t) \\ &= (A - KC)\vec{x} - Dw(t) + Kv(t) \end{aligned} \quad (13.21)$$

以上に得られた，(13.18) と (13.21) をまとめて書くと，

$$\begin{aligned} \frac{d}{dt} \begin{pmatrix} x \\ \vec{x} \end{pmatrix} &= \begin{pmatrix} A - BF & -BF \\ 0 & A - KC \end{pmatrix} \begin{pmatrix} x \\ \vec{x} \end{pmatrix} + P(t) \\ \text{ただし, } P(t) &= \begin{pmatrix} D \\ -D \end{pmatrix} w(t) + \begin{pmatrix} 0 \\ K \end{pmatrix} v(t) \end{aligned} \quad (13.22)$$

を得る．一般に，線形システム (13.22) の安定性は，係数行列の固有値で決まる．さらに，係数行列は上三角のブロック行列なので，状態ベクトル x の安定性は，左上の行列 $(A - BF)$ の固有値で定まる．また，これとは独立に，誤差ベクトル \vec{x} の安定性は，右下の行列 $(A - KC)$ の固有値で定まる^{†1}．

したがって，2 つのゲイン F と K に着目すると，状態 x の安定性は状態フィードバックゲイン F だけに依存し，推定誤差 \vec{x} の安定性はカルマンゲイン K だけに依存することが分かる．この性質を，(制御と状態推定の) 分離定理という^{†2}．

算法 13.3 (分離定理) 定常 LQG レギュレータ (13.16) において，状態フィードバックゲイン F と，カルマンゲイン K は，互いに影響を与えずに，別々に設計することができる．

13.4 数 値 例

図 1.6 p5 の制御方法を明らかにする．問題を再掲する． m は質量， k はばね定数， c は減

^{†1} (12.44) p113 と同じ結果に帰着したことを意味する．

^{†2} 加藤⁷⁾ p157 など．

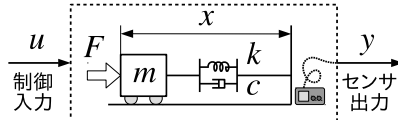


図 1.1 の 1 自由度機械振動系

衰係数とし，質点の実際の変位を x で表した．簡単のため， $m = 1 \text{ kg}$ とし， $u = F$ の場合を考えた（制御入力外力そのもの）．そして，対応する状態方程式と，次のような観測方程式を考えた．

$$\begin{cases} \dot{x} = Ax + Bu, & x := \begin{pmatrix} x \\ \dot{x} \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 1 \\ -k & -c \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ y = Cx = x_1 = x, & C = (1, 0) \end{cases} \quad (13.23)$$

さて，図 1.1 の制御系とは，(13.23) の制御系に，システム雑音 $w(t)$ と，観測雑音 $v(t)$ を加味した次の LQG 制御系である．

$$\begin{cases} \dot{x} = Ax + Bu + Dw, & D = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ y = Cx + v, \\ u = u^{\text{LQG}} := -F\hat{x}, & F = (F_1, F_2) \end{cases} \quad (13.24)$$

パラメータ条件は， $m = 1$ ， $c = k = 0.1$ とした．また， $w(t)$ と $v(t)$ の分散はそれぞれ $Q = 0$ および $R = 0.01$ とした（この問題ではいずれもスカラ）．ここで， $F = (0, 0.45)$ とし推定値の速度成分をフィードバックした LQG 制御の結果が，図 1.6 p5 である．図中の黒線が LQG 制御 $u(t) = u^{\text{LQG}} := -F\hat{x}$ を受けた質点の運動 $x(t)$ である．

以上の数値例の作成には，本書レポジトリ (iv 頁) の Ch13.ipynb を用いた．

カルマンフィルタは，状態の推定だけでなく，パラメータの推定にも使える．ただし，そのために解くべき問題は一般に非線形になるので，カルマンフィルタの拡張が必要である．

なお，以下では適宜，次の略号を用いる．ちょっともう，見るからに紛らわしいのだが，広く認知された略号なのでご容赦いただきたい．

略号 (元の英語)	日本語
KF (Kalman filter)	カルマンフィルタ
KBF (Kalman–Bucy filter)	カルマンブーシーフィルタ
EKF (Extended Kalman filter)	拡張カルマンフィルタ
EKBF (Extended Bucy–Kalman filter)	拡張カルマンブーシーフィルタ
EnKF (Ensemble Kalman filter)	アンサンブルカルマンフィルタ

14.1 カルマンフィルタによるパラメータ推定

簡単な例を挙げる．一定振幅 a の周期外力 $F := a \sin(\beta t)$ を受ける 1 自由度振動系と，その観測を考える (図 14.1) ．

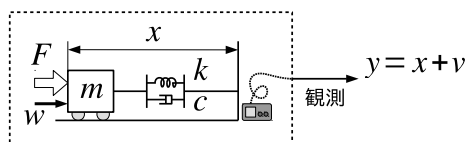


図 14.1 1 自由度機械振動系

この系は，雑音 $w_1(t)$, $w_2(t)$, $v(t)$ を考慮すると，次のように表せる．

$$\begin{aligned}
 \dot{x}_1 &= x_2 + w_1(t), \\
 \dot{x}_2 &= -k x_1 - c x_2 + a \sin(\beta t) + w_2(t), \\
 y &= x + v(t)
 \end{aligned} \tag{14.1}$$

簡単のため $m = 1$ とした．その他のパラメータは，デフォルト値を $k = 1$, $c = 0.2$, $a = 1$, $\beta = 1.5$ としておく．

もしかすると，第 1 式に対する外乱 $w_1(t)$ を奇異に感じる人がいるかもしれない（外力の入りは第 2 式なので）．力学的には確かにそうなのだが，よりシステム工学的な視点で見ると，この種の雑音項は，外乱というより，むしろ，システムの不確かさ（モデル化誤差）を表している．以下，後者の視点で話を進める．

14.1.1 線形なパラメータ推定問題

まずは，カルマンフィルタの拡張なしに解ける問題を見てみよう．

外力振幅 a が測れないときに，観測データ $y(t)$ から推定する問題を考える．この種の問題は，カルマンフィルタで解けるのである．

そのための常套手段として， a が未知の定数であることを，次の微分方程式で表す．

$$\dot{a}(t) = 0 \quad (14.2)$$

確かに，上式を満足する $a(t)$ は，定数だが不定値となる．この (14.2) を元の (14.1) に連立すると，次式のような確率システムが作れる．これを拡大系という．

$$\begin{aligned} \dot{x}_1 &= x_2 + w_1(t), \\ \dot{x}_2 &= -k x_1 - c x_2 + a x_3 + w_2(t), \\ \dot{x}_3 &= 0 + w_3(t), \\ y &= x + v(t) \end{aligned} \quad (14.3)$$

