–分かる確率統計とカルマンフィルタ– [1]

杉本末雄
立命館大学 名誉教授

2021 年 6 月 24 日

[1]@sueo.sugimoto 無断転載を禁じる

目 次

第1章　確率・統計学の基礎

1.1　確率とは

新聞には，明日の大阪の午前の降水確率は 30%(= 0.3) であると記載されたり，硬貨を 1 回投げると,「表」の出る確率は $\frac{1}{2}$ であるとか,「確率」(probability) という言葉が日常よく使用されているが，あらためて「確率」とは何ですかといわれると，答えに困ることがある．

実際「確率」という概念が数学的に，明確に整理され，確率論が数学として確立したのは，ロシアの数学者コルモゴロフ (A. Kolmogorov) によるところが多い．彼の著書 Grundbegriffe der Wahrseheinlichkeitsrechung (1933) で，確率の公理 (axioms of probability) が提案され,「確率」が明確化したことによる．コルモゴロフは，従来から用いられていた確率という言葉を，以下で述べる 3 つの公理によりまとめあげた．

確率の公理を述べる前に,「確率」と対となる言葉に,「事象」(event) がある．不確定（不確実な）現象，たとえば「大阪の明日の午後に雨が降る」とか，1 回投げた硬貨の「表が出る」が事象に相当する．本書での目的の 1 つは，ある「事象」が起こる「確率」を求めることにある．確率を求めることは，ある事象の起こることについての「予測」ともいえる．すなわち，確率を求めることの大きな目的の 1 つは生じるであろう「事象」の予測，予知ともいえる．

明日の午後の天気で「雨が降る」，硬貨投げでの「表が出る」等のように，不確実な現象，不確実な試行 (trial), 不規則実験 (random experiment) 等により生じた事項を**「事象」**（event）という．

次の硬貨投げの不規則実験を通じて，事象およびコルモゴロフの確率の公理について説明することにする．

1.2　硬貨投げの不規則実験

硬貨を一度づつ，3 回投げるものとする．一度の試行では，硬貨の「表」(head) か「裏」(tail) のいずれかが起こるものとする．このとき，表を H，裏を T と記せば，HHH, HHT, HTH, THH, HTT, THT, TTH, TTT の事象のいずれかが起こる．起こる事象の全ての集合を**見本空間** (sample space) あるいは**標本空間**と呼び，Ω で表すものとする．すなわち，集合論の記号（集合論のまとめについては付録 1 を参照のこと）で表すと

$$\Omega \equiv \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\} \equiv \{\omega_1, \omega_2, \cdots, \omega_8\} \tag{1.1}$$

である．逆に，Ω の要素，$\omega_1, \cdots \omega_8$ を，各々**根元事象** (elementary event) あるいは**見本** (sample)，**見本点** (sample point)，**生成点** (generic point) などと呼ぶ．

すなわち 不規則実験（試行）で生じる全ての結果 (outcomes) の全体の集合を見本空間，あるいは標本空間と呼び，Ω で表す．Ω の要素を見本 (sample) あるいは見本点 (sample point) という．

有限個の要素をもつ見本空間を**有限見本空間** (finite sample space)，また可算無限個の要素もつ見本空間を**可算見本空間** (countably infinite sample space)，これらを併せて，**離散見本空間** (discrete sample space) とよぶ．また非可算無限個の要素をもつ見本空間を**非可算無限見本空間** (noncountably infinite sample space) と呼ぶ．

上記の例では，例えば，2 度「表」が出る確率はいくらか？等の確率計算が必要となる場合がある．このような場合，{2 度「表」が出る } という事項も 1 つの事象である．すなわち，この事象は根元

事象の集合 $\{HHT, HTH, THH\} = \{\omega_2, \omega_3, \omega_4\}$ に相当し，見本空間 Ω の部分集合となっている．そこで，事象全体の集合 $\mathbf{F}$ を定義しておくと，確率の計算に好都合である．すなわち，見本空間の部分集合（：事象）全体の集合として次の性質をもつ集合 $\mathbf{F}$ を定義する．

(1)　$\Omega \in \mathbf{F}$

(2)　$A \in \mathbf{F} \Longrightarrow A^C \in \mathbf{F}$

(3)　$A_i \in \mathbf{F} \Longrightarrow \cup_{i=1}^{\infty} A_i \in \mathbf{F}$

以上のような構造をもつ集合 $\mathbf{F}$ は σ 集合体 (σ-field)，あるいは σ 代数（σ-algebra）と呼ばれる．

見本空間の部分集合には，特殊な集合として，見本空間それ自身 Ω および空集合 (空事象, 不可能事象)(empty set, empty event) ϕ （演習問題とその解を参照のこと）を含めることとする．$\mathbf{F}$ を事象空間 (events space) と呼ぶこととする．また上記の A^C は集合（事象）A の補集合（補事象）(complementary set, complement) であり，集合論 (set theory) から

$$A^C \cup A = \Omega, A^C \cap A = \phi$$

の性質がある．

1.3　確率の公理

以上の例題を通じて，ある不規則実験あるいは試行を行うと，見本空間（Ω），事象空間（$\mathbf{F}$）がそのつど構成されることが分った．そこで 2 つの空間 $\Omega, \mathbf{F}$ に加えて，以下の確率の公理を満足する，集合に対する実関数 P の 3 つをまとめて，$(\Omega, \mathbf{F}, P)$ と記し，これを**確率空間** (probability space, prabability triple) という．

確率 P の公理

$\mathbf{F}$ の要素に関して以下の公理 (1) – (3) を満たす集合に対する実関数を確率 (probability, probability measure) と呼ぶ．

公理 (1)　$P(\Omega) = 1$

公理 (2)　$A \in \mathbf{F} \Longrightarrow 0 \leq P(A) \leq 1$

公理 (3)　$A_i \in \mathbf{F}\ (i = 1, 2, \ldots)$,　$A_i \cap A_j = \phi (i \neq j)$

$$\Longrightarrow P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

以上の確率の公理は，コルモゴロフの確率の公理の提案以前に，日常用いられていた「確率」という概念の根本を数学的に鋭くえぐりだしたものである．このたった 3 つの公理により，従来，自明として用いられていたさまざまな確率の性質（定理）を数学的に導くことができる．また，公理 (3) 内での $A_i \cap A_j = \phi$ である事象 A_i, A_j を**互いに排反な事象** (mutually exclusive events) という．

演習問題 1

事象空間，確率の公理を用いて，以下の問題 (1-1)-(1-4) を解け．

(1-1) $\phi \in \mathbf{F}$ であることを示せ．

(1-2) $A_i \in \mathbf{F}$　$(i = 1, 2, \ldots)$　であるとき,$\cap_{i=1}^{\infty} A_i \in \mathbf{F}$ となることを示せ．

(1-3) $A, B \in \mathbf{F}$, のとき $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ であることを示せ．

(1-4) $A \subset B$　$(A, B \in \mathbf{F})$ のとき, $P(A) \leq P(B)$ となることを示せ．

(1-3) の解答

$$A \cup B = (A \cap B) \cup (A^C \cap B) \cup (A \cap B^C)$$

と3つの排反な事象に分割できる．したがって

$$P(A \cup B) = P(A \cap B^C) + P(A^C \cap B) + P(A \cap B^C)$$

が成立する．一方，

$$\begin{aligned} P(A) &= P((A \cap B) \cup (A \cap B^C)) = P(A \cap B) + P(A \cap B^C) \\ P(B) &= P((A \cap B) \cup (A^C \cap B)) = P(A \cap B) + P(A^C \cap B) \end{aligned}$$

すなわち，

$$P(A \cap B^C) = P(A) - P(A \cap B) \qquad P(A^C \cap B) = P(B) - P(A \cap B)$$

が成立する．したがって，これらを代入すると

$$P(A \cup B) = P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B)$$

となり，(1-3) を示せる．

条件付確率

$A, B \in \mathbf{F}$，すなわち A, B を不規則実験での事象とし，$P(A) > 0$ と仮定する．この時，事象 A が生じたという条件のもとでの，事象 B が生じる確率を $P_A(B)$ または $P(B|A)$ と記す．この場合，事象 A が生じた事が既知であるので元の見本空間 Ω に代わって見本空間は A となり，$P(B|A)$ は

$$P_A(B) \equiv P(B|A) \equiv \frac{P(A \cap B)}{P(A)} = \frac{P(A, B)}{P(A)} \tag{1.2}$$

と定義する，あるいは

$$P(A \cap B) = P(A|B)P(B) \tag{1.3}$$

と定義する．また $P(B|A)$ は事象 A が与えられたもとでの B の条件付確率 (conditional probability of B given A)，という．演習問題 (2-1) のように，$P_A(\cdot)$ （または $P(\cdot|A)$）は確率の公理を満たしていることが示せる．

演習問題 2

(2-1) $\mathbf{F}$ に対して，$P_A(\cdot)$ （または $P(\cdot|A)$) は確率の公理 (1)-(3) を満たすことが示せ．

解答

$$P_A(\Omega) = \frac{P(\Omega \cap A)}{P(A)} = 1,$$

また，$C \in \mathbf{F}$ に対して，

$$P_A(C) = \frac{P(C \cap A)}{P(A)} \geq 0$$

最後に，$C_1, C_2, \cdots \in \mathbf{F}$ に対して，$A \cap C_1, A \cap C_1, \cdots$ は互いに排反事象なので

$$\begin{aligned} P_A(\cup_{i=1}^{\infty} C_i) &= \frac{P(\cup_{i=1}^{\infty}(A \cap C_i)}{P(A)} \\ &= \sum_{i=1}^{\infty} \frac{P(A \cap C_i)}{P(A)} \qquad (1.4) \\ &= \sum_{i=1}^{\infty} P_A(C_i) \qquad (1.5) \end{aligned}$$

例題　**1.1**

サイコロを 1 回投げたとき，(a) 4 以下の目が出る確率を求めよ．　(b) 投げた結果が奇数の目であることを知ったもとで，4 以下の目が出る確率を求めよ．ただし，1 の目－6 の目は等確率 $\frac{1}{6}$ で生じるものとする．

解答

サイコロを 1 回投げたとき，見本空間は $\Omega = \{1$ の目$, 2$ の目$, \cdots, 6$ の目 $\} = \{\omega_1, \cdots, \omega_6\}$ となり，いま，「奇数の目がでる」という事象を A,「4 以下の目がでる」という事象を B とすると，$\omega_i \cap \omega_j = \phi\ (i \neq j)$ であるので

$$P(B) = P(\{\omega_1, \cdots, \omega_4\}) = P(\omega_1) + P(\omega_2) + \cdots + P(\omega_4) = \frac{4}{6} = \frac{2}{3}$$

は明らか，また

$$P(A \cap B) = P(\{\omega_1, \omega_3\}) = P(\omega_1) + P(\omega_3) = \frac{2}{6} = \frac{1}{3}$$

であるので，

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

以上の結果を用いて次の定理を示すことができる．

例題　**1.2**

3 つの事象 A_1, A_2, A_3 に対して，

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1 \cap A_2) P(A_3 | A_1 \cap A_2) \\ &= P(A_1) P(A_2 | A_1) P(A_3 | A_1 \cap A_2) \end{aligned}$$

が成立する．

解　$A_1, A_2 \in \mathbf{F}$　であり，$A_1 \cap A_2 \in \mathbf{F}$　であるので，はじめの等式は明らか，第 2 番目の等式も条件付確率の定義より明らか．

例題　**1.3**

事象 $A_1, A_2, \ldots, A_n$ が互いに排反な事象 (mutually exclusive events)，かつそれらの合併（和）集合が見本空間 Ω となるとき，すなわち

$$A_i \cap A_j = \phi \quad (i \neq j), \quad \cup_{i=1}^{n} A_i = \Omega$$

であるとき,（簡潔に,「事象 $A_1, A_2, \ldots, A_n$ は見本空間 Ω の分割」ともいう ）

$$\begin{aligned} P(A) &= P(A \cap \{A_1 \cup A_2 \cup \cdots \cup A_n\} \\ &= P(\{A \cap A_1\} \cup \{A \cap A_2\} \cup \cdots \cup \{A \cap A_n\}) \\ &= \sum_{k=1}^{n} P(A \cap A_k) \end{aligned}$$

が成立する．

独立事象 (independent events)

$P(B|A) = P(B)$ であるとき, 2 つの事象 A と B は，独立な事象 (independent events) という．またこのことは

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

が成立することと等価である．

例題　1.4

事象 A と B が独立な事象であれば，事象 A と B^c もまた独立な事象であることを示せ．

証明:

$$P(A) = P(A \cap \Omega) = P(A \cap \{B \cup B^c\}) = P(B)P(A|B) + P(B^c)P(A|B^c)$$