拡大系の状態ベクトル $x := (x, \dot{x}, a)^T$ は，推定したいパラメータ a を成分に含んでいる．これを，拡大状態ベクトルという．

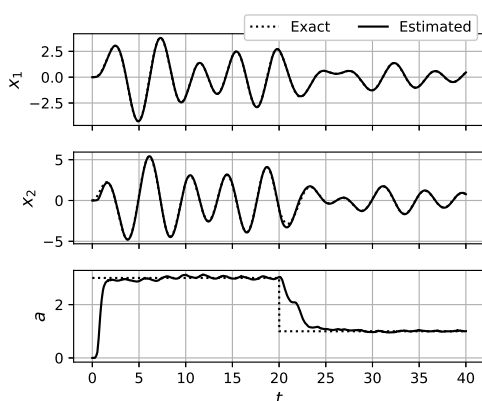
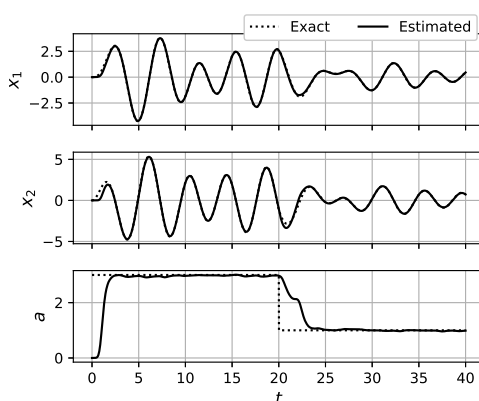
なお，第 3 式に $x_3 = a$ の不確かさを表す $w_3(t)$ を追加したのだが，お気づきだろうか．これを忘れると，推定値の収束性が極端に悪化する場合がある．

さて，(14.3) の拡大系は，（幸運にも）線形なので行列で表示できる．

$$\begin{aligned} \dot{x} &= A(t)x + Dw(t), \quad y = Cx + v(t), \quad x = (x_1, x_2, x_3)^T := (x, \dot{x}, a)^T, \\ A(t) &:= \begin{pmatrix} 0 & 1 & 0 \\ -k & -c & \sin(\beta t) \\ 0 & 0 & 0 \end{pmatrix}, \quad D := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad C := (1, 0, 0) \end{aligned} \quad (14.4)$$

時変の[†]行列 $A(t)$ を含むので，定常カルマンフィルタは使えない．とはいえ，線形ではある

[†] 「時間変化する」の意．

図 14.2 外力振幅 a の推定 (KF)図 14.3 外力振幅 a の推定 (KBF)

ので，算法 11.6 が算法 12.3 で解くことができる．その結果， $(x, \dot{x}, a)^T$ の推定値が得られる．すなわち， x, \dot{x} と一緒に a の値が推定できる．

数値例を示そう．まず，元の連続時間系 (14.1) の数値解を求め，観測量のサンプルデータ $y(t)$ を作成した．その際， a の値をステップ状に次のように変化させた．

$$a(t) = \text{STEP}(t; a_1, a_2, t_1) = \begin{cases} a_1 & (0 \leq t < t_1) \\ a_2 & (t_1 \leq t) \end{cases} \quad (14.5)$$

次に，(14.4) の行列でカルマンフィルタを構成し，サンプルデータ $y(t)$ に対する濾波推定量 $(\hat{x}(t), \hat{\dot{x}}(t), \hat{a}(t))$ を求めた．

図 14.2 は，KF (算法 11.6) による推定結果である．連続時間系のデータ $y(t)$ に対して，離散時間の算法 11.6 を適用するので，算法 12.2 を介してカルマンフィルタを構成した．図中の点線が真値，実線が濾波推定値を表す．上の 2 段が状態量の結果，3 段目がパラメータの結果である．状態の推定量 \hat{x}_1, \hat{x}_2 は収束が速く，これに少し遅れる形で，パラメータ \hat{x}_3 が真値付近に追従している．いずれにしても，KF によってパラメータが推定できた．

図 14.3 は，同様にして求めた，KBF (算法 12.3) による推定結果である．今度はフィルタ自体が連続時間なので，算法 12.2 の適用は必要ない．KF との違いは，思ったほど見られないが，前半の \hat{a} のパタつきが，だいぶ改善されている．

大きな違いは，両者の計算時間である．離散型の KF の計算時間は 0.16 s 程度だったのに対して，連続型の KBF は 2.5 s 程度を要した．つまり，KF の計算速度は，KBF より 15 倍速かった．むろん，計算速度はプログラミング技術にも左右されるので，著者の KBF がダメダメだったせいかもしれない，優劣を断言するのは容易ではない．

なお，図 14.2，図 14.3 の結果は，Ch14-2_KF.ipynb と Ch14-2_KBF.ipynb で作成した．

コードを見ると分かると思うが、時変カルマンフィルタのコードを得るために 14.2 節の EKF 用のコードを流用している。しかし、これはプログラミング上の問題であって、アルゴリズムは算法 11.6 および算法 12.3 のものである。

14.1.2 非線形なパラメータ推定問題

次に、カルマンフィルタの拡張が必要になる場合を述べる。

同じシステム (14.1) について、今度は、減衰係数 c を推定したいとしよう。前節と同じ方法で、未知な定数を $\dot{c} = \dot{x}_3 = 0$ で表し、元のシステム (14.1) と連立する。

$$\begin{aligned}\dot{x}_1 &= x_2 + w_1(t), \\ \dot{x}_2 &= -k x_1 - x_3 x_2 + a \sin(\beta t) + w_2(t), \\ \dot{x}_3 &= 0 + w_3(t), \quad y = x + v(t)\end{aligned}\tag{14.6}$$

残念なことに、非線形項 $x_3 x_2$ が生じてしまった。そのせいで、(14.6) は行列表示できないので、通常の KF や KBF は適用できない。

14.1.3 カルマンフィルタの拡張

KF による濾波推定は (KBF もそうだが)、現代的には、データ同化の一手法に位置づけられる。データ同化とは、実現象をシミュレーションする際に、時々刻々、シミュレータが出力する数値解を現実の観測値と照合することで、数値解を補正していく方法論のことである。図 11.1 を見ると確かにそうになっている。予測推定値 $\hat{x}_{t/t-1}$ が上述の修正前の数値解に相当し、濾波推定値 $\hat{x}_{t/t}$ が補正後の数値解にあたる。

ここで特筆すべきは、カルマンフィルタのシミュレータ相当部分は、標本路を一切使うことなく、条件付き期待値 $\hat{x}_{t/t-1}$ を予測している。これにより、カルマンフィルタの持ち味である高速演算が実現するのだが、欠点がある。条件付き期待値を直接出力可能なシミュレータは、システムが線形でないと作れない。同様に、共分散シミュレータも無理である。

そこで、様々な近似を取り込む形で、カルマンフィルタの非線形対応がなされていった。代表的なものとしては、拡張カルマンフィルタ (EKF)、Unscented カルマンフィルタ (UKF)、アンサンブルカルマンフィルタ (EnKF) などがある²⁾。

本書の残りの部分では、両極端と思しき EKF と EnKF を取り上げたいと思う。EKF は、標本路を一切使わない方針を保ちながら、非線形項 $\langle f(x_t) \rangle$ を、大胆にも $f(\langle x_t \rangle)$ で近似してしまう手法である。一方で、EnKF は、非線形項 $\langle f(x_t) \rangle$ の値を 100 本程度の標本路で近似する手法である。

なお、いずれの手法も著者はまだ初心者の方際であり、以下の内容は、一読者による名著²⁾の学習ノートにすぎない。恥ずかしながら事前にご容赦いただきたい。

14.2 拡張カルマンフィルタ (EKF および EKBF)

14.2.1 問題設定

非線形確率システム (離散時間) の与え方には色々あるが、本書では次の形式を考える。

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_t, t) + \mathbf{G}_t \mathbf{w}_t, \\ \mathbf{y}_t = \mathbf{H}(\mathbf{x}_t, t) + \mathbf{v}_t, \\ \langle \mathbf{w}_t \rangle = \langle \mathbf{v}_t \rangle = \mathbf{0}, \quad \langle \mathbf{w}_t \mathbf{w}_s^T \rangle = \mathbf{Q}_t \delta_{ts}, \\ \langle \mathbf{v}_t \mathbf{v}_s^T \rangle = \mathbf{R}_t \delta_{ts}, \quad t = 0, 1, 2, \dots \end{cases} \quad (14.7)$$

ただし、 $\mathbf{x}_t \in \mathbb{R}^n$, $\mathbf{y}_t, \mathbf{v}_t \in \mathbb{R}^p$, $\mathbf{w}_t \in \mathbb{R}^r$ である。線形問題では行列 \mathbf{F}_t , \mathbf{H}_t だった部分が、非線形のベクトル値関数 $\mathbf{F}(\mathbf{x}_t, t)$, $\mathbf{H}(\mathbf{x}_t, t)$ に変わった。

14.2.2 非線形項の線形近似

EKF のアルゴリズムは、元になる KF (算法 11.6) に若干の変更を加えたものである。元になる KF を以下に示す。下線は、EKF で拡張される部分を表す。

$$\begin{aligned} \mathbf{K}_t &= \underline{\Sigma}_t \underline{H}_t^T (\underline{H}_t \underline{\Sigma}_t \underline{H}_t^T + \mathbf{R}_t)^{-1} \\ \hat{\mathbf{x}}_{t/t} &= \hat{\mathbf{x}}_t + \mathbf{K}_t (\mathbf{y}_t - \underline{H}_t \hat{\mathbf{x}}_t) \\ \hat{\mathbf{x}}_{t+1} &= \underline{F}_t \hat{\mathbf{x}}_{t/t} \\ \underline{\Sigma}_{t+1} &= \underline{F}_t \underline{\Sigma}_t \underline{F}_t^T + \mathbf{G}_t \mathbf{Q}_t \mathbf{G}_t^T - \underline{F}_t \mathbf{K}_t \underline{H}_t \underline{\Sigma}_t \underline{F}_t^T \end{aligned}$$

下線部は、(14.7) に対応する行列がないので、別なものを用意する必要がある。

これらの下線部は 2 つのタイプに分類できる。

- ベクトル値型・・・ $\underline{F}_t \hat{\mathbf{x}}_{t/t}$, $\underline{H}_t \hat{\mathbf{x}}_t$
- 行列単独型・・・ \underline{F}_t , \underline{H}_t

EKF では、これらを次のような考え方で拡張する。

(a) ベクトル値型 ベクトル値型は、状態/観測方程式の期待値に由来する。

$$\langle \mathbf{x}_{t+1} \rangle = \langle \mathbf{F}_t \mathbf{x}_t + \mathbf{G}_t \mathbf{w}_t \rangle = \langle \mathbf{F}_t \mathbf{x}_t \rangle \stackrel{\text{線形性}}{=} \mathbf{F}_t \langle \mathbf{x}_t \rangle = \mathbf{F}_t \hat{\mathbf{x}}_{t/t} \quad (14.8)$$

$$\langle \mathbf{y}_t \rangle = \langle \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t \rangle = \langle \mathbf{H}_t \mathbf{x}_t \rangle \stackrel{\text{線形性}}{=} \mathbf{H}_t \langle \mathbf{x}_t \rangle = \mathbf{H}_t \hat{\mathbf{x}}_t \quad (14.9)$$

しかし、同じことを、非線形システム (14.7) に適用しようとする、

$$\langle x_{t+1} \rangle = \langle F(x_t, t) + G_t w_t \rangle = \langle F(x_t, t) \rangle \stackrel{\text{線形性}}{=} \text{無理} \quad (14.10)$$

$$\langle y_t \rangle = \langle H(x_t, t) + v_t \rangle = \langle H(x_t, t) \rangle \stackrel{\text{線形性}}{=} \text{無理} \quad (14.11)$$

となり、途中で計算が止まってしまう。

そこで、EKF では、非線形関数 $F(x, t)$ のテイラー展開を利用する。そのために、確率ベクトル x を期待値とそれまわりの変動成分に分ける。

$$x := \langle x \rangle + \vec{x} \quad (14.12)$$

\vec{x} は変動を表す確率ベクトルである。これを代入して \vec{x} でテイラー展開し、2 次以上の項を無視すると、次の近似を得る。

$$F(x, t) \approx F(\langle x \rangle, t) + \frac{\partial F(\langle x \rangle, t)}{\partial x} \vec{x} \quad (14.13)$$

$\frac{\partial f}{\partial x}(\cdot)$ はヤコビ行列である。

▶▶ (ヤコビ行列) $f(x)$ の i 成分を $f_i(x_1, \dots, x_n)$ とする。このとき、偏微分係数 $\partial f_i / \partial x_j$ を ij 成分とする行列をヤコビ行列といい、次のように書く。

$$\frac{\partial f(a)}{\partial x} \quad (14.14)$$

本書では、上式は、偏微分した後に $x = a$ を代入したヤコビ行列を表す。

両辺の期待値をとり、 $\langle \vec{x} \rangle$ が十分に小さいと仮定すると、

$$\langle F(x, t) \rangle \approx F(\langle x \rangle, t) \quad (14.15)$$

$$\text{同様にして、} \langle H(x, t) \rangle \approx H(\langle x \rangle, t) \quad (14.16)$$

は、妥当な近似となる。

この近似方法は、確率論的には、2 次モーメント $\langle x^2 \rangle$ を、期待値の 2 乗 $\langle x \rangle^2$ で近似するような暴挙なのだが[†]、フィルタに搭載すると、不思議と上手く機能するのである。

(b) 行列単独型 行列単独型は、状態/観測方程式に含まれる線形変換に由来する。

$$x' = F_t x, \quad x' = H_t x \quad (14.17)$$

このうちの行列部分 F_t, H_t は、値ではなく作用を表す。非線形システム (14.7) における相当物は、非線形変換である。

[†] x が平均 0 のガウス確率変数だとすると、 $\langle x^2 \rangle = \langle x \rangle^2$ ならば x の分散は 0 になる。すなわち、 x は原点に集中しており、ランダムとはいえない。

$$\mathbf{x}' = \mathbf{F}(\mathbf{x}, t), \quad \mathbf{x}' = \mathbf{H}(\mathbf{x}, t) \quad (14.18)$$

これらの値の近似ではなく，作用 $\mathbf{F}(\bullet, t)$, $\mathbf{H}(\bullet, t)$ の近似を求めたい．

十分に小さなベクトル \vec{x} を用いて， $\mathbf{x} := \mathbf{a} + \vec{x}$ と書けるとき， \mathbf{x} は \mathbf{a} の近傍にある．これを $\mathbf{x}' = \mathbf{F}(\mathbf{x}, t)$ に代入した結果を， $\mathbf{x}' := \mathbf{F}(\mathbf{a}, t) + \vec{x}'$ で表す．ここで (14.13) を使うと，

$$\mathbf{F}(\mathbf{a}, t) + \vec{x}' = \mathbf{x}' = \mathbf{F}(\mathbf{a} + \vec{x}, t) \approx \mathbf{F}(\mathbf{a}, t) + \frac{\partial \mathbf{F}(\mathbf{a}, t)}{\partial \mathbf{x}} \vec{x} \quad (14.19)$$

となるが，最初と最後の $\mathbf{F}(\mathbf{a}, t)$ が相殺されて，次の近似式が得られる．

$$\vec{x}' \approx \frac{\partial \mathbf{F}(\mathbf{a}, t)}{\partial \mathbf{x}} \vec{x} \quad (14.20)$$