が成立する．ただし，最後の2つの等式では，$A{\cap}B$ と $A{\cap}B^C$ は互いに排反な事象;$(A{\cap}B){\cap}(A{\cap}B^c) = \phi$，および条件付確率の定義を用いている．ここでまた，$P(A|B) = P(A)$，$P(B^c) = 1 - P(B)$ であるので，

$$P(A) = P(B)P(A) + [1 - P(B)]P(A|B^c)$$

となる．すなわち

$$P(A)[1 - P(B)] = [1 - P(B)]P(A|B^c)$$

が成立し，

$$P(A) = P(A|B^c)$$

となる．

定理 1.1 ベイズの定理 *(Bayes' Theorem (Rule))* いま，事象 $A_1, A_2, \ldots, A_n$ が互いに排反，かつそれらの合併（和）集合が見本空間 Ω となるとき，すなわち，事象 $A_1, A_2, \ldots, A_n$ は見本空間 Ω の分割であるとき，事象 A のもとでの条件付確率について

$$P(A_k|A) = \frac{P(A_k \cap A)}{P(A)} = \frac{P(A_k)P(A|A_k)}{P(A \cap \Omega)} = \frac{P(A_k)P(A|A_k)}{\sum_{j=1}^{n} P(A_j)P(A|A_j)}$$

が成立する．

上記の定理は，与えられた条件となる事象 A のもとでの，事象 A_k の確率が，条件と結果が反転した $P(A|A_j)$ の計算により求められることを示しており，原因確率の定理 (a theorem of probability of cause) とも呼ばれる．

1.3.1　確率変数

確率変数 $X(\omega)$ とは，$\omega \in \Omega$ から実数 (real number) (複素数 (complex number) の場合にも拡張できるが）への一価関数，かつ，

$$\{\omega : X(\omega) \leq x\} \in \mathbf{F}$$

なる実関数と定義される．測度論では，以上の性質を満たせば，確率変数 X は可測関数 (measurable function) であるいわれる．

前章での「硬貨を3回投げた場合の不規則実験の例」について説明しよう．この不規則実験では，見本空間は $\Omega \equiv \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\} \equiv \{\omega_1, \omega_2, \cdots, \omega_8\}$ であり，変数 X を，たとえば

$$X(\omega_1) = 3, \quad X(\omega_2) = 2, \quad X(\omega_3) = 2, \quad X(\omega_4) = 2,$$

$$X(\omega_5) = 1, \quad X(\omega_2) = 1, \quad X(\omega_3) = 1, \quad X(\omega_4) = 0,$$

と定義すると，X は確率変数 (random variable) となる．以上の確率変数は，実は表の生じた回数として，整数 $0, 1, 2, 3$ を対応させたものである．同じく，変数 Y を

$$Y(\omega_1) = 0, \quad Y(\omega_2) = 1, \quad Y(\omega_3) = 1, \quad Y(\omega_4) = 1,$$

$$Y(\omega_5)=2,\quad Y(\omega_2)=2,\quad Y(\omega_3)=2,\quad Y(\omega_4)=3,$$

として，裏の起こった回数を整数 $0,1,2,3$ を対応させた変数 Y も確率変数となる．

このとき，変数の取るべき値によって確率変数を，以下のように**離散型確率変数** (discrete random variable) および**連続型確率変数** (continuous random variable) に分類できる．すなわち，確率変数が有限個の値，あるいは加算無限個の値をとれば離散型確率変数，非可算無限個（例えば 0 から 10 の間の実数）の値をとれば連続型確率変数という．したがって，上記の例での確率変数 X,Y はいずれも離散型確率変数ということになる．カルマンフィルタの理論では、連続型確率変数，特に正規型（ガウス）確率変数が重要であるので，以後、連続型確率変数についてのみ説明する．

1.3.2 確率分布 (probability distribution)

連続型確率変数 (continuous random variable)

確率変数 $X(\omega)$ が絶対連続 (absolutely continuous)，あるいは簡単に連続 (continuous) であるとは，分布関数 $F(x)$ が，以下の性質 1) − 2) を満たす実関数 $f(x)$ を用いて，

$$F(x)=P(X\le x)=\int_{-\infty}^{x}f(u)du \qquad (-\infty<x<\infty) \tag{1.6}$$

と表現できる場合である．ただし，$f(x)$ は

1) $f(x)\ge 0$

2) $\int_{-\infty}^{\infty}f(x)dx=1.$

(1.6) 式より，X が連続型確率変数であると，X が特定の値をとる確率は 0 となる．なぜなら，いま特定の値を，例えば a とすると

$$P(X=a)=\lim_{h\to 0_+}P(a<X\le a+h)$$

であるので，

$$\lim_{h\to 0_+}P(a<X\le a+h)=\lim_{h\to 0_+}P(X\le a+h)-P(X\le a)$$

となり

$$P(X=a)=\lim_{h\to 0_+}F(a+h)-F(a)=0$$

となるからである．したがって，連続型確率変数 X がある 2 つの値 a,b の間の値をとる確率（各不等号は等号を含めてもよい）は

$$P(a<X<b)=\int_a^b f(x)dx$$

となる．

また，連続型確率変数に対する実関数 $f(x)$ は**確率関数** (probability function)，**確率分布** (probability distribution)，とも呼ばれるが，通常は**確率密度関数** (probability density function; 略して p.d.f) あるいは単に**密度関数** (density function) と呼ぶことが多い．上記の性質 1), 2) を満たす関数は密度関数であり，必要な確率は () 式により求めることができる．

分布関数 (distribution functions)

確率変数 $X(\omega)$ と任意の実数 x に対して，

$$F(x)\equiv P[X(\omega)\le x]=P[\{\omega\};X(\omega)\le x]$$

なる関数，$F(x)$ を**累積分布関数** (cumulative distribution function)，あるいは簡単に**分布関数** (distribution function) という．このとき $F(x)$ は以下の性質をもっている．

1). $x \le y$ ならば $F(x) \le F(y)$

2). $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to\infty} F(x) = 1$

3). $F(x)$ は右連続関数（すなわち，すべての x に対して $\lim_{h\to 0^+} F(x+h) = F(x)$ が成立する）．

例

いま，連続型確率変数 X が区間 x と $x+h$ の間の値をとる確率，すなわち

$$P(x < X \le x+h) = P(x \le X \le x+h) = \int_x^{x+h} f(u)du \tag{1.7}$$

とすると，十分小なる h に対して

$$F(x+h) - F(x) = f(x)h + o(h) \tag{1.8}$$

となる．ただし $o(h)$ はランダウ記号

$$\lim_{h\to 0}\frac{o(h)}{h} = 0$$

である．

$$\lim_{h\to 0}\frac{F(x+h)-F(x)}{h} = \frac{dF(x)}{dx} = f(x)$$

となる．

1.3.3 結合分布 (joint distribution)

2 つの確率変数 X, Y を考える．

X, Y が連続型確率変数の場合，X, Y の**結合**（同時）**確率密度関数** (joint probability density function) あるいは**結合確率関数** (joint probability function)$p_{X,Y}(x,y)$ は性質 2 での和を積分に替え，以下の性質を満たすものとして定義する．

$$\begin{aligned} &1) \qquad p_{X,Y}(x,y) \ge 0 \\ &2) \qquad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{X,Y}(x,y)dxdy = 1 \end{aligned} \tag{1.9}$$

いま，結合確率密度関数 $p_{X,Y}(x,y)$ を用いると，X が a, b の間の値をとり，同時に Y が c, d の間の値をとる確率は

$$P(a < X < b, c < Y < d) = \int_a^b\int_c^d p_{X,Y}(x,y)dxdy$$

で表せる．また，離散型確率変数の場合と類似して，連続型確率変数 X, Y の結合分布関数 (joint distribution function) は

$$F(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dud$$

により定義する．このとき，

$$\frac{\partial^2 F}{\partial x\partial y} = p_{X,Y}(x,y)$$

が成立する（演習問題とその解を参照）．また各々の確率変数 X, Y についての周辺分布関数 (marginal distribution function)，$F_X(x), F_Y(y)$ は

$$P(X \le x) = F_X(x) = \int_{u=-\infty}^{x}\int_{-\infty}^{\infty} F(u,v)dudv = \int_{-\infty}^{x} du\int_{-\infty}^{\infty} F(u,v)dv$$

$$P(Y \leq y) = F_Y(y) = \int_{-\infty}^{x} \int_{v=-\infty}^{y} F(u,v) du dv = \int_{-\infty}^{x} du \int_{-\infty}^{y} F(u,v) dv$$

で定義される．**周辺確率密度関数** (marginal probability density function)，$p_X(x), p_Y(y)$ は，離散型確率変数と類似して

$$P(X \leq x) = p_X(x) = \int_{-\infty}^{\infty} f(x,y) dy$$

$$P(Y \leq y) = p_Y(y) = \int_{-\infty}^{\infty} f(x,y) dx$$

で定義される．

独立確率変数

いま，X, Y を離散型確率変数とする．このとき，$X = x$ をとる事象と $Y = y$ をとる事象とが独立事象 (independent events) であれば，

$$P(X = x, Y = y) = P(X = x) P(Y = y)$$

が成立し，確率変数 X, Y は独立な確率変数 (independent random variables) という．等価的に

$$p_{X,Y}(x,y) = p_X(x) p_Y(y) \tag{1.10}$$

も成り立つ．

また同様に，連続型確率変数 X, Y についても，すべての x, y について，事象 $X \leq x$ と事象 $Y \leq y$ とが独立事象 (independent events) であれば，

$$P(X \leq x, Y \leq y) \equiv P(\{X \leq x\} \cap \{Y \leq y\}) = P(X \leq x) P(Y \leq y)$$

となる．上式は分布関数で表現すると

$$F(x,y) = F_X(x) F_Y(y) \tag{1.11}$$

と等価である．このとき，確率変数 X, Y は独立な確率変数 (independent random variables) という．また，逆にすべての x, y に対して，(1.10) 式あるいは (1.11) 式が成立すれば，X, Y は独立な確率変数である．また，結合確率密度関数 $p_{X,Y}(x,y)$ についても，

$$p_{X,Y}(x,y) = p_X(x) p_Y(y)$$

が成立すれば，連続型確率変数 X, Y は独立である．

1.4 確率変数の変換

X の pdf を $p_X(x)$ とする．一対一関数 (these corresponds one and only one) $g(x)$ により $Y = g(X)$ となる確率変数 Y を考える．$g(x)$ の逆関数を $h(y)$ とすると，

$$X = g^{-1}(Y) = h(Y) \tag{1.12}$$

である．ここで，次の等式が成立し，

$$P(y < Y \leq y + dy) = P(x < X \leq x + dx) \tag{1.13}$$

すなわち

$$p_Y(y)|dy| = p_X(x)|dx| \tag{1.14}$$

が成り立つ．したがって Y の確率密度関数 $p_Y(y)$ は

$$p_Y(y)\big|dy\big| = p_X(x)\big|dx\big| \tag{1.15}$$

すなわち

$$p_Y(y) = p_X(x)\Big|\frac{dx}{dy}\Big| = p_X[h(y)]\Big|h'(y)\Big| \tag{1.16}$$

となる．次に以上の結果を，2 変数系に拡張する．

X_1, X_2 を連続型確率変数とし, jpdf を $p_{X_1,X_2}(x_1,x_2)$ とする．このとき

$$Y_1 = g_1(X_1,X_2) \tag{1.17}$$

$$Y_2 = g_2(X_1,X_2) \tag{1.18}$$

とし, その逆関数は

$$X_1 = h_1(Y_1,Y_2) \tag{1.19}$$

$$X_2 = h_2(Y_1,Y_2) \tag{1.20}$$

で与えられるものとする．このとき Y_1,Y_2 の 結合密度関数 $p_{Y_1,Y_2}(y_1,y_2)$ は

$$p_{Y_1,Y_2}(y_1,y_2)\big|dy_1dy_2\big| = p_{X_1,X_2}(x_1,x_2)\big|dx_1dx_2\big| \tag{1.21}$$

すなわち

$$p_{Y_1,Y_2}(y_1,y_2) = p_{X_1,X_2}(x_1,x_2)\Big|\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)}\Big| = p_{X_1,X_2}[h_1(y_1,y_2),h_2(y_1,y_2)]\big|J\big| \tag{1.22}$$

となる．ただし J は ヤコブの行列式 (Jacobian determinant(あるいは，簡単に ヤコビヤン (Jacobian)) であり,

$$J = \frac{\partial(x_1,x_2)}{\partial(y_1,y_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} \tag{1.23}$$

で与えられる．

例

X は $\sim N(0,1)$，すなわち標準正規分布に従う確率変数とする．$Y = g(X) = X^2$ であるとき，確率変数 Y の pdf $p_Y(y)$ を求めよ．