この近似式により，(14.18) の作用は，ヤコビ行列で次のように近似できる．

$$\mathbf{x} = \mathbf{a} \text{ の近傍では } \mathbf{F}(\bullet, t) \approx \frac{\partial \mathbf{F}(\mathbf{a}, t)}{\partial \mathbf{x}} \bullet \quad (14.21)$$

$$\mathbf{x} = \mathbf{b} \text{ の近傍では } \mathbf{H}(\bullet, t) \approx \frac{\partial \mathbf{H}(\mathbf{b}, t)}{\partial \mathbf{x}} \bullet \quad (14.22)$$

14.2.3 基本アルゴリズム

前節の結果 (14.15)，(14.16)，(14.21)，(14.22) を参考に，KF の各項を 2 種類のルール

$$F_t \hat{\mathbf{x}}_{t/t} \mapsto \mathbf{F}(\hat{\mathbf{x}}_{t/t}, t), \quad H_t \hat{\mathbf{x}}_t \mapsto \mathbf{H}(\hat{\mathbf{x}}_t, t), \quad (14.23)$$

$$F_t \mapsto \frac{\partial \mathbf{F}(\hat{\mathbf{x}}_{t/t}, t)}{\partial \mathbf{x}}, \quad H_t \mapsto \frac{\partial \mathbf{H}(\hat{\mathbf{x}}_t, t)}{\partial \mathbf{x}} \quad (14.24)$$

で置き換えたものが，EKF である．次のように整理できる．

算法 14.1 (EKF) 非線形確率システム (離散時間)：

$$\begin{cases} \mathbf{x}_{t+1} = \mathbf{F}(\mathbf{x}_t, t) + G_t \mathbf{w}_t, & \mathbf{x}_t \in \mathbb{R}^n, \mathbf{w}_t \in \mathbb{R}^r \\ \mathbf{y}_t = \mathbf{H}(\mathbf{x}_t, t) + \mathbf{v}_t, & \mathbf{y}_t, \mathbf{v}_t \in \mathbb{R}^p, t = 0, 1, 2, \dots \end{cases} \quad (14.25)$$

の最小分散濾波推定値 $\hat{\mathbf{x}}_{t/t}$ は，次のように逐次計算できる．

(0) 初期条件：

$$\hat{\mathbf{x}}_0 = \bar{\mathbf{x}}_0, \quad \Sigma_0 = \Sigma_0 \quad (14.26)$$

(1) 濾波推定： $\mathbf{y}_t, \hat{\mathbf{x}}_t, \Sigma_t$ を既知として，

$$H_t = \frac{\partial \mathbf{H}(\hat{\mathbf{x}}_t, t)}{\partial \mathbf{x}} \quad \text{【線形化観測行列】} \quad (14.27)$$

$$K_t = \Sigma_t H_t^T (H_t \Sigma_t H_t^T + R_t)^{-1} \text{【カルマンゲイン】} \quad (14.28)$$

$$\hat{x}_{t/t} = \hat{x}_t + K_t \{y_t - H(\hat{x}_t, t)\} \text{【濾波推定値】} \quad (14.29)$$

ただし，逆行列が存在しない場合は，疑似逆行列を用いる．

(2) 予測推定：

$$F_t = \frac{\partial F(\hat{x}_{t/t}, t)}{\partial x} \text{【線形化状態行列】} \quad (14.30)$$

$$\hat{x}_{t+1} = F(\hat{x}_{t/t}, t) \text{【予測推定値】} \quad (14.31)$$

$$\Sigma_{t+1} = F_t \Sigma_t F_t^T + G_t Q_t G_t^T - F_t K_t H_t \Sigma_t F_t^T \text{【予測推定誤差】} \quad (14.32)$$

同様の考え方で，KBF を非線形拡張したものを EKBF という．

算法 14.2 (EKBF) 非線形確率システム (連続時間)：

$$\begin{cases} \dot{x} = A(x, t) + D(t)w(t), & x(t) \in \mathbb{R}^n, w(t) \in \mathbb{R}^r, \\ y = C(x, t) + v(t), & y(t), v(t) \in \mathbb{R}^p \end{cases} \quad (14.33)$$

$$\begin{cases} \langle w(t) \rangle = 0, & \langle w(t)w(s)^T \rangle = Q(t)\delta(t-s), \\ \langle v(t) \rangle = 0, & \langle v(t)v(s)^T \rangle = R(t)\delta(t-s) \quad (s \geq t) \end{cases} \quad (14.34)$$

の最小分散濾波推定値 $\hat{x}(t)$ は，次の微分方程式系の解として得られる．すなわち，既知の観測量 $y(t)$ と，適当な初期条件 $\hat{x}(0) = \bar{x}_0$ ， $\Sigma(0) = \Sigma_0$ に対して，

$$C(t) := \frac{\partial C(\hat{x}, t)}{\partial x} \text{ (線形化観測行列)} \quad (14.35a)$$

$$K(t) := \Sigma C(t)^T R(t)^{-1} \text{ (カルマンゲイン)} \quad (14.35b)$$

$$\dot{\hat{x}} = A(\hat{x}, t) + K(t)\{y - C(\hat{x}, t)\} \text{ (濾波推定値)} \quad (14.35c)$$

$$A(t) := \frac{\partial A(\hat{x}, t)}{\partial x} \text{ (線形化状態行列)} \quad (14.35d)$$

$$\dot{\Sigma} = A(t)\Sigma + \Sigma A(t)^T + D(t)Q(t)D(t)^T - K(t)C(t)\Sigma \text{ (誤差)} \quad (14.35e)$$

14.2.4 数 値 例

EKF と EKBF は非線形項を受け付けるので，さきほど断念した減衰係数 c の推定問題 (14.6) を解くことができる．算法 14.2 の EKBF の形式で書くと，

$$\dot{\mathbf{x}} = \mathbf{A}(\mathbf{x}, t) + D\mathbf{w}(t), \quad y = C(\mathbf{x}, t) + v(t), \quad \mathbf{x} = (x_1, x_2, x_3)^T := (x, \dot{x}, c)^T,$$

$$\mathbf{A}(\mathbf{x}, t) := \begin{pmatrix} x_2 \\ -k x_1 - x_3 x_2 + a \sin(\beta t) \end{pmatrix}, \quad D := \text{単位行列}, \quad C(\mathbf{x}, t) := x_1 \quad (14.36)$$

となる．

図 14.4 は，(14.36) を EKF で解いた結果である．問題が連続時間，フィルタが離散時間なので，算法 12.2 を介して求めた．同じく，図 14.5 は EKBF による結果である．

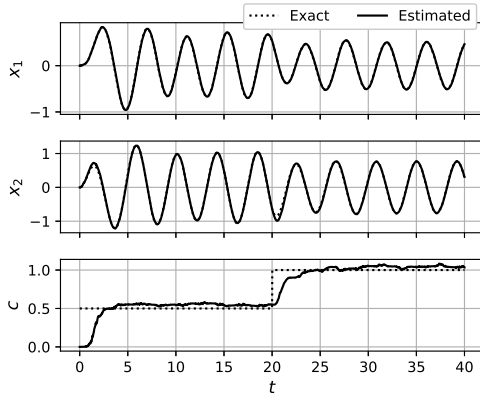


図 14.4 減衰係数 c の推定 (EKF)