解

$x = g^{-1}(y) = \pm\sqrt{y}$ であり，$|\frac{dx}{dy}| = |\pm\frac{1}{2\sqrt{y}}| = |\frac{1}{2\sqrt{y}}|$ となる．したがって，$y \geq 0$ に対して，

$$\begin{aligned} p_Y(y) &= p_X(\sqrt{y})\frac{1}{2\sqrt{y}} + p_X(-\sqrt{y})\frac{1}{2\sqrt{y}} \\ &= \sqrt{2\pi}e^{-\frac{y}{2}}\frac{1}{2\sqrt{y}} + \sqrt{2\pi}e^{-\frac{y}{2}}\frac{1}{2\sqrt{y}} \\ &= \sqrt{2\pi y}e^{-\frac{y}{2}} \end{aligned} \tag{1.24}$$

となる．この pdf は後に示す，自由度 1 の χ^2 分布 (カイジジョウブンプと読む) である．

例 確率変数の和の分布

確率変数 X_1, X_2 の結合 pdf を $p_{X_1,X_2}(x_1,x_2)$ とする．このとき，X_1 と X_2 の和の確率変数 Y

$$Y = X_1 + X_2$$

を考える．Y の pdf $p_Y(y)$ を求めてみる．

$$
\begin{aligned}
Y &= X_1 + X_2 \\
Z &= X_1
\end{aligned}
$$

なる補助的な確率変数 Z を導入する．このとき

$$
\begin{aligned}
X_1 &= Z \\
X_2 &= Y - Z
\end{aligned}
$$

となり，

$$
p_{Y,Z}(y,z) = p_{X_1,X_2}(z, y-z)|\left|\frac{\partial(x_1,x_2)}{\partial(z,y)}\right|| \tag{1.25}
$$

である．

$$
|\left|\frac{\partial(x_1,x_2)}{\partial(z,y)}\right|| = |\begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix}| = 1 \tag{1.26}
$$

であるので，Y の周辺 pdf $p_Y(y)$ を求めると

$$
p_Y(y) = \int_{-\infty}^{\infty} p_{Y,Z}(y,z)dz = \int_{-\infty}^{\infty} p_{X_1,X_2}(z, y-z)dz \tag{1.27}
$$

となる．特に，特に確率変数 X_1, X_2 が独立であれば，

$$
p_{X_1,X_2}(z, y-z) = p_{X_1}(z)p_{X_2}(y-z)
$$

が成立すので，(1.27) 式は

$$
p_Y(y) = \int_{-\infty}^{\infty} p_{X_1}(z)p_{X_2}(y-z)dz \tag{1.28}
$$

となり，$p_Y(y)$ は 2 つの pdf $p_{X_1}(y), p_{X_2}(y)$ の合成積 (convolution, 畳込み積分) となっている．

1.5 期待値

確率変数の期待値（expectation of a random variables $X(\omega)$）

確率変数 X の**期待値** (expectation) は

$$
\mathbf{E}(X) \equiv \int_{-\infty}^{\infty} x f(x)dx \equiv \mu_X \tag{1.29}
$$

により定義する．また確率変数 $X(\omega)$ の**分散** (variance) は

$$
\mathrm{Var}(X) \equiv \mathbf{E}[(X-\mu_X)(X-\mu_X)^T] \tag{1.30}
$$

により与えられる．すなわち、

$$
\mathrm{Var}(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(x-\mu_X)^T p_X(x)dx \equiv \sigma_X^2 \tag{1.31}
$$

で与えられる．

結合分布に対する分散，共分散

確率変数が 2 つ以上の場合についての期待値について述べる．今簡単のため X，Y を 2 つの連続型確率変数としてその結合 pdf を $p_{X,Y}(x,y)$ とする．このとき，X および Y の平均値（あるいは期待値）は以下で定義する．

$$\mu_X = E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x p_{X,Y}(x,y)dxdy \tag{1.32}$$

$$\mu_Y = E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y p_{X,Y}(x,y)dxdy \tag{1.33}$$

また，分散は

$$\sigma_X^2 = E[(X-\mu_X)^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu x)^2 p_{X,Y}(x,y)dxdy \tag{1.34}$$

$$\sigma_Y^2 = E[(Y-\mu_Y)^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y-\mu y)^2 p_{X,Y}(x,y)dxdy \tag{1.35}$$

これらは周辺 pdf $p_X(x), p_Y(y)$ を用いると，

$$\mu_X = \int_{-\infty}^{\infty} x p_X(x)dx, \qquad \mu_y = \int_{-\infty}^{\infty} y p_Y(y)dy \tag{1.36}$$

$$\sigma_X^2 = \int_{-\infty}^{\infty}(x-\mu_X)^2 p_X(x)dx, \qquad \sigma_Y^2 = \int_{-\infty}^{\infty}(y-\mu_Y)^2 p_Y(y)dy \tag{1.37}$$

と一致することは，

$$p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)dy, \qquad p_Y(y) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)dx \tag{1.38}$$

の関係から明かである．
X，Y の共分散 (covariance)σ_{XY} は

$$\sigma_{XY} = \mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \tag{1.39}$$

で定義される．すなわち，

$$\sigma_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)p_{X,Y}(x,y)dxdy \equiv \rho\sigma_X\sigma_Y \tag{1.40}$$

であり，ρ は**相関係数** (correlation coefficient) と呼ばれる．

平均値 μ_X，μ_Y，分散 σ_X^2，σ_Y^2，共分散 σ_{XY} について以下の関係式が示せる．
(1) $\sigma_{XY} = E(XY) - E(X)E(Y) = E(XY) - \mu_X\mu_Y$
(2) X と Y が独立な (independent) 確率変数であるとき，$\sigma_{XY} = \mathrm{Cov}(X,Y) = 0$
(3) $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm \mathrm{Cov}(X,Y)$ あるいは，$\sigma_{X\pm Y}^2 = \sigma_X^2 + \sigma_Y^2 \pm 2\sigma_{XY}$
(4) $|\sigma_{XY}| \leq \sigma_X\sigma_Y$

(4) はシュワルツ（Schwarz）の不等式を用いて，以下のように示すことが出来る．

$$\begin{aligned}
|\sigma_{XY}| &= \left|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)p_{X,Y}(x,y)dxdy\right| \\
&= \left|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)\sqrt{p_{X,Y}(x,y)}(y-\mu_Y)\sqrt{p_{X,Y}(x,y)}dxdy\right| \\
&\leq \Big[\int_{-\infty}^{\infty}(x-\mu_X)^2 p_{X,Y}(x,y)dxdy\Big]^{1/2} \times \Big[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y-\mu_Y)^2 p_{X,Y}(x,y)\Big]^{1/2} \\
&= \sigma_X\sigma_Y
\end{aligned} \tag{1.41}$$

相関係数 (correlation coefficient)

確率変数 X，Y が独立であれば，

$$\mathrm{Cov}(X,Y)=\sigma_{XY}=0$$

となる．一方 X，Y が完全に従属している．例えば $X=Y$ とすると，

$$\mathrm{Cov}(X,Y)=\sigma_{XY}=\sigma_X\sigma_Y$$

となることを示せる．したがって確率変数 X，Y の（線形的）従属性の尺度として，

$$\rho=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

を定義すると，ρ を相関係数 (correlation coefficient or coefficient of correlation) という．$|\sigma_{XY}|\le\sigma_X\sigma_Y$ より $|\rho|\le 1$ であることが示せる．特に，$\rho=0$（すなわち共分散が０）の場合，X と Y は無相関 (uncorrelated) という．

条件付確率と条件付期待値

事象 A が与えられたもとでの事象 B の条件付確率 (conditional probability) $P(B|A)$ は

$$P(B|A)=\frac{P(A\bigcap B)}{P(A)}=\frac{P(A,B)}{P(A)} \tag{1.42}$$

すなわち，

$$P(A,B)=P(B|A)P(A) \tag{1.43}$$

が成り立つ．この時、連続型確率変数 X,Y に対する**結合確率密度関数** (joint probability density function (: jpdf))$p_{X,Y}(x,y)$ に対して

$$p_{X,Y}(x,y)\equiv\lim_{\Delta x,\Delta y\to 0}\frac{P[x<X(\omega)\le x+\Delta x, y<Y(\omega)\le y+\Delta y]}{\Delta x\Delta y} \tag{1.44}$$

$$p_{X,Y}(x,y)=\lim_{\Delta x,\Delta y\to 0}\frac{P[x<X\le x+\Delta x\big|y<Y\le y+\Delta y]P[y<Y\le y+\Delta y]}{\Delta x\Delta y} \tag{1.45}$$

であり，**条件付確率密度**

$$\begin{aligned} p(x|y) &\equiv \lim_{\Delta x,\Delta y\to 0}\frac{P[x<X\le x+\Delta x\big|y<Y\le y+\Delta y]}{\Delta x} \\ &= \lim_{\Delta x\to 0}\frac{P[x<X\le x+\Delta x\big|Y=y]}{\Delta x} \end{aligned} \tag{1.46}$$

で定義する．ここで Y の周辺確率密度関数

$$p_Y(y)=\lim_{\Delta y\to 0}\frac{P(y<Y\le y+\Delta y)}{\Delta y} \tag{1.47}$$

であることを利用すると

$$p_{X,Y}(x,y)=p_{X|Y}(x|y)p_Y(y) \tag{1.48}$$

が成立する．すなわち，X，Y の結合 *pdf* を $p_{X,Y}(x,y)$ とすると，$Y=y$ が与えられたもとでの条件付 *pdf* $p_{X|Y}(x|y)$ は

$$p_{X|Y}(x|y)=\frac{p_{X,Y}(x,y)}{p_X(x)} \tag{1.49}$$

で与えられた．このとき，$Y=y$ が与えられたもとでの X の条件付期待値 (conditional expectation)(あるいは，条件付平均値 (conditional mean)) は

$$E[X|Y=y]=\int_{-\infty}^{\infty}xp_{X|Y}(x|y)dy \tag{1.50}$$

で与えられる．$Y=y$ は連続型確率変数の場合，$y<X\leq y+dy$ と解釈する．このとき，次の性質が成り立つ．

(P1)　X と Y が独立ならば，

$$E[X|Y=y]=E[X] \tag{1.51}$$

(証明)

$$\begin{aligned}
E[X|Y=y] &= \int_{-\infty}^{\infty}xp(x|y)dy && (1.52)\\
&= \int_{-\infty}^{\infty}x\frac{p_{X,Y}(x,y)}{p_Y(y)}dy && (1.53)\\
&= \int_{-\infty}^{\infty}x\frac{p_X(x)p_Y(y)}{p_X(x)}dy && (1.54)\\
&= E[X] && (1.55)
\end{aligned}$$

(P2)

$$E(X)=\int_{-\infty}^{\infty}E(X|Y=y)p_Y(y)dx \tag{1.56}$$

(証明)

$$\begin{aligned}
\int_{-\infty}^{\infty}E(X|Y=y)p_Y(y)dy &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xp(x|y)dxp_Y(y)dy && (1.57)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xp_{X,Y}(x,y)dxdy && (1.58)\\
&= E[X] && (1.59)
\end{aligned}$$

また，Y が与えられたもとでの X の a まわりでの r 次の条件付モーメント (rth conditional moment) は

$$E[(X-a)^r|Y=x]=\int_{-\infty}^{\infty}(x-a)^rp(x|y)dx \tag{1.60}$$

で与えられる．

確率ベクトルの結合確率密度関数の表現

X を n 次元確率ベクトルとする．このとき，X の確率密度関数をスカラー確率変数と同じく，$p_X(x)$ で表す．また n 次元確率ベクトル X の期待値を $E[X]$ と記し，

$$E[X]=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}p_{X_1,\cdots,X_n}(x_1,\cdots,x_n)dx_1\cdots dx_n\equiv\int_{-\infty}^{\infty}xp_X(x)dx \tag{1.61}$$

と記す．

例　正規性確率ベクトル

n 次元確率ベクトル X が平均 μ，共分散行列 R の正規性 (ガウス性) 確率ベクトルとは，その (結合)pdf が

$$p_X(x)=\frac{1}{(2\pi)^{n/2}\left|R\right|^{1/2}}\exp\{-\frac{1}{2}[x-\mu]^{\mathrm{T}}R^{-1}[x-\mu]\} \tag{1.62}$$

である．X がスカラー確率変数の場合と同じく

$$X \sim N[\mu,\ R]$$

と記す．次に確率ベクトル X, Y に対しての条件付期待値の重要な性質について述べる．

条件付期待値

m 次元確率ベクトル $Y = y$ であるとき，n 次元確率ベクトル X の条件付期待値 (conditional expectation) は

$$E(X|Y=y) \equiv \int_{-\infty}^{\infty} x p_{X|y}(x|y) dx \tag{1.63}$$

で与えられる．ここで，次の評価関数を最小にする $g(y)$ を求めることを考えよう．すなわち

$$E[X-g(Y)]^T[X-g(Y)] \equiv \int\int [x-g(y)]^T[x-g(y)] p_{X,Y}(x,y) dxdy \tag{1.64}$$