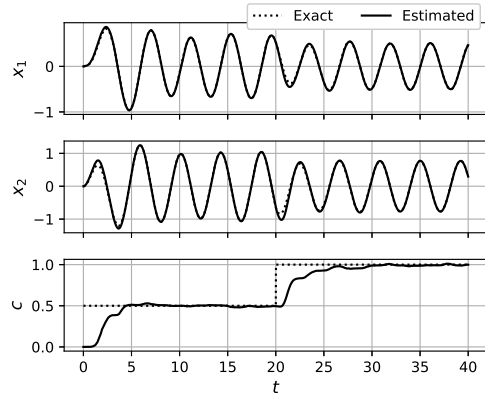


図 14.5 減衰係数 c の推定 (EKBF)

この結果においては，離散時間型と連続時間型で差が出ている．連続時間型の EKBF による推定値のほうが，真値に近く，安定している．

14.3 アンサンブルカルマンフィルタ (EnKF)

14.3.1 基本的なアイデア

標本路を一切使わない EKF や EKBF とは逆に，EnKF では，期待値や共分散の時間的推移を，標本路から求めてしまう．すなわち，生の状態方程式 $\mathbf{x}_{t+1} = \mathbf{F}(\mathbf{x}_t, t) + G\mathbf{w}_t$ から多数の数値解を生成して，その標本平均や標本分散を利用する．これなら非線形でも問題ない．

標本路を使うのは，カルマンゲインの部分である．(11.44) を再掲する．

$$K_t = \underbrace{\left\langle \bar{\mathbf{x}}_{t/t-1} \bar{\mathbf{y}}_{t/t-1}^T \right\rangle}_{\text{データ表現}} \left\langle \bar{\mathbf{y}}_{t/t-1} \bar{\mathbf{y}}_{t/t-1}^T \right\rangle^{-1}$$

$$= \underbrace{\Sigma_{t/t-1} H_t^T (R_t + H_t \Sigma_{t/t-1} H_t^T)^{-1}}_{\text{行列表現}} \quad (14.37)$$

通常のカルマンフィルタ (KF や KBF) では、後半の行列表現を使うことで、アルゴリズムから生の標本路を排除した (観測量 \mathbf{y}_t だけは例外) . しかしそれでは、非線形だと、 $\Sigma_{t/t-1}$ の漸化式が使えなくなる . そこで拡張カルマンフィルタ (UKF や UKBF) では、非線形関数をヤコビ行列で近似することにより、 $\Sigma_{t/t-1}$ の漸化式を復活させた .

どうせ近似するなら、別の方法もあるんじゃない? というわけで、EnKF では、前半のデータ表現のほうを用いて、これを数値シミュレーションで近似する . するともう何も要らない、といったら語弊があるが、何らかの方法で $\langle \tilde{\mathbf{x}}_{t/t-1} \tilde{\mathbf{y}}_{t/t-1}^T \rangle$, $\langle \tilde{\mathbf{y}}_{t/t-1} \tilde{\mathbf{y}}_{t/t-1}^T \rangle$ さえ計算できれば用が足りることになる .

14.3.2 EnKF のアルゴリズム

そのための手順²⁾ を以下に示す . 非線形確率システム (14.7) を対象とする .

(0) 初期化ステップ (14.7) の初期値 \mathbf{x}_0 をランダムに生成し、 M 個の標本をとる .

$$\mathcal{X}_0 := \left\{ \mathbf{x}_0^{(1)}, \mathbf{x}_0^{(2)}, \dots, \mathbf{x}_0^{(\omega)}, \dots, \mathbf{x}_0^{(M)} \right\} \quad (14.38)$$

$\omega = 1, 2, \dots, M$ は標本番号である . 標本の分布は問題に即して定める . 以下では、慣例に従って、こうした標本を「アンサンブル」と呼び、各標本点を「粒子」と呼ぶ .

EnKF の世界では、初期値のアンサンブル (14.38) は次の添字で表記され、予測推定値 $\hat{\mathbf{x}}_{t/t-1}|_{t=0}$ を表すアンサンブル、とみなされる .

$$\mathcal{X}_{0/-1} := \left\{ \mathbf{x}_{0/-1}^{(1)}, \mathbf{x}_{0/-1}^{(2)}, \dots, \mathbf{x}_{0/-1}^{(\omega)}, \dots, \mathbf{x}_{0/-1}^{(M)} \right\} = \mathcal{X}_0 \quad (14.39)$$

このアンサンブルの重心が $\hat{\mathbf{x}}_{t/t-1}|_{t=0}$ を表すことになる .

(1) 観測更新ステップ 予測推定アンサンブル $\mathcal{X}_{0/-1}$ を、出力方程式に代入すると、

$$\mathbf{y}_{0/-1}^{(\omega)} = \mathbf{H}(\mathbf{x}_{0/-1}^{(\omega)}, 0) + \mathbf{v}_0^{(\omega)}, \quad \omega = 1, 2, \dots, M \quad (14.40)$$

を要素とする予測出力アンサンブルが得られる .

$$\mathcal{Y}_{0/-1} := \left\{ \mathbf{y}_{0/-1}^{(1)}, \mathbf{y}_{0/-1}^{(2)}, \dots, \mathbf{y}_{0/-1}^{(\omega)}, \dots, \mathbf{y}_{0/-1}^{(M)} \right\} \quad (14.41)$$

以上で、1 段目の数値シミュレーションは完了である .

得られた $\mathcal{X}_{0/-1}$, $\mathcal{Y}_{0/-1}$ を用いて、 $\langle \tilde{\mathbf{x}}_{0/-1} \tilde{\mathbf{y}}_{0/-1}^T \rangle$, $\langle \tilde{\mathbf{y}}_{0/-1} \tilde{\mathbf{y}}_{0/-1}^T \rangle$ の近似値を求める . まず、それぞれの標本平均 (重心) によって、 $\hat{\mathbf{x}}_{0/-1}$, $\hat{\mathbf{y}}_{0/-1}$ を近似する .

$$\hat{\mathbf{x}}_{0/-1} \approx \langle \mathcal{X}_{0/-1} \rangle := \frac{1}{M} \sum_{\omega=1}^M \mathbf{x}_{0/-1}^{(\omega)}, \quad \hat{\mathbf{y}}_{0/-1} \approx \langle \mathcal{Y}_{0/-1} \rangle := \frac{1}{M} \sum_{\omega=1}^M \mathbf{y}_{0/-1}^{(\omega)} \quad (14.42)$$

さらに, $\mathcal{X}_{0/-1}$, $\mathcal{Y}_{0/-1}$ の不偏共分散により, $\langle \vec{x}_{0/-1} \vec{y}_{0/-1}^T \rangle$, $\langle \vec{y}_{0/-1} \vec{y}_{0/-1}^T \rangle$ を近似する.

$$\begin{aligned} \langle \vec{x}_{0/-1} \vec{y}_{0/-1}^T \rangle &\approx \text{Cov}[\mathcal{X}, \mathcal{Y}]_{0/-1} \\ &:= \frac{1}{M-1} \sum_{\omega=1}^M \left(x_{0/-1}^{(\omega)} - \langle \mathcal{X}_{0/-1} \rangle \right) \left(y_{0/-1}^{(\omega)} - \langle \mathcal{Y}_{0/-1} \rangle \right)^T \\ \langle \vec{y}_{0/-1} \vec{y}_{0/-1}^T \rangle &\approx \text{Cov}[\mathcal{Y}, \mathcal{Y}]_{0/-1} \\ &:= \frac{1}{M-1} \sum_{\omega=1}^M \left(y_{0/-1}^{(\omega)} - \langle \mathcal{Y}_{0/-1} \rangle \right) \left(y_{0/-1}^{(\omega)} - \langle \mathcal{Y}_{0/-1} \rangle \right)^T \end{aligned} \quad (14.43)$$