このとき

$$\begin{aligned}
& E[X-g(Y)]^T[X-g(Y)] \\
= & \int\int [x^T x - x^T g(y) - g^T(y)x + g^T(y)g(y)] p_{X|Y}(x|y) p_{X,Y}(y) dxdy \\
= & \int\int x^T x p_{X,Y}(x,y) dxdy \\
& + \int p(y)dy \int [-x^T g(y) - g^T(y)x + g^T(y)g(y)] p(x|y) dx \\
= & \int\int x^T x p_{X,Y}(x,y) dxdy + \int p(y)dy \{-E(X^T|Y=y)g(y) - g^T(y)E(X|Y=y) + g^T(y)g(y)\} \\
= & -\int p(y)dy \Big\{ \big(g(y) - E(X|Y=y)\big)^T \big(g(y) - E(X|Y=y)\big) \Big\} \\
& + \int\int x^T x p_{X,Y}(x,y) dxdy - \int p(y)dy \Big\{ \big(E(X|Y=y)\big)^T \big(E(X|Y=y)\big) \Big\}
\end{aligned} \tag{1.65}$$

となる．したがって，

$$g(y) = E(X|Y=y) \tag{1.66}$$

とすれば，(1.64) を最小化でき，

$$\min E[X-g(Y)]^T[X-g(Y)] = \int x^T x p_X(x) dx - \int E(X^T|Y=y)E(X|Y=y)p(y)dy \tag{1.67}$$

となる．

1.6 推定論

1.6.1 点推定

例えば, 大きさ n の標本 $O_n(X_1, \ldots, X_n)$

$$X_t = \mu + V_t \tag{1.68}$$

が得られたとする．ただし，$V_t \sim N[0, \sigma^2]$，$V_1, \cdots, V_n$ は互いに互いに独立な確率変数とする．このとき標本 $O_n(X_1, \ldots, X_n)$ から，未知母数 (未知パラメータ), μ および σ^2 を推定するため, 標本

の平均 (sample mean) $\bar{X}$, **標本の分散** (sample variance) S^2 を考え, これを母数 μ, σ^2 の推定量 (estimator) とする. すなわち, 標本平均, 標本分散をそれぞれ

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1.69}$$

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{1.70}$$

とする. この時, 統計量 $\bar{X}$ の平均値は

$$E[\bar{X}] = E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{1}{n}n\mu = \mu \tag{1.71}$$

となり, μ と一致する.

このように, 母数 θ に対する推定量 $T(X_1, \ldots, X_n)$ の平均が, 次式のように母数 θ に等しいとき

$$E[T(X_1, \ldots, X_n)] = \theta \tag{1.72}$$

$T(X_1, \ldots, X_n)$ を **不偏推定量** (unbiased estimator) とよび, その実現値 $T(x_1, \ldots, x_n)$ を **不偏推定値** (unbiased estimate) という.

次に, 標本分散 S^2 が σ^2 の不偏推定量であるかどうかを調べるため, $E[S^2]$ を計算する. そのために, 次の関係式に注目する.

$$\begin{aligned}\sum_{i=1}^{n}(X_i - \mu)^2 &= \sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \mu)^2 \\ &= \sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{n}2(X_i - \bar{X})(\bar{X} - \mu) + \sum_{i=1}^{n}(\bar{X} - \mu)^2 \end{aligned} \tag{1.73}$$

であり, 第 2 項目は

$$2\sum_{i=1}^{n}(X_i - \bar{X})(\bar{X} - \mu) = 2(\bar{X} - \mu)\sum_{i=1}^{n}(X_i - \bar{X}) = 2(\bar{X} - \mu)(n\bar{X} - n\bar{X}) = 0 \tag{1.74}$$

となるので,

$$\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 \tag{1.75}$$

となる. 上式を n で割ることにより,

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 + (\bar{X} - \mu)^2 \tag{1.76}$$

を得るが, 右辺第 1 項目は, S^2 であり, 移項して両辺の期待値をとると

$$E[S^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right] - E\left[(\bar{X} - \mu)^2\right] \tag{1.77}$$

となる. ここで

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right] = \frac{1}{n}n\sigma^2 = \sigma^2 \tag{1.78}$$

また

$$
\begin{aligned}
E\left[(\bar{X}-\mu)^2\right] &= E\left[\left(\frac{\sum_{i=1}^n X_i}{n}-\mu\right)^2\right] \\
&= E\left[\left(\frac{\sum_{i=1}^n X_i - n\mu}{n}\right)^2\right] \\
&= E\left[\left(\frac{\sum_{i=1}^n (X_i-\mu)}{n}\right)^2\right] \\
&= \frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n} \qquad (1.79)
\end{aligned}
$$

となり，(1.78)(1.79) 式を用いると，(1.77) 式は

$$
\begin{aligned}
E[S^2] &= \sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2=(1-\frac{1}{n})\sigma^2 \\
&\neq \sigma^2 \qquad (1.80)
\end{aligned}
$$

となる．したがって，標本分散 S^2 は σ^2 の不偏推定量ではない. しかし

$$
U^2 \equiv \frac{1}{n-1}\sum_{i=1}^n (X_i-\bar{X})^2=\frac{n}{n-1}S^2 \qquad (1.81)
$$

なる推定量 $U^2(X_1,\ldots,X_n)$ は σ^2 の不偏推定量となる.

一般に, 母数 θ の不偏推定量 $T(X_1,\ldots,X_n)$ の中では, 推定量の分散 $E[(T-E(T))^2]=E[(T-\theta)^2]$ がより小さいほど, よい推定であるといえる. すなわち $E[(T-\theta)^2]$ が最小となるような不偏推定量 T が最良な θ の推定量であるといえる. さらに, 一般的に, T' が不偏推定量でなくても, 真の母数 θ のまわりのばらつき (2 次モーメント) $E[(T'-\theta)^2]$ がより小さいほど, よい推定, あるいはより有効な推定量であるといえる.

また, T' が θ の不偏推定量でない場合でも, 標本の大きさを $n\to\infty$ としたとき, T' が θ へ **確率収束** (convergence in probability) するならば, すなわち, $\forall\varepsilon>0$ に対して

$$
\lim_{n\to\infty} P\left(\left|T'(X_1,\cdots,X_n)-\theta\right|\geq\varepsilon\right)=0 \qquad (1.82)
$$

であるならば, $T'(X_1,\cdots,X_n)$ は θ の一致推定量 (consistence estimator) であるという. 式 (1.82) が成立するための十分条件は, チェビシェフの不等式 (Chebycheff's inequality) を用いることにより

$$
\lim_{n\to\infty} E\Big\{T'(X_1,\cdots,X_n)\Big\}=\theta
$$

かつ

$$
\lim_{n\to\infty} E\left\{\left(T'(X_1,\cdots,X_n)-E\Big[T'(X_1,\cdots,X_n)\Big]\right)^2\right\}=0
$$

が成立することである.

したがって, 上記の母分散 σ^2 に対する統計量 $S^2(X_1,\cdots,X_n)$ は σ^2 の不偏推定量ではないが, 一致推定量であることを示せる. また, U^2 も σ^2 の一致推定量である．さて後に示す, **クラメール・ラオ** (Cramer-Rao) の不等式より, 母数 θ の任意の不偏推定量 $T(X_1,\cdots,X_n)$ の分散について

$$
\begin{aligned}
\mathrm{Var}(T) &\equiv E\left\{\Big(T(X_1,\cdots,X_n)-\theta\Big)^2\right\} \\
&\geq \frac{1}{nE\left[\left(\frac{\partial \log f(x,\theta)}{\partial\theta}\right)^2\right]}=\frac{1}{-nE\left[\frac{\partial^2 \log f(x,\theta)}{\partial\theta^2}\right]} \\
&\equiv C_B
\end{aligned}
$$

の関係があるので, θ の不偏推定量 T の中で分散が C_B となるような推定量があれば, この推定量 T は θ の不偏かつ最良の推定量である. このような推定量を **有効推定量** (efficient estimator) あるいは **最小分散推定値** (minimum variance estimator) という.

以上の例を通じて, 未知母数 θ の推定量 $T(X_1, \cdots, X_n)$ の望ましい性質は, 次の ( )〜( ) としてまとめることができる. すなわち

推定量の望ましい性質

( ) 不偏性： $E[T] = \theta$

( ) 一致性： $\forall \varepsilon > 0$ に対して, $\lim_{n\to\infty} P(|T - \theta| \geq \varepsilon) = 0$

( ) 有効性

以上，ここでは, 実際に得られた標本データ $O_n(x_1, \cdots, x_n)$ にもとづき, θ の推定量 $T(X_1, \cdots, X_n)$ の 1 つの実現値 $T(x_1, \cdots, x_n)$ を得ることになるので, このような推定を未知母数 θ の**点推定** (point estimation) という. 点推定に対比する概念が，以下に述べる区間推定 (interval estimation) である.

1.6.2 区間推定

区間推定とは, 与えられた標本 $O_n(X_1, \cdots, X_n)$ より, 区間の上限, 下限となる 2 つの統計量 $\hat{\theta}_u(X_1, \cdots, X_n)$ および $\hat{\theta}_l(X_1, \cdots, X_n)$ を見い出し, 未知母数 θ が与えられた確率 $(1-\alpha), 0 < \alpha < 1$, 区間 $[\hat{\theta}_l, \hat{\theta}_u]$ にあることを保証しようとするものである. すなわち, 与えられた $0 \leq \alpha \leq 1$ に対して

$$P\Big(\hat{\theta}_l(X_1, \cdots, X_n) \leq \theta \leq \hat{\theta}_u(X_1, \cdots, X_n)\Big) = 1 - \alpha$$

なる, 区間 $[\hat{\theta}_l, \hat{\theta}_u]$ を見出そうとするものである. このとき, 区間 $[\hat{\theta}_l, \hat{\theta}_u]$ は, **信頼度** (confident)$1-\alpha$ の**信頼区間** (confidence interval)，またはパーセンテージ表現し，**信頼水準** (confident level)$(1-\alpha)\times 100$ % の信頼区間という．$\hat{\theta}_l, \hat{\theta}_u$ をあわせて $1-\alpha$ の**信頼限界** (confidence limits) という. 区間 $(\hat{\theta}_l, \hat{\theta}_u)$ はなるべく狭い方が望ましい. $\Big[$ 区間 $[(\hat{\theta}_l, \hat{\theta}_u]$ は母数 θ を含むことは $(1-\alpha)\times 100$ %信用できる. $\Big]$

1.6.3 母分散 σ^2 が既知の正規母集団の母平均 μ の区間推定

いま, 分散 σ^2 を既知, 母平均 μ が未知である正規母集団 $N(\mu, \sigma^2)$ から無作為抽出された大きさ n の標本 $O_n(X_1, \cdots, X_n)$ を考える. このとき, 標本平均 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ であるので,

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

とすると, Z は標準正規分布 (standard normal distribution) $N(0,1)$ に従う. したがって, 与えられた $0 \leq \alpha \leq 1$ に対して

$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha \tag{1.83}$$

が成立する. ただし, $z_{\frac{\alpha}{2}}$ は

$$\frac{\alpha}{2} = \int_{z_{\frac{\alpha}{2}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{n^2}{2}} dx$$

なる関係をもつ, 標準正規分布での上側確率 $\frac{\alpha}{2}$ 点の値である.

式 (1.83) において, P 内の不等式を変形すると

$$P\left(\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\frac{\alpha}{2}}\le\mu\le\bar{X}+\frac{\sigma}{\sqrt{n}}z_{\frac{\alpha}{2}}\right)=1-\alpha$$

となる. すなわち $1-\alpha$ の信頼係数の母平均 μ の信頼区間は

$$\left[\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\frac{\alpha}{2}}\ ,\ \bar{X}\frac{\sigma}{\sqrt{n}}z_{\frac{\alpha}{2}}\right]$$

ということになる. 例えば, $\alpha=0.05$ のとき, $Z_{\frac{\alpha}{2}}\cong 1.96$ となり, $1-\alpha=0.95$, すなわち 95% の信頼係数の μ の信頼区間は

$$\left[\bar{X}-1.96\frac{\sigma}{\sqrt{n}}\ ,\ \bar{X}+1.96\frac{\sigma}{\sqrt{n}}\right]$$

ということになる.

1.6.4 最小 2 乗法 (least square method, minimum mean square method)

least square method を直訳すると，本節表題の最小 2 乗法となる．一方, minimum mean square method は最小平均 2 乗法となるが，これを混用して，どちらも最小 2 乗法あるいは最小自乗法と呼ぶことが多い．

least square method(最小 2 乗法) は，どちらかというと確定論的な考え方からのパラメータ推定法である．次の問題を考える。すなわち，2 つの観測データの組 $\{y_t,\ x_t;\ t=1,\cdots,n\}$ に基づいて，θ を定めて未知パラメータ θ と x の関数 $h(x,\ \theta)$ を y に近似する，すなわち

$$y_t\approx h(x_t,\theta) \tag{1.84}$$

の近似問題を考える．このとき，θ の最適な推定値は，以下のように y_t と $h(x_t,\ \theta)$ との間の誤差

$$e_t=y_t-h(x_t,\theta),\quad t=1,\cdots,n \tag{1.85}$$

を考える．e_t は観測データについての近似した場合の誤差であり，残差 (residual) と呼ばれる．残差の 2 乗値の和を最小にするパラメータ θ の推定値を最小自乗推定値 (least square estimate)$\hat{\theta}$ という．すなわち

$$\hat{\theta}=J(\theta)\equiv\sum_{t=1}^{n}e_t^2\sum_{t=1}^{n}\Big[y_t-h(x_t,\theta)\Big]^2 \tag{1.86}$$

と定義し，$J(\theta)$ を最小化する θ の値を $\hat{\theta}$ とする．すなわち，数学の記号を用いると

$$\hat{\theta}=\arg\min J(\theta) \tag{1.87}$$

と表現できる．このとき，$\hat{\theta}$ が満たすべき条件（必要条件）は

$$\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\Big[y_i-h(x_,\theta)^2\Big]\Big|_{\theta=\hat{\theta}}=0 \tag{1.88}$$

である.

例

$$y_t\sim ax_t+b,\qquad t=1,2,\ldots,n \tag{1.89}$$

となる，$\theta = [a,\ b]^{\mathrm{T}}$ の最小自乗推定値 $\hat{\theta} = [\hat{a},\ \hat{b}]^{\mathrm{T}}$ を求めてみよう．すなわち，損失関数 (評価関数) を

$$J(\theta) \equiv \min_{\theta} \sum_{t=1}^{n} \Big[y_t - (ax_t + b) \Big]^2 \tag{1.90}$$

として，

$$J(\hat{\theta}) = \min_{\theta} J(\theta) \equiv \min_{\theta} \sum_{t=1}^{n} \Big[y_t - (ax_t + b) \Big]^2 \tag{1.91}$$

となる $\hat{\theta} = [\hat{a},\ \hat{b}]^{\mathrm{T}}$ を求める．(上式での $\hat{\theta}$ は arg の記号を用いて $\hat{\theta} = \arg\min J(\theta)$ と書ける)　ここで，

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad H = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, \quad \theta = \begin{bmatrix} a \\ b \end{bmatrix} \tag{1.92}$$

と表現すると

$$J(\theta) \equiv \sum_{t=1}^{n} \big[y_t - ax_t + b \big]^2 = (y - H\theta)^{\mathrm{T}}(y - H\theta) \tag{1.93}$$

となり，平方完成を利用すると

$$\begin{aligned} J(\theta) &= y^{\mathrm{T}}y - y^{\mathrm{T}}H\theta - \theta^{\mathrm{T}}H^{\mathrm{T}}H\theta \\ &= \Big[\theta - (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}y\Big]^{\mathrm{T}} H^{\mathrm{T}}H \Big[\theta - (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}y\Big] + y^{\mathrm{T}}y - y^{\mathrm{T}}H(H^{\mathrm{T}}H)^{-1}Hy \end{aligned} \tag{1.94}$$

となり，

$$\arg\min_{\theta} J(\theta) = (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}y \tag{1.95}$$

$$\min_{\theta} J(\theta) = y^{\mathrm{T}}y - y^{\mathrm{T}}H(H^{\mathrm{T}}H)^{-1}Hy \tag{1.96}$$

となる．

1.6.5 最小分散推定値について

確率変数 $X,\ Y$ に対して，$Y = y$　の条件のもとでの X の推定値 $\hat{x}$ を考える．推定誤差の大小の目安として，通常以下の規範を考える．

$$E\left\{ \|X - \hat{x}\|^2 | Y = y \right\}, \quad \text{ただし} \quad \|a\|^2 = a^T a$$

最小誤差分散推定値 $\hat{x}$ は以下の条件を満たすものとする．すなわち，y の関数であるどのようなベクトル z に対して

$$E\left\{ \|X - \hat{x}\|^2 | Y = y \right\} \leq E\left\{ \|X - z\|^2 | Y = y \right\}$$

が成立するものとする．このとき

$\hat{x}$ は**最小平均 2 乗推定値** (minimum mean-square estimates) と呼ばれる．

定理 1.2 *2*つの確率ベクトル X, Y に対して，確率ベクトル Y が y の値をとったと観測したものとする．また $\hat{x}$ を上記で定義した X の最小分散推定値とする．このとき，$\hat{x}$ は $Y = y$ が与えられたもとでの X 条件付平均値（期待値）として一意的に定まる．すなわち

$$\hat{x} = E\left[X|Y=y\right] = \int_{-\infty}^{\infty} x p_{X|Y}(x|y)dx$$

で与えられる．

定理 の証明

z を y の関数であってもよいが x の関数でないものとする．このとき

$$\begin{aligned} E\left\{\|X-z\|^2|Y=y\right\} &= \int_{-\infty}^{\infty}(x-z)^T(x-z)p_{X|Y}(x|y)dx \\ &= \int_{-\infty}^{\infty} x^T x p_{X|Y}(x|y)dx - \int_{-\infty}^{\infty} x^T p_{X|Y}(x|y)dxz \\ &\quad -z^T\int_{-\infty}^{\infty} x p_{X|Y}(x|y)dx + z^T z \\ &= \left[z^T - \int_{-\infty}^{\infty} x^T p_{X|Y}(x|y)dx\right]\left[z - \int_{-\infty}^{\infty} x p_{X|Y}(x|y)dx\right] \\ &\quad + \int_{-\infty}^{\infty} x^T x p_{X|Y}(x|y)dx - ||\int_{-\infty}^{\infty} x p_{X|Y}(x|y)dx||^2 \qquad (*) \end{aligned}$$

となる．ここで最後の等式の右辺は z の関数であり，

$$z = \int_{-\infty}^{\infty} x p_{X|Y}(x|y)dx = E\left[X|Y=y\right] \equiv \hat{x}$$

とおけば，右辺は最小となり，これを実現する z の値は**条件付期待値** (conditional mean estimate) となる．これを $\hat{x}$ と表記している．またこのとき，この推定値 $\hat{x}$ による推定値の 2 乗誤差の平均値は $((*))$ 式に $z = \hat{x}$ を代入することにより

$$\begin{aligned} E\left\{\|X-\hat{x}\|^2|Y=y\right\} &= \int_{-\infty}^{\infty} x^T x p_{X|Y}(x|y)dx - ||\int_{-\infty}^{\infty} x p_{X|Y}(x|y)dx||^2 \\ &= E\left\{\|X\|^2|Y=y\right\} - \|\hat{x}\|^2 \qquad (1.97) \end{aligned}$$

となる．

1.6.6　条件付期待値の性質

$Y = y$ が与えられたもとでの X の条件付期待値は y の関数であり，$\varphi(y) \equiv E\left\{X|Y=y\right\}$ この関数 $\varphi(y)$ を用いて，確率変数 $\varphi(Y)$ を考え

$$\varphi(Y) \equiv E\left\{X|Y\right\}$$

と表現する．このとき，次の関係が成立する．

a)　　$E\{\varphi(Y)\} \equiv E\left\{E[X|Y]\right\} = E[X].$

b) $$E\{E\{g(X,Y)|Y\}\} = E\{g(X,Y)\}$$

a) の証明:

$$\begin{aligned}
&\int_{-\infty}^{\infty} E[X|y]p_Y(y)dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x p_{X|Y}(x|y)dx \cdot p_Y(y)dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x p_{X,Y}(x,y)dxdy \\
&= \int_{-\infty}^{\infty} x p_X(x)dx = E[X]
\end{aligned}$$

b) の証明:

$$\begin{aligned}
&\int_{-\infty}^{\infty} p_Y(y)\Big[\int_{-\infty}^{\infty} g(x,y)p_{X|Y}(x|y)dx\Big]dy \\
= &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)p_{X,Y}(x,y)dxdy = E\{g(X,Y)\}
\end{aligned}$$

1.6.7 正規性確率ベクトルに対する条件付確率

X,Y を結合正規分布に従う 2 つに確率ベクトルとし，平均値，共分散を

$$E\left\{\begin{bmatrix} X \\ Y \end{bmatrix}\right\} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \mu$$

$$\mathrm{Cov}\left\{\begin{bmatrix} X \\ Y \end{bmatrix}\right\} = E\left\{\begin{bmatrix} X-\mu_x \\ Y-\mu_y \end{bmatrix}\cdot\begin{bmatrix}(X-\mu_x)^T & (Y-\mu_y)^T\end{bmatrix}\right\} = \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix} \equiv R$$

とする．このとき，以下に示すように

$$E[X|Y=y] = \mu_x + R_{xy}R_{yy}^{-1}(y-\mu_y) \tag{1.98}$$

および

$$E[X|Y] = \mu_x + R_{xy}R_{yy}^{-1}(Y-\mu_y) \tag{1.99}$$

が成立する．また

$$\begin{aligned}
E\{E[X|Y]\} &= \mu_x + R_{xy}R_{yy}^{-1}\left[E(Y)-\mu_y\right] \\
&= \mu_x = E[X]
\end{aligned}$$

となる．

(1.98) の証明

X: $n\times 1$,　Y: $m\times 1$　のサイズとする．このとき X, Y は結合正規分布に従う確率ベクトルであるので，条件付確率密度は

$$
\begin{aligned}
p_{X|Y}(x|y) &= \frac{p_{XY}(x,y)}{p_Y(y)} \\
&= \frac{\frac{1}{(2\pi)^{\frac{n+m}{2}}|R|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\begin{bmatrix} x^T-\mu_x^T & y^T-\mu_y^T \end{bmatrix} R^{-1}\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}\right\}}{\frac{1}{(2\pi)^{\frac{m}{2}}|R_{yy}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\begin{bmatrix} y^T-\mu_y^T \end{bmatrix} R_{yy}^{-1}\begin{bmatrix} y-\mu_y \end{bmatrix}\right\}}
\end{aligned}
$$

すなわち

$$
\begin{aligned}
p_{X|Y}(x|y) &= \frac{1}{(2\pi)^{\frac{n}{2}}}\cdot\frac{|R_{yy}|^{\frac{1}{2}}}{|R|^{\frac{1}{2}}} \\
&\quad\times\exp\left\{-\frac{1}{2}\begin{bmatrix} x^T-\mu_x^T & y^T-\mu_y^T \end{bmatrix} R^{-1}\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}+\frac{1}{2}\begin{bmatrix} y^T-\mu_y^T \end{bmatrix} R_{yy}^{-1}\begin{bmatrix} y-\mu_y \end{bmatrix}\right\}
\end{aligned} \tag{1.100}
$$

となる．

ここで，以下の等式が成立することに注目する．

$$
\begin{aligned}
&\begin{bmatrix} I & -R_{xy}R_{yy}^{-1} \\ O & I \end{bmatrix}\begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}\begin{bmatrix} I & O \\ -R_{yy}^{-1}R_{xy}^T & I \end{bmatrix} \\
=&\begin{bmatrix} R_{xx}-R_{xy}R_{yy}^{-1}R_{yx} & O \\ R_{yx} & R_{yy} \end{bmatrix}\begin{bmatrix} I & O \\ -R_{yy}^{-1}R_{xy}^T & I \end{bmatrix} \\
=&\begin{bmatrix} R_{xx}-R_{xy}R_{yy}^{-1}R_{yx} & O \\ O & R_{yy} \end{bmatrix}
\end{aligned}
$$

したがって

$$
\begin{aligned}
R &= \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix} \\
&= \begin{bmatrix} I & -R_{xy}R_{yy}^{-1} \\ O & I \end{bmatrix}^{-1}\begin{bmatrix} R_{xx}-R_{xy}R_{yy}^{-1}R_{yx} & O \\ O & R_{yy} \end{bmatrix}\begin{bmatrix} I & O \\ -R_{yy}^{-1}R_{xy}^T & I \end{bmatrix}^{-1}
\end{aligned} \tag{1.101}
$$

となり

$$
\begin{aligned}
R^{-1} &= \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}^{-1} \\
&= \left(\begin{bmatrix} I & -R_{xy}R_{yy}^{-1} \\ O & I \end{bmatrix}^{-1}\begin{bmatrix} R_{xx}-R_{xy}R_{yy}^{-1}R_{yx} & O \\ O & R_{yy} \end{bmatrix}\begin{bmatrix} I & O \\ -R_{yy}^{-1}R_{xy}^T & I \end{bmatrix}^{-1}\right)^{-1} \\
&= \left(\begin{bmatrix} I & O \\ -R_{yy}^{-1}R_{xy}^{\mathrm{T}} & I \end{bmatrix}\begin{bmatrix} \left(R_{xx}-R_{xy}R_{yy}^{-1}R_{yx}\right)^{-1} & O \\ O & R_{yy}^{-1} \end{bmatrix}\begin{bmatrix} I & -R_{xy}R_{yy}^{-1} \\ O & I \end{bmatrix}\right) \\
&= \begin{bmatrix} \left(R_{xx}-R_{xy}R_{yy}^{-1}R_{yx}\right)^{-1} & -\left(R_{xx}-R_{xy}R_{yy}^{-1}R_{yx}\right)^{-1}R_{xy}R_{yy}^{-1} \\ -R_{yy}^{-1}R_{xy}^{\mathrm{T}}\left(R_{xx}-R_{xy}R_{yy}^{-1}R_{yx}\right)^{-1} & R_{yy}^{-1}R_{xy}^{\mathrm{T}}\left(R_{xx}-R_{xy}R_{yy}^{-1}R_{yx}\right)^{-1}R_{xy}R_{yy}^{-1}+R_{yy}^{-1} \end{bmatrix}
\end{aligned} \tag{1.102}
$$