この時点で, アンサンブルに基づくカルマンゲインが, 次のように定まる.

$$K_0 = \langle \vec{x}_{t/t-1} \vec{y}_{t/t-1}^T \rangle \langle \vec{y}_{t/t-1} \vec{y}_{t/t-1}^T \rangle^{-1} \approx \text{Cov}[\mathcal{X}, \mathcal{Y}]_{0/-1} \left(\text{Cov}[\mathcal{Y}, \mathcal{Y}]_{0/-1} \right)^{-1} \quad (14.44)$$

この K_0 を用いて, 現時刻の観測値 y_0 に対する各粒子の濾波推定値を, 次式で求め,

$$\mathbf{x}_{0/0}^{(\omega)} := \mathbf{x}_{0/-1}^{(\omega)} + K_0 \left(y_0 - \mathbf{y}_{0/-1}^{(\omega)} \right), \quad \omega = 1, 2, \dots, M \quad (14.45)$$

これらを要素とする濾波アンサンブルを構成する.

$$\mathcal{X}_{0/0} := \left\{ \mathbf{x}_{0/0}^{(1)}, \mathbf{x}_{0/0}^{(2)}, \dots, \mathbf{x}_{0/0}^{(\omega)}, \dots, \mathbf{x}_{0/0}^{(M)} \right\} \quad (14.46)$$

この標本平均が, EnKF による濾波推定値とされる.

$$\hat{\mathbf{x}}_{0/0} := \langle \mathcal{X}_{0/0} \rangle \quad (14.47)$$

(2) 時間更新ステップ 2 段目の数値シミュレーションとして, 濾波アンサンブル $\mathcal{X}_{0/0}$ を初期値として (14.7) を解く. その結果,

$$\mathbf{x}_{1/0}^{(\omega)} = \mathbf{F} \left(\mathbf{x}_{0/0}^{(\omega)}, 0 \right) + G_0 \mathbf{w}_0^{(\omega)}, \quad \omega = 1, 2, \dots, M \quad (14.48)$$

を要素とする, 次時刻の予測推定アンサンブルが得られる.

$$\mathcal{X}_{1/0} := \left\{ \mathbf{x}_{1/0}^{(1)}, \mathbf{x}_{1/0}^{(2)}, \dots, \mathbf{x}_{1/0}^{(\omega)}, \dots, \mathbf{x}_{1/0}^{(M)} \right\} \quad (14.49)$$

この $\mathcal{X}_{1/0}$ に対して, 再び (14.40) ~ (14.49) の計算を繰り返すと, 次の濾波推定値 $\hat{\mathbf{x}}_{1/1}$ と予測推定アンサンブル $\mathcal{X}_{2/1}$ が得られる. 以下, 帰納的に, 濾波推定値の列, $\hat{\mathbf{x}}_{0/0}, \hat{\mathbf{x}}_{1/1}, \dots, \hat{\mathbf{x}}_{t/t}, \dots$ が得られる ($\mathcal{X}_{t/t-1}$ の列とともに).

以上が, EnKF による濾波推定アルゴリズムである.

14.3.3 行列計算による EnKF 実装へのヒント

EnKF の計算式は、ちょっとマニアックな行列形式で書かれることが多いような気がする。高速な行列演算ライブラリを意識したものと思われるが、ちょっと真似してみる。

アンサンブル $\mathcal{X} = \{x^{(1)}, x^{(2)}, \dots, x^{(M)}\}$ の全成分を一括で保持するために、各粒子 $x^{(\omega)}$ を列ベクトルとする次の行列を導入する。

$$X := \begin{pmatrix} x^{(1)} & x^{(2)} & \dots & x^{(M)} \end{pmatrix} \quad (14.50)$$

これを、アンサンブル行列という。その標本平均は、次の行列演算で得られる。

$$\langle X \rangle = \frac{1}{M} X \mathbb{I} \quad (14.51)$$

ただし、 $\mathbb{I} := (1, 1, \dots, 1)^T$ である。不偏共分散は次のように得られる。

$$\begin{aligned} \text{Cov}[X, Y] &= \frac{M}{M-1} \left(\frac{1}{M} X Y^T - \langle X \rangle \langle Y \rangle^T \right) \\ &= \frac{1}{M-1} X Y^T - \frac{M}{M-1} \langle X \rangle \langle Y \rangle^T \end{aligned} \quad (14.52)$$

以上の表記で整理すると、EnKF の計算方法は、次のように簡潔に表せる[†]。

算法 14.3 (EnKF) 非線形確率システム

$$x_{t+1} = F(x_t) + G_t w_t, \quad y_t = H(x_t, t) + v_t$$

の濾波推定値 $\hat{x}_{t/t}$ は、次のように逐次計算できる。

(1) 観測更新ステップ

- 前時刻の予測推定アンサンブルを既知とする。

$$X_{t/t-1} := \begin{pmatrix} x_{t/t-1}^{(1)} & x_{t/t-1}^{(2)} & \dots & x_{t/t-1}^{(M)} \end{pmatrix}$$

- 予測出力アンサンブルを求める。

$$Y_{t/t-1} = \begin{pmatrix} y_{t/t-1}^{(1)} & \dots & y_{t/t-1}^{(M)} \end{pmatrix}, \quad y_{t/t-1}^{(\omega)} = H(x_{t/t-1}^{(\omega)}, t) + v_t^{(\omega)}$$

- $X_{t/t-1}, Y_{t/t-1}$ の標本平均と不偏共分散を求める。 $\mathbb{I} := (1, 1, \dots, 1)^T$ とする。

$$\begin{aligned} \langle X_{t/t-1} \rangle &= \frac{1}{M} X_{t/t-1} \mathbb{I}, \quad \langle Y_{t/t-1} \rangle = \frac{1}{M} Y_{t/t-1} \mathbb{I}, \\ \text{Cov}[X, Y]_{t/t-1} &= \frac{X_{t/t-1} Y_{t/t-1}^T}{M-1} - \frac{M \langle X_{t/t-1} \rangle \langle Y_{t/t-1} \rangle^T}{M-1}, \\ \text{Cov}[Y, Y]_{t/t-1} &= \frac{Y_{t/t-1} Y_{t/t-1}^T}{M-1} - \frac{M \langle Y_{t/t-1} \rangle \langle Y_{t/t-1} \rangle^T}{M-1} \end{aligned}$$

[†] こんな書き方よく見ますよね。

- カルマンゲイン K_t を求める .

$$K_t = \text{Cov}[X, Y]_{t/t-1} \left(\text{Cov}[Y, Y]_{t/t-1} \right)^{-1}$$

- 観測値 y_t に基づき、濾波アンサンブルを求める .

$$X_{t/t} = \left(\mathbf{x}_{t/t}^{(1)}, \mathbf{x}_{t/t}^{(2)}, \dots, \mathbf{x}_{t/t}^{(M)} \right), \quad \mathbf{x}_{t/t}^{(\omega)} = \mathbf{x}_{t/t-1}^{(\omega)} + K_t \left(y_t - \mathbf{y}_{t/t-1}^{(\omega)} \right)$$

- 濾波推定値を得る .

$$\hat{\mathbf{x}}_{t/t} = \langle X_{t/t} \rangle = \frac{1}{M} X_{t/t} \mathbb{I}$$

(2) 時間更新ステップ

- 次時刻の予測アンサンブルを求める .

$$X_{t+1/t} := \left(\mathbf{x}_{t+1/t}^{(1)}, \mathbf{x}_{t+1/t}^{(2)}, \dots, \mathbf{x}_{t+1/t}^{(M)} \right), \quad \mathbf{x}_{t+1/t}^{(\omega)} = \mathbf{F} \left(\mathbf{x}_{t/t}^{(\omega)}, t \right) + G_t \mathbf{w}_t^{(\omega)}$$

(1) 観測更新ステップへ戻る

14.3.4 数 値 例

図 14.2, 図 14.4 と同じ問題を EnKF で解いた結果を, 図 14.6, 図 14.7 に示す. EnKF は離散時間, 問題は連続時間なので, 算法 12.2 で問題を変換してから解いた.