が成立する．ただし，三角行列に関する，以下の性質を用いている．すなわち，

$$
\begin{aligned}
\begin{bmatrix} I & O \\ B & I \end{bmatrix}^{-1} &= \begin{bmatrix} I & O \\ -B & I \end{bmatrix}, \qquad \begin{bmatrix} I & B \\ O & I \end{bmatrix}^{-1} = \begin{bmatrix} I & -B \\ O & I \end{bmatrix} \\
\left|\begin{bmatrix} I & O \\ B & I \end{bmatrix}\right| &= 1, \qquad \left|\begin{bmatrix} I & B \\ O & I \end{bmatrix}\right| = 1
\end{aligned}
$$

である．また (1.101) 式より

$$
\begin{aligned}
|R| &= |R_{xx}-R_{xy}R_{yy}^{-1}R_{yx}|\,|R_{yy}| && (1.103)\\
|R|^{\frac{1}{2}} &= |R_{xx}-R_{xy}R_{yy}^{-1}R_{yx}|^{\frac{1}{2}}|R_{yy}|^{\frac{1}{2}} && (1.104)
\end{aligned}
$$

となる．

ここで，(1.100) 式に (1.102) 式の関係を代入すると，(1.100) 式の指数の肩の第 1 項は，

$$
\begin{aligned}
&-\frac{1}{2}\left[\begin{array}{cc} x^{\mathrm{T}}-\mu_x^T & y^T-\mu_y^T \end{array}\right]R^{-1}\left[\begin{array}{c} x-\mu_x \\ y-\mu_y \end{array}\right]\\
=&-\frac{1}{2}\Big\{(x-\mu_x)^{\mathrm{T}}\big(R_{xx}-R_{xy}R_{yy}^{-1}R_{yx}\big)^{-1}(x-\mu_x)\\
&-(x-\mu_x)^{\mathrm{T}}\big(R_{xx}-R_{xy}R_{yy}^{-1}R_{yx}\big)^{-1}R_{xy}R_{yy}^{-1}(y-\mu_y)\\
&-(y-\mu_y)^{\mathrm{T}}R_{yy}^{-1}R_{xy}^{\mathrm{T}}\big(R_{xx}-R_{xy}R_{yy}^{-1}R_{yx}\big)^{-1}(x-\mu_x)^{\mathrm{T}}\\
&+(y-\mu_y)^{\mathrm{T}}\Big[R_{yy}^{-1}R_{xy}^{\mathrm{T}}\big(R_{xx}-R_{xy}R_{yy}^{-1}R_{yx}\big)^{-1}R_{xy}R_{yy}^{-1}+R_{yy}^{-1}\Big](y-\mu_y)\Big\}\\
=&-\frac{1}{2}\left[(x-\mu_x)-R_{xy}R_{yy}^{-1}(y-\mu_y)\right]^T(R_{xx}-R_{xy}R_{yy}^{-1}R_{yx})^{-1}\left[(x-\mu_x)-R_{xy}R_{yy}^{-1}(y-\mu_y)\right]\\
&-\frac{1}{2}\left[(y-\mu_y)^TR_{yy}^{-1}(y-\mu_y)\right] \qquad (1.105)
\end{aligned}
$$

となり，最終的に

$$
\begin{aligned}
p_{X|Y}(x|y) &= \frac{1}{(2\pi)^{\frac{n}{2}}|R_{xx}-R_{xy}R_{yy}^{-1}R_{yx}|^{\frac{1}{2}}}\\
&\quad\times\exp\left\{-\frac{1}{2}[x-\hat{x}]^T(R_{xx}-R_{xy}R_{yy}^{-1}R_{yx})^{-1}[x-\hat{x}]\right\}
\end{aligned}
$$

を得る．ただし

$$\hat{x}=\mu_x+R_{xy}R_{yy}^{-1}(y-\mu_y)$$

である．すなわち

$$E[X|Y=y]=\int_{-\infty}^{\infty}xp_{X|Y}(x|y)dx=\hat{x}$$

を得る．

ここで，$n\times 1$ 確率ベクトル X と $m\times 1$ 確率ベクトル Y について，独立 (independent)，無相関 (uncorrelate)，また直交 (orthogonal) についての定義を述べる．

定義 1　(独立な確率ベクトル)

連続型確率ベクトル X,Y が独立であるとは，X,Y の結合確率 pdf $p_{X,Y}(x,y)$ が各々の周辺確率密度関数の積で表される場合である．すなわち

$$p_{X,Y}(x,y)=p_X(x)p_Y(y) \qquad (1.106)$$

が成立すれば，X,Y は独立な確率ベクトルである．

定義 2　(無相関な確率ベクトル)

連続型確率ベクトル X,Y が無相関であるとは，XY^{T} の期待値が各々の期待値の積で表される場合である．すなわち

$$E\big[XY^{\mathrm{T}}\big]=E\big[X\big]E\big[Y^{\mathrm{T}}\big] \qquad (1.107)$$

が成立すれば，X,Y は無相関な確率ベクトルである．あるいは $\mathrm{Cov}\big\{\mathrm{X,Y}\big\}=0$ であれば無相関である．

定義 2 (直交する確率ベクトル)

連続型確率ベクトル X, Y の各要素が直交するとは，X の各要素 X_i と Y の各要素の内積 $E\big[X_iY_j\big]$ がゼロである場合である．すなわち

$$E\big[XY^{\mathrm{T}}\big] = O_{n,m}: \quad n \times m \text{ ゼロ行列} \tag{1.108}$$

が成立すれば，X, Y は直交する確率ベクトルである．

よく知られているように，確率ベクトル X, Y が独立であれば，無相関である．また X, Y が正規性確率ベクトルであれば，X, Y が無相関であれば，独立である．すなわち正規性確率ベクトル X, Y に対しては，無相関と独立は等価である．

1.7 線形最小分散推定 (linear minimum (error) variance estimation)

定理 1.3 *2* つの確率ベクトル $[X^T \quad Y^T]^T$ の平均値ベクトルおよび共分散行列は以下で与えられているものとする．

$$E\left\{\begin{bmatrix} X \\ Y \end{bmatrix}\right\} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \mu$$

$$\mathrm{Cov}\left\{\begin{bmatrix} X \\ Y \end{bmatrix}\right\} = E\left\{\begin{bmatrix} X-\mu_x \\ Y-\mu_y \end{bmatrix} \cdot \big[(X-\mu_x)^T \quad (Y-\mu_y)^T\big]\right\} = \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix} \equiv R$$

このとき，線形最小分散推定量は

$$\hat{E}[X|Y] \equiv \arg\min_{\hat{X}} E\left\{\|X - \hat{X}\|^2\right\} = \mu_x + R_{xy}R_{yy}^{-1}(Y - \mu_y)$$

となる．

証明: X の線形推定量を

$$\hat{X} = AY + b$$

とする．この推定量は，厳密にはアフィン推定量 (affine estimator) とよぶべきものであるが，習慣的に**線形推定量** (linear estimator) とよばれている．このときまず推定誤差の分散を計算すると以下のようになる．

$$
\begin{aligned}
& E\left\{\|X-\hat{X}\|^2\right\} \\
= \quad & E\left\{\mathrm{trace}\left[[X-\hat{X}][X-\hat{X}]^T\right]\right\} \\
= \quad & \mathrm{trace}E\left\{[X-AY-b][X-AY-b]^T\right\} \\
= \quad & \mathrm{trace}E\left\{[X-AY-(\mu_x-A\mu_y)+(\mu_x-A\mu_y)-b][X-AY-(\mu_x-A\mu_y)+(\mu_x-A\mu_y)-b]^T\right\} \\
= \quad & \mathrm{trace}\left[R_{xx}-AR_{yx}-R_{xy}A^T+AR_{yy}A^T\right]+\mathrm{trace}[(\mu_x-A\mu_y-b)(\mu_x-A\mu_y-b)^T] \\
& +\mathrm{trace}E\left\{[X-AY-(\mu_x-A\mu_y)][\mu_x-A\mu_y-b]^T+[\mu_x-A\mu_y-b][X-AY-(\mu_x-A\mu_y)]^T\right\} \\
= \quad & \mathrm{trace}\left[R_{xx}-AR_{yx}-R_{xy}A^T+AR_{yy}A^T\right]+\mathrm{trace}[(\mu_x-A\mu_y-b)(\mu_x-A\mu_y-b)^T] \\
= \quad & \mathrm{trace}[R_{xx}-AR_{yx}-R_{xy}A^T+AR_{yy}A^T]+\|\mu_x-A\mu_y-b\|^2 \\
= \quad & \mathrm{trace}\left[(A-R_{xy}R_{yy}^{-1})R_{yy}(A^T-R_{yy}^{-1}R_{yx})+R_{xx}-R_{xy}R_{yy}^{-1}R_{yx}\right]+\|\mu_x-A\mu_y-b\|^2 \\
= \quad & \mathrm{trace}\left[(A-R_{xy}R_{yy}^{-1})R_{yy}(A^T-R_{yy}^{-1}R_{yx})\right] \\
& +\mathrm{trace}[R_{xx}-R_{xy}Rz_{yy}^{-1}R_{yx}]+\|\mu_x-A\mu_y-b\|^2
\end{aligned}
$$

ただし，ベクトル α に対して，$\mathrm{trace}[\alpha\alpha^T]=\alpha^T\alpha$，が成立することを用いている．また

$$
\left\{\begin{array}{l} A^0=R_{xy}R_{yy}^{-1} \\ \mu_x-A^0\mu_y-b^0=\mathbf{0} \end{array}\right\} \Rightarrow b^0=\mu_x-A^0\mu_y
$$

であるので，最終的に (1.109) 式より

$$
\begin{aligned}
\hat{E}[X|Y] \quad = \quad & \arg\min_{\hat{X}} E\left\{\|X-\hat{X}\|^2\right\} \\
= \quad & A^0Y+b^0 \\
= \quad & A^0Y+\mu_x-A^0\mu_y \\
= \quad & \mu_x+A^0(Y-\mu_y) \\
= \quad & \mu_x+R_{xy}R_{yy}^{-1}(Y-\mu_y)
\end{aligned}
$$

なる関係を得る．以上の議論より，結合正規分布に従う確率変数 (ベクトル) に対する以下の有名な結果を得る．

定理 1.4 確率ベクトル X, Y が結合正規分布に従うとき, 最小分散推定量 *(minimum variance estimator)* と線形最小分散推定量 *(linear minimum variance estimator)* は一致する．

定理 1.5 直交原理 *(orthogonality principle)*

X と Y を結合確率ベクトルとする．このとき，線形推定誤差: $X-\hat{E}[X|Y]$ は Y に直交する．すなわち

$$
E\{[X-\hat{E}[X|Y]]Y^T\}=\mathbf{O}_{n,m}
$$

が成り立つ．

また逆にある行列 A とベクトル b に対して $E\{[X-AY-b]Y^T\} = \mathbf{O}_{n,m}$ 　(: $n \times m$ ゼロ行列 が成立すると

$$\begin{aligned} A &= R_{xy}R_{yy}^{-1} \\ b &= \mu_x - A\mu_y \end{aligned} \tag{1.109}$$

となる．

補足 (ベクトル確率変数に対する直交性について)

2 つの確率ベクトル X, Y (ベクトルサイズを $X : n \times 1, Y : m \times 1$ とする) に対して，X の各要素が Y の各要素に直交するとき, すなわち $E[x_i y_j] = 0;\ 1 \le i \le n, \quad 1 \le j \le m$ であるとき，確率ベクトル X と Y は直交するといい,.

$$E[\mathbf{XY}^T] = \mathbf{O}_{nm} : \quad n \times m \quad \text{ゼロ行列}$$

が成立する．

定理 1.6 X と Y を正規性確率ベクトルとする．このとき，推定誤差: $Z \equiv X - E[X|Y]$ とすると，Z と Y は独立である．

証明

Z も Y も正規性確率ベクトルであり，$E[Z]=0$，$E[Y]=\mu_y$ であるので，

$$\begin{aligned} \mathrm{Cov}\Big\{X - E[X|Y], Y\Big\} &= E\Big\{\big[X - E[X|Y]\big]\big[Y-\mu_y\big]^{\mathrm{T}}\Big\} \\ &= E\Big\{\big[X - \hat{E}[X|Y]\big]\big[Y-\mu_y\big]^{\mathrm{T}}\Big\} \\ &= E\Big\{\big[X - \mu_x - R_{xy}R_{yy}^{-1}\big]\big[Y-\mu_y\big]^{\mathrm{T}}\Big\} \\ &= R_{xy} - R_{xy}R_{yy}^{-1}R_{yy} \\ &= O \end{aligned} \tag{1.110}$$