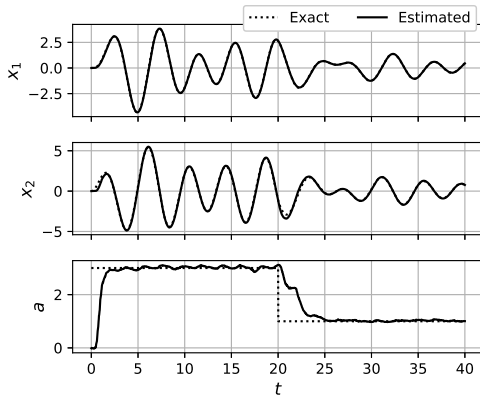


図 14.6 外力振幅 a の推定 (EnKF)

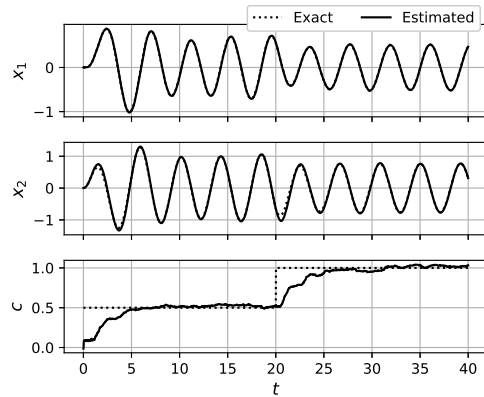


図 14.7 減衰係数 c の推定 (EnKF)

得られた推定精度としては, 離散時間型の仲間である EKF よりも, 連続時間型の EKBF に近い結果となった.

EnKF の計算時間は、いずれの結果も 60 s 前後を要した。これは EKF の 300 倍、EKBF の 24 倍である。すなわち、標本路を用いる EnKF は、計算速度についてはあまり期待しない手法なのだと思う。それと引き換えに、EnKF では、システムのヤコビ行列を求める必要がない。これは大規模システムではメリットとなるようで、EnKF をネットで検索すると、気象予測分野の情報が多くヒットする。

なお、本書で触れなかった Unsented カルマンフィルタは、両者の中間的な性質を持つようである。

付 録

A.1 集合論の概要

確率論の理解には、集合論の基礎知識が不可欠である。多数回の試行結果を、一括で表現したり、操作したりするからだ。ここでは、集合論に不慣れな読者のために、簡単な手引を用意しておく。ごく初等的な内容なので、集合論に慣れている人達は読まなくても大丈夫である。

A.1.1 数学的な集合とは？

数学的でいうところの集合、すなわち数学的に演算可能な集合とは、「非メンバーと明確に線引きされた、メンバーの集まり」のことである。 x が集合 X の要素（メンバー）であることを、 $x \in X$ と書く。このとき、「 $x \in X$ 」という主張は、本当か嘘か（真か偽か）の二者択一になる。 $x \in X$ が嘘であることを、 $x \notin X$ と書く（ x は X の要素でないと読む）。

A.1.2 集合の表記方法

集合の表記方法には、大きく 2 種類ある。

【外延的表記】1 つめは、要素を具体的に列挙する方式である。列挙した要素は中括弧で囲う。

$$A := \{x_1, x_2, x_3\}, \quad B := \{x_3, x_4, x_5, x_6\} \quad (\text{A.1})$$

あるいは、列挙したふりをする表記もよく用いられる。

$$N := \{1, 2, 3, \dots\} \quad (\text{A.2})$$

【内包的表記】2 つめは、要素の選抜条件を書く方式である。

$$X := \{x \mid x \text{ が満たすべき条件}\} \quad (\text{A.3})$$

この表記は、全要素が列挙できない場合にも使える。例えば、

$$C := \{c \mid c \text{ は人間で、悪玉コレステロール値が } 120 \text{ 以上}\} \quad (\text{A.4})$$

などは数学的に演算可能な集合となる。該当者を全て列挙するのは現実的には不可能だが、条件は客観的なので、「 $x \in C$ 」という主張は、本当か嘘かの二者択一になる。ゆえに C は集合である。例えば、「 $\text{犬} \in C$ 」は嘘であり、「 $\text{著者} \in C$ 」は本当である[†]。

次の集合と記号がよく用いられる。本書でも用いる。

- $N := \{n \mid n \text{ は自然数}\}$ （自然数の全体集合）
- $R := \{x \mid x \text{ は実数}\}$ （実数の全体集合）

[†] なんてね・・・

- $[a, b] := \{x \mid a \leq x \leq b, x \in \mathbb{R}\}$ (閉区間)
- $(a, b) := \{x \mid a < x < b, x \in \mathbb{R}\}$ (开区間)
- $[a, b) := \{x \mid a \leq x < b, x \in \mathbb{R}\}$ (半开区間)

内包的表記より, $x \in \mathbb{R}$ は, n が実数であることを意味する. このことから, 「 x は実数である」と書く代わりに「 $x \in \mathbb{R}$ 」と書ける. 同様に, 「 x は a 以上 b 以下の数である」は「 $x \in [a, b]$ 」と書ける.

▶▶ 开区間 (a, b) は, 2 次元横ベクトルと同じ表記だが, 慣例なのでしかたない. 文脈で判断する.

その他の表記として, 空の集合 $\{\}$ を \emptyset と書く. また, $\{a\}$ と a は意味が異なるので注意が必要だ. $\{a\}$ は独りぼっちの集合であり, a はむき出しの要素である. 独りぼっちだからといって, $\{a\}$ を a と混同してはならない. 例えば,

- 【誤った表現】 $a \subset A, \{a\} \in A$
- 【正しい表現】 $a \in A, \{a\} \subset A$

である. “ \subset ” の使い方は次節で述べる.

A.1.3 等号と部分集合

集合どうしの等号 (イコール) を導入したい. まず, 外延的表記の場合は, 全要素を照合すればよい. 構成メンバーが同じであれば, 同じ集合である. 例えば,

$$\{a, b, c\} = \{a, c, b\} \quad (\text{A.5})$$

である. 要素を表記する順序は問われない. 構成メンバーが同じなら同じ集合である.

その一方で, 内包的表記において, 全数検査が不可能な場合は, 次の定義が用いられる.

$$A = B \stackrel{\text{定義}}{\iff} \text{「} A \text{ の任意の要素を } B \text{ が含み, かつ } B \text{ の任意の要素を } A \text{ が含むこと」} \quad (\text{A.6})$$

また, カギ括弧内の条件を片方ずつ使ったものを, 部分集合と呼び, 次のように表記する.

$$A \subset B \stackrel{\text{定義}}{\iff} \text{「} A \text{ の任意の要素を } B \text{ が含むこと」} \quad (\text{A.7})$$

$$A \supset B \stackrel{\text{定義}}{\iff} \text{「} B \text{ の任意の要素を } A \text{ が含むこと」} \quad (\text{A.8})$$

したがって, 「 $A = B$ 」は「 $A \subset B$ かつ $A \supset B$ 」と同じことである.