となり，Z と Y 無相関な確率ベクトル，したがって正規性より独立な確率ベクトルとなる．

1.7.1　最尤推定法とスカラー線形回帰モデル

互いに独立なスカラー連続型確率変数 V_k, $(k=1,\ldots,n)$ に対して，その変換 Y_k, $(k=1,\ldots,n)$ を考える．すなわち，

$$Y_k = h_k\theta + V_k, \qquad k = 1,2,\ldots,n \tag{1.111}$$

を考える．ただし h_k は既知の確定的変数，θ は未知数とする．このとき，V_k が確率変数であるので Y_k も確率変数になる．

観測データ $y_1, y_2, \ldots, y_n$ を確率変数 $Y_1, Y_2, \ldots, Y_n$ の実現値と見なし，この確率変数に対する結合確率密度関数 $p_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n)$ に観測データの具体的に得られた実現値 (観測値) を代入し，

推定したい未知パラメータ (θ) の関数とみなす．これを**尤度関数** (likelihood function)$L(\theta)$ と呼ぶが，$L(\theta)$ の値を最大化する θ の値を**最尤推定値** (maximum likelihood estimate)$\hat{\theta}_{ML}$ という．すなわち，

$$\hat{\theta}_{ML} \equiv \arg\max_{\theta} L(\theta) \equiv \arg\max_{\theta} p_{V_1,\cdots,V_n}(v_1,\cdots,v_n|\theta)$$

また，対数関数の単調増加性より

$$\begin{aligned}\hat{\theta}_{ML} &\equiv arg\max_{\theta} L(\theta)\\ &= arg\max_{\theta} \ln L(\theta)\\ &= arg\max_{\theta} l(\theta)\end{aligned} \tag{1.112}$$

が成立し，$l(\theta) \equiv \ln L(\theta)$ は**対数尤度関数** (log-likelihood function) と呼ばれる．尤度関数の最大化の計算では，対数尤度関数の最大化の方が計算が容易な場合が多く，対数尤度関数がよく利用される．

いま，(1.111) 式に対して，V_k が正規性白色雑音の場合，観測データ Y_k の実現値 (観測値)$\{Y_k = y_k, k = 1, 2, \cdots, n\}$ が与えられた場合の，未知数 θ の最尤推定値を，以下で求める．

確率変数の変換による pdf の変換公式を用いると

$$\begin{aligned}Y_1 &= h_1\theta + V_1\\ &\vdots\\ Y_n &= h_n\theta + V_n\end{aligned}$$

より

$$\begin{aligned}V_1 &= Y_1 - h_1\theta\\ &\vdots\\ V_n &= Y_n - h_n\theta\end{aligned}$$

となり，

$$\begin{aligned}p_{Y_1,\cdots,Y_n}(y_1,\cdots,y_n) &= \left[p_{V_1,\cdots,V_n}(v_1,\cdots,v_n)\left|\frac{\partial(v_1,\cdots,v_n)}{\partial(y_1,\cdots,y_n)}\right|\right]_{v_1=y_1-h_1\theta,\cdots,v_n=y_n-h_n\theta}\\ &= p_{V_1}(v_1)\cdots p_{V_n}(v_n)\Big|_{v_1=y_1-h_1\theta,\cdots,v_n=y_n-h_n\theta}\\ &= \prod_{k=1}^{n} p_{V_k}(v_k)\Big|_{v_k=y_k-h_k\theta}\end{aligned} \tag{1.113}$$

ただし，Jacobian の絶対値は

$$||\frac{\partial(v_1,\cdots,v_n)}{\partial(y_1,\cdots,y_n)}|| = |\begin{vmatrix}\frac{\partial v_1}{\partial y_1} & \cdots & \frac{\partial v_1}{\partial y_n}\\ \vdots & & \vdots\\ \frac{\partial v_n}{\partial y_1} & \cdots & \frac{\partial v_n}{\partial y_n}\end{vmatrix}| \tag{1.114}$$

$$= |\begin{vmatrix}1 & & \\ & \ddots & \\ & & 1\end{vmatrix}| \tag{1.115}$$

$$= 1 \tag{1.116}$$

となっている．

ここで，尤度関数 $L(\theta)$，および対数尤度関数 $l(\theta)$ を以下で与えられる．

$$
\begin{aligned}
L(\theta) &= p_{Y_1,\cdots,Y_n}(y_1,\cdots,y_n|\theta) = \prod_{k=1}^{n} p_{V_k}(v_k|\theta) \\
&= \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_k-h_k\theta)^2}{2\sigma^2}} \qquad (1.117) \\
l(\theta) &= \ln L(\theta) = \ln p_{Y_1,\cdots,Y_n}(y_1,\cdots,y_n) = \ln \prod_{k=1}^{n} p_{V_k}(v_k) = \ln \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_k-h_k\theta)^2}{2\sigma^2}} \\
&= \sum_{k=1}^{n} \Big\{ -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{\big(y_k-h_k\theta\big)^2}{2\sigma^2} \Big\} \qquad (1.118)
\end{aligned}
$$

したがって，最尤推定値 $\hat{\theta}_{ML}$ は

$$
\begin{aligned}
\hat{\theta}_{ML} &= \arg\max_{\theta} \sum_{k=1}^{n} \Big\{ -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{\big(y_k-h_k\theta\big)^2}{2\sigma^2} \Big\} = \arg\max_{\theta} \sum_{k=1}^{n} \Big\{ -\frac{\big(y_k-h_k\theta\big)^2}{2\sigma^2} \Big\} \\
&= \arg\min_{\theta} \frac{1}{2\sigma^2} \sum_{k=1}^{n} \big\{ y_k - h_k\theta \big\}^2 \qquad (1.119)
\end{aligned}
$$

により求まる．すなわち，

$$
\begin{aligned}
\frac{d}{d\theta}\Big\{ \sum_{k=1}^{n} \big(y_k-h_k\theta\big)^2 \Big\} &= 2\sum_{k=1}^{n} \big(y_k-h_k\theta\big)(-h_k) \\
&= -2\Big\{ \sum_{k=1}^{n} h_k y_k + \sum_{k=1}^{n} h_k^2\theta \Big\} \equiv 0 \qquad (1.120)
\end{aligned}
$$

となり，最尤推定値は

$$
\hat{\theta}_{ML} = \frac{\sum_{k=1}^{n} h_k y_k}{\sum_{k=1}^{n} h_k^2} \qquad (1.121)
$$

となる．

1.7.2 クラメール・ラオ (Cramer-Rao) の不等式

実数のパラメータ θ に対する推定量 $\hat{\theta}(y_1,\ \cdots,\ y_n) \equiv \hat{\theta}(Y^n)$ の分散について考える．

定理 1.7 $\hat{\theta}(Y^n)$ が θ の不偏推定量，すなわち $E\{\hat{\theta}(Y^n)\} = \theta$ とする．

このとき

$$
Var[\ \hat{\theta}(Y^n) - \theta\ ] \ \geq\ \frac{1}{E\{[\frac{d\ \ln\ \ f(Y^n|\ \theta)}{d\theta}]^2\}} \equiv C_B
$$

（$Var[\ \hat{\theta}(Y^n)\ ]$ でよい．なぜなら $\hat{\theta}(Y^n)$ は不偏推定量であるので）

不等式は，また

$$
Var[\ \hat{\theta}(Y^n) - \theta\ ] \ \geq\ \frac{1}{E\{-\frac{d^2\ \ln\ \ f(Y^n|\ \theta)}{d\theta^2}\}} \ \equiv\ C_B
$$

とも表現できる．ただし $\frac{\partial f}{\partial\theta}$，$\frac{\partial^2 f}{\partial\theta^2}$ が存在し，かつ Y^n について絶対可積分とする．
右辺の値 C_B をクラメール・ラオの下界という．
したがって，ある不偏推定量 $\check{\theta}(Y^n)$ があって，

$$\mathrm{Var}(\check{\theta}(Y^n)) \;=\; C_B$$

であることが証明できれば $\check{\theta}(Y^n)$ は θ の不偏かつ最良の推定量．このような推定量を有効推定量 *(efficient estimator)* あるいは最小分散推定値 *(mininum variance estimator)* という．

証明
は以下の通り．

$\hat{\theta}(Y^n)$ は不偏推定量であるので

$$E[\,\hat{\theta}(Y^n)-\theta\,] \;=\; \int_{-\infty}^{\infty}[\,\hat{\theta}(Y^n)-\theta\,]\,f(Y^n|\,\theta)\,dY^n \;=\; 0 \tag{1.122}$$

が成立する．(1.122) 式を θ について微分すると，

$$\begin{aligned}
\frac{d}{d\theta}\int_{-\infty}^{\infty}(\,\hat{\theta}-\theta\,)\,f dY^n \;&=\; \int_{-\infty}^{\infty}\frac{d}{d\theta}\,\{(\,\hat{\theta}-\theta\,)f\}\,dY^n\\
&=\; \int_{-\infty}^{\infty}-f(\,Y^n|\,\theta\,)\,dY^n+\int_{-\infty}^{\infty}(\,\hat{\theta}-\theta\,)\frac{df(Y^n|\,\theta\,)}{d\theta}\,dY^n\\
&=\; 0
\end{aligned}$$

が成立する．上式中，

$$\int_{-\infty}^{\infty}-f(\,Y^n|\,\theta\,)\,dY^n \;=\; -1$$

であるので，

$$\int_{-\infty}^{\infty}(\,\hat{\theta}-\theta\,)\frac{df(\,Y^n|\,\theta\,)}{d\theta}\,dY^n \;=\; 1$$

となる．また，ここで，

$$\frac{d\ \ln\ f}{d\theta} \;=\; \frac{1}{f}\,\frac{df}{d\theta}$$

であるので，

$$\frac{df}{d\theta} \;=\; f\ \cdot\frac{d\ \ln\ f}{d\theta}$$

が成立し，

$$\int_{-\infty}^{\infty}(\,\hat{\theta}-\theta\,)\frac{d\ \ln\ f}{d\theta}f\,dY^n \;=\; 1 \tag{1.123}$$

となる．ここで有名なシュワルツ (Schwarz) の不等式，すなわち

$$\Big|\ \int_{-\infty}^{\infty}f(x)g(x)dx\ \Big| \;\le\; \Big(\ \int_{-\infty}^{\infty}f^2(x)dx\ \Big)^{\frac{1}{2}}\Big(\ \int_{-\infty}^{\infty}g^2(x)dx\ \Big)^{\frac{1}{2}}$$

ただし等号成立は $f(x)=kg(x)$ のときのみ．
一般的には，内積 $(\cdot\ ,\ \cdot)$ とノルム $\|\cdot\|$ を用いると，

$$|\,(\ f\ ,\ g\ )\,| \;\le\; \|f\|\ \cdot\ \|g\|$$

である．ここで (1.123) 式にシュワルツの不等式を適用すると，

$$\begin{aligned}
1 \;&=\; \Big|\ \int_{-\infty}^{\infty}(\,\hat{\theta}-\theta\,)\frac{d\ \ln\ f}{d\theta}f\,dY^n\Big| \;=\; \Big|\ \int_{-\infty}^{\infty}(\,\hat{\theta}-\theta\,)\sqrt{f}\cdot\frac{d\ \ln\ f}{d\theta}\sqrt{f}\,dY^n\ \Big|\\
&\le\; \Big(\ \int_{-\infty}^{\infty}(\,\hat{\theta}-\theta\,)^2 f\,dY^n\ \Big)^{\frac{1}{2}}\ \Big(\ \int_{-\infty}^{\infty}(\ \frac{d\ \ln\ f}{d\theta}\ )^2 f\,dY^n\ \Big)^{\frac{1}{2}}
\end{aligned}$$

したがって

$$\int_{-\infty}^{\infty} (\ \hat{\theta} - \theta\ )^2\ f(Y^n|\ \theta\ )\ dY^n\ \geq\ \frac{1}{\int_{-\infty}^{\infty}\ \left(\ \frac{d\ \ln\ f}{d\theta}\ \right)^2 f\ dY^n}$$

が成立する．ただし等号成立は，

$$\frac{d\ \ln\ f(\ Y^n|\ \theta\ )}{d\theta}\ =\ k(\theta)\ (\hat{\theta}(\ Y^n\ ) - \theta)$$

の場合であり，$k(\theta)$ は Y^n の関数でないことである．

また

$$\int_{-\infty}^{\infty} f(\ Y^n|\ \theta\ )\ dY^n\ =\ 1$$

であり，両辺を θ で微分すると，

$$\int_{-\infty}^{\infty} \frac{df}{d\theta}\ dY^n\ =\ \int_{-\infty}^{\infty} \frac{d\ \ln\ f}{d\theta} f\ dY^n\ =\ 0$$