以上より, 内包的表記の集合 $A := \{x \mid x \text{ は条件 } A \text{ を満たす}\}, B := \{x \mid x \text{ は条件 } B \text{ を満たす}\}$ に対して, $A = B$ を証明するには, 次のような 2 段階の手順を踏めばよい.

1. 条件 A を満足する x を任意にとる. この x が条件 B を満足することを示す. ($A \subset B$ の証明)
(内包的表記より, 条件 A を満足する x は, 集合 A の要素に他ならない. 以下同じ)
 2. 条件 B を満足する x を任意にとる. この x が条件 A を満足することを示す. ($B \subset A$ の証明)
- これにより, $A \subset B$ と $A \supset B$ の両立が示されれば, $A = B$ が証明されたことになる.

簡単な例で試してみよう. 次の集合 A, B について, $A = B$ の成立を示せ.

$$A := \{x \mid (x+1)(x-1) < 0\}, \quad B := \{x \mid -1 < x < 1\} \quad (\text{A.9})$$

▶ 証明 $(x+1)(x-1) < 0$ を満足する任意の x をとると, これは $-1 < x < 1$ を満足する (2 次不等式の解). ゆえに, この x は集合 B の条件を満足するので, $A \subset B$ が示された. 逆に, 任意の $-1 < x < 1$ をとると, これは集合 A の条件を満足する (2 次不等式の解). ゆえに $B \subset A$ も示された. したがって, $A = B$ が示された.

A.1.4 集 合 演 算

集合から別の集合を作る演算を集合演算という．最も基本的なものとして，集合和 (合併) \cup ，集合積 (共通部分) \cap ，補集合 $\bar{}$ ，集合差 \setminus が広く用いられている．外延的表記の集合演算は言葉通りなので，以下，内包的表記の場合を考える．

$$A_1 := \{x \mid x \text{ は条件 1 を満たす}\}, \quad A_2 := \{x \mid x \text{ は条件 2 を満たす}\} \quad (\text{A.10})$$

これらの集合演算を，内包的表記で表すと次のようになる．

$$\text{集合和 } A_1 \cup A_2 := \{x \mid x \text{ は条件 1 を満たす, または, } x \text{ は条件 2 を満たす}\} \quad (\text{A.11})$$

$$\text{集合積 } A_1 \cap A_2 := \{x \mid x \text{ は条件 1 を満たす, かつ, } x \text{ は条件 2 を満たす}\} \quad (\text{A.12})$$

$$\begin{aligned} \text{補集合 } \bar{A}_1 &:= \{x \mid x \text{ は条件 1 を満たさない}\}, \\ \bar{A}_2 &:= \{x \mid x \text{ は条件 2 を満たさない}\} \end{aligned} \quad (\text{A.13})$$

$$\begin{aligned} \text{集合差 } A_1 \setminus A_2 &:= A_1 \cap \bar{A}_2 \\ &= \{x \mid x \text{ は条件 1 を満たす, かつ, } x \text{ は条件 2 を満たさない}\} \end{aligned} \quad (\text{A.14})$$

このように，内包的表記に対する集合演算は，条件部分の論理演算 (または，かつ，否定) に帰着する．

A.1.5 直 積 集 合

2 つの集合 A, B の要素の対 (a, b) の全体を，次のように表記する．

$$A \times B := \{(a, b) \mid a \in A, b \in B\} \quad (\text{A.15})$$

これを， A, B の直積集合と呼ぶ．

▶▶ 開区間 (a, b) と同じ表記だが，慣例なのでしかたない．文脈で判断する．

同様に，

$$A \times B \times C := \{(a, b, c) \mid a \in A, b \in B, c \in C\} \quad (\text{A.16})$$

は，3 つ組 (a, b, c) の全体を表す．

よく用いられる表現として次のようなものがある．これは本来は，実数の n 個組の全体を表す．

$$R^n := \underbrace{R \times R \times \cdots \times R}_{n \text{ 個}} = \{(x_1, x_2, \cdots, x_n) \mid x_1, x_2, \cdots, x_n \in R\} \quad (\text{A.17})$$

これを n 次元数ベクトル空間とするには，等号，加法，スカラー倍を付与する必要があるが，これらを暗黙に仮定して，単に R^n と書いて n 次元数ベクトル空間を表す風習がある．

また，区間の直積 $[a, b] \times [c, d]$ は，矩形領域 (長方形の領域) を表す．実際，定義をたどると，

$$\begin{aligned} [a, b] \times [c, d] &= \{(x, y) \mid x \in [a, b], y \in [c, d]\} \\ &= \{(x, y) \mid a \leq x \leq b, c \leq y \leq d\} \end{aligned} \quad (\text{A.18})$$

となる．

A.1.6 限定記号

実は、次の等式は、厳密にいうと意味が定まらない。値が 0 の定数関数を意味するのか、特定の x で 0 となる方程式なのか、これだけでは判断できない。

$$f(x) = 0 \quad (\text{A.19})$$

このようなときに、次の記号が使えると便利である。

- 全称記号 $\forall x \cdots$ 全ての x について、任意の x について、どの x についても
- 存在記号 $\exists x \cdots$ ある x が (1 つ以上) 存在して、適当な x を選べば

これらを使うと、定数関数は、例えば次のように表示できる。

$$f(x) = 0, \quad \forall x \in \mathbb{R} \quad (\text{A.20})$$

他方、零点を (少なくとも 1 つ) 有する関数は、例えば次のように表示できる。

$$f(x) = 0, \quad \exists x \in \mathbb{R} \quad (\text{A.21})$$

表記上は些細な違いだが、意味するところは大きく異なる。

例題を挙げておこう。次の式は本当か嘘か？ 判定せよ。

$$\sin x = 0, \quad \forall x \in \mathbb{R} \quad (\text{A.22})$$

$$\sin x = 0, \quad \exists x \in \mathbb{R} \quad (\text{A.23})$$

日本語で読み下すと、上の式は「全ての実数 x について $\sin x = 0$ となる」、下の式は「適当な実数 x を選べば、 $\sin x = 0$ とできる」。sin のグラフ形状から、それぞれの真偽は明らかであろう。

引用・参考文献

- 1) 片山徹. 応用カルマンフィルタ. 朝倉書店, 2000.
- 2) 片山徹. 非線形カルマンフィルタ. 朝倉書店, 2011.
- 3) A. Papoulis. 確率とランダム変数. パボリス応用確率論. 東海大学出版会, 1992. 中山謙二, 根本幾, 町田東一訳.
- 4) 津野義道. Kalman-Bucy のフィルター理論. 共立出版, 2006.
- 5) 久米均. 統計解析への出発. シリーズ入門統計的方法. 岩波書店, 1989.
- 6) 砂原善文. 確率システム理論. コロナ社, 1979.
- 7) 加藤一郎. 最適制御入門: レギュレータとカルマン・フィルタ. 東京大学出版会, 1987.

本書の内容は、確かに上記の本で勉強したのですが、どの本で勉強したか特定できず、引用不十分となってしまった箇所があるように思います。小生の不見識の致すところであり、何卒ご容赦ください。