となり，再び θ で微分すると

$$\int_{-\infty}^{\infty} \frac{d^2\ \ln\ f}{d\theta^2} f\ dY^n + \int_{-\infty}^{\infty} \frac{d\ \ln\ f}{d\theta}\ \frac{df}{d\theta}\ dY^n\ =\ 0$$

となる．2 項について，

$$\int_{-\infty}^{\infty} \frac{d\ \ln\ f}{d\theta}\ \frac{df}{d\theta}\ dY^n\ =\ \int_{-\infty}^{\infty} \frac{d\ \ln\ f}{d\theta}\ \frac{d\ \ln\ f}{d\theta} f\ dY^n$$

が成立するので，

$$\int_{-\infty}^{\infty} \frac{d^2\ \ln\ f}{d\theta^2} f\ dY^n\ =\ -\int_{-\infty}^{\infty} \left(\ \frac{d\ \ln\ f}{d\theta}\ \right)^2 f\ dY^n$$

が成立し，

$$E\ \left\{\ \left[\frac{d\ \ln\ f(Y^n|\ \theta)}{d\theta}\ \right]^2\right\}\ =\ -E\ \left\{\frac{d^2\ \ln\ f(Y^n|\ \theta)}{d\theta^2}\right\}$$

が成立する．

不規則パラメータに対する最小 2 乗誤差推定量の 2 乗誤差の下界

定理 1.8 θ を確率変数とし，$Y^n = [y_1,\ \cdots,\ y_n]$ を観測ベクトルとする．このとき θ のすべての推定量 $\hat{\theta}(\ Y^n\ )$ は以下の不等式を満たす．

$$\begin{aligned} E\ \{\ [\ \hat{\theta}(\ Y^n\ ) - \theta\ ]^2\ \} &\geq (\ E\{\ (\frac{\partial\ \ln\ f(Y^n,\ \theta)}{\partial\theta})^2\}^{-1}) \\ &= \{\ E\ [\ \frac{\partial^2\ \ln\ f(Y^n,\ \theta)}{\partial\theta^2}\ ]\ \}^{-1} \end{aligned}$$

ただし $E[\ \ ]$ は Y^n と θ についての期待値をとるものとし，次の条件を仮定する．

1. $\frac{\partial\ f(Y^n,\ \theta)}{\partial\theta}$ は Y^n と θ について絶対可積分とする．
2. $\frac{\partial^2\ f(Y^n,\ \theta)}{\partial\theta^2}$ は Y^n と θ について絶対可積分とする．
3. θ が与えられたもとでの誤差の条件付期待値：

$$B(\theta) = \int_{-\infty}^{\infty} [\ \hat{\theta}(Y^n) - \theta\ ]\ f(Y^n|\ \theta)\ dY^n$$

に対して

$$\lim_{\theta\to\infty}\ B(\theta)\ f(\theta)\ =\ 0$$
$$\lim_{\theta\to-\infty}\ B(\theta)\ f(\theta)\ =\ 0$$

が成立するものとする．

証明

$$
\begin{aligned}
\frac{d\, f(\theta)\ B(\theta)}{d\theta} &= \frac{d}{d\theta}\left[\ f(\theta)\int_{-\infty}^{\infty}[\ \hat{\theta}(Y^n)-\theta\ ]\ f(Y^n|\ \theta)\ dY^n\ \right] \\
&= \frac{d}{d\theta}\left[\ \int_{-\infty}^{\infty}[\ \hat{\theta}(Y^n)-\theta\ ]\ f(Y^n,\ \theta)\ dY^n\ \right] \\
&= -\int_{-\infty}^{\infty} f(Y^n,\ \theta)\ dY^n + \int_{-\infty}^{\infty}[\ \hat{\theta}(Y^n)-\theta\ ]\ \frac{\partial\ f(Y^n,\ \theta)}{\partial\theta}\ dY^n
\end{aligned}
$$

両辺を θ について積分すると

$$
f(\theta)\ B(\theta)\ \Big|_{\theta=-\infty}^{\infty} = -1 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\ \hat{\theta}(Y^n)-\theta\ ]\ \frac{\partial\ \ln\ f(Y^n,\ \theta)}{\partial\ \theta}\ f(Y^n,\ \theta)\ dY^n\ d\theta
$$

$B(\theta)\ f(\theta)$ は $\theta\ =\ \pm\infty$ で 0 であるので

$$
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\ \hat{\theta}(Y^n)-\theta\ ]\ \frac{\partial\ \ln\ f(Y^n,\ \theta)}{\partial\theta}\ f(Y^n,\ \theta)\ dY^n\ d\theta = 1
$$

シュワルツの不等式を適用すると

$$
\left[\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\ \hat{\theta}(Y^n)-\theta\ ]^2\ f(Y^n,\ \theta)\ dY^n\ d\theta\ \right]^{\frac{1}{2}} \times \left[\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\frac{\partial\ \ln\ f(Y^n,\ \theta)}{\partial\theta})^2\ f(Y^n,\ \theta)\ dY^n\ d\theta\ \right]^{\frac{1}{2}} \geqq 1
$$

したがって

$$
E\ \{\ [\ \hat{\theta}(Y^n)-\theta\ ]^2\}\ \ \geq \{\Big[\ \frac{\partial\ \ln\ f(Y^n,\ \theta)}{\partial\theta}\ \Big]^2\}^{-1}
$$

ここで

$$
\int\int f(Y^n,\ \theta)\ dY^n\ d\theta\ =\ 1
$$

を θ で微分することにより，

$$
\int\int \frac{\partial\ f}{\partial\ \theta}\ dY^n\ d\theta\ =\ \int\int \frac{\partial\ \ln\ f}{\partial\ \theta}\ f\ dY^n\ d\theta\ =\ 0
$$

を得る．もう一度 θ で微分することにより，

$$
\int\int \frac{\partial^2\ \ln\ f}{\partial\ \theta^2}\ f\ dY^n\ d\theta\ +\ \int\int \frac{\partial\ \ln\ f}{\partial\ \theta}\ \frac{\partial\ \ln\ f}{\partial\ \theta}\ f\ dY^n\ d\theta\ =\ 0
$$

となる．すなわち

$$
\int\int (\frac{\partial\ \ln\ f}{\partial\ \theta})^2\ f\ dY^n\ d\theta\ = -\int \frac{\partial^2\ \ln\ f}{\partial\ \theta^2}\ f\ dY^n\ d\theta
$$

となる．したがって

$$
\begin{aligned}
E\left[\ (\frac{\partial\ \ln\ f(\ Y^n,\ \theta\ )}{\partial\ \theta})^2\ \right] &= -E\left[\ \frac{\partial^2\ \ln\ f}{\partial\ \theta^2}\ f\ dY^n\ d\theta\ \right] \\
&= -E\left[\ \frac{\partial^2\ \ln\ f(\ Y^n|\ \theta\ )}{\partial\ \theta^2}\ +\ \frac{\partial^2\ \ln\ f(\ \theta\ )}{\partial\ \theta^2}\ \right]
\end{aligned}
$$

が成立する．

したがって

$$
E\{\ [\ \hat{\theta}(\ Y^n\ )-\theta\ ]^2\ \}\ \geq\ \{-E\left[\ \frac{\partial^2\ \ln\ f(\ Y^n|\ \theta\ )}{\partial\ \theta^2}\ \right] - E\left[\ \frac{\partial^2\ \ln\ f(\ \theta\ )}{\partial\ \theta^2}\ \right]\}^{-1}
$$

となる．また等号が成立するのは，

$$\frac{\partial\ \ln\ f(\ Y^n,\ \theta\ )}{\partial\ \theta}\ =\ k\ [\ \hat{\theta}(\ Y^n\ )-\theta\ ]$$

の関数が成立するときのみである．これを θ で再び微分すると，

$$\frac{\partial^2\ \ln\ f(\ Y^n,\ \theta\ )}{\partial\ \theta^2}\ =\ -k$$

この関係は事後確率 $f(\ Y^n|\ \theta\ )$ に対する関係式 $f(\ Y^n,\ \theta\ )\ =\ f(\ Y^n|\ \theta\ )f(\ Y^n\ )$ より

$$\frac{\partial^2\ \ln\ f(\ Y^n|\ \theta\ )}{\partial\ \theta^2}\ =\ -k$$

が成立する．
したがってこの関係から

$$\ln\ f(\ Y^n|\ \theta\ )\ =\ -k\theta^2+c_1\theta+c_2$$

を得る．したがってこの関係から，有効推定量が存在するためには θ についての事後確率密度関数は正規分布でなければならない．

1.8 線形回帰式

次の線形回帰式を考えよう．

$$y_k=H_k\theta+v_k,\ k=0,1,\cdots,t \tag{1.124}$$

ただし

$$y_k\in\mathbf{R}^m,\ \theta\in\mathbf{R}^n$$

とする．また v_k は正規性白色雑音とし，

$$E[v_k]=0,\qquad E[v_kv_k^{\mathrm{T}}]=R_k$$

とする．このとき，尤度関数 $L(\theta)$，および対数尤度関数 $l(\theta)$ を以下で与えられる．

$$\begin{aligned}
L(\theta) &= p_{Y_0,\cdots,Y_t}(y_0,\cdots,y_t|\theta)=\prod_{k=0}^{t}p_{V_k}(y_k-H_k\theta)\\
&= \prod_{k=0}^{t}\frac{1}{\sqrt{2\pi}^p\big|R_k\big|^{1/2}}\exp\Big\{-\frac{1}{2}(y_k-H_k\theta)^{\mathrm{T}}R_k^{-1}(y_k-H_k\theta)\Big\}\\
l(\theta) &= \ln L(\theta)=\ln\prod_{k=0}^{t}\frac{1}{\sqrt{2\pi}^p\big|R_k\big|^{1/2}}\exp\Big\{-\frac{1}{2}(y_k-H_k\theta)^{\mathrm{T}}R_k^{-1}(y_k-H_k\theta)\Big\}\\
&= \sum_{k=0}^{t}\Big\{-\frac{1}{2}\ln[(2\pi)^p|R_k|]-\frac{1}{2}(y_k-H_k\theta)^{\mathrm{T}}R_k^{-1}(y_k-H_k\theta)\Big\}
\end{aligned} \tag{1.125}$$

となる．すなわち，最尤推定値は

$$\begin{aligned}
\hat{\theta}_t &= \arg\ \min_{\theta}\sum_{k=0}^{t}(y_k-H_k\theta)^{\mathrm{T}}R_k^{-1}(y_k-H_k\theta)=\arg\ \min_{\theta}J(\theta)\\
& \quad v_k=y_k-H_k\theta
\end{aligned}$$

により求めることを考える．また，上記は重み付き最小自乗法による推定値でもある．

最小化のため，上式右辺を θ の平方完成を目指して以下のように変形できる．

$$\begin{aligned}
& J(\theta) \\
\equiv\ & \sum_{k=0}^{t}(y_k - H_k\theta)^{\mathrm{T}}R_k^{-1}(y_k - H_k\theta) \\
=\ & \sum_{k=0}^{t} y_k^{\mathrm{T}}R_k^{-1}y_k - \sum_{k=0}^{t} y_k^{\mathrm{T}}R_k^{-1}H_k\theta - \sum_{k=0}^{t}\theta^{\mathrm{T}}H_k^{T}R_k^{-1}y_k + \theta^{\mathrm{T}}\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}H_k\theta \\
=\ & \left[\theta - \left(\sum_{k=0}^{t}H_k^{\mathrm{T}}R_kH_k\right)^{-1}\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}y_k\right]^{\mathrm{T}}\left(\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}H_k\right)\left[\theta - \left(\sum_{k=0}^{t}H_k^{\mathrm{T}}R_kH_k\right)^{-1}\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}y_k\right] \\
& + \sum_{k=0}^{t} y_k^{\mathrm{T}}R_k^{-1}y_k - \sum_{k=0}^{t} y_k^{\mathrm{T}}R_k^{-1}H_k\left(\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}H_k\right)^{-1}\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}y_k
\end{aligned}$$

したがって

$$\hat{\theta}_t = \left(\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}H_k\right)^{-1}\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}y_k \tag{1.126}$$

とすると，$J(\theta)$ を最小化でき，

$$\min_{\theta} J(\theta) = \sum_{k=0}^{t} y_k^{\mathrm{T}}R_k^{-1}y_k - \sum_{k=0}^{t} y_k^{\mathrm{T}}R_k^{-1}H_k\left(\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}H_k\right)^{-1}\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}y_k \tag{1.127}$$

となる．また

$$y_k = H_k\theta + v_k \tag{1.128}$$

であるので，(1.126) 式に (1.128) 式を代入すると

$$\hat{\theta}_t = \theta + \left(\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}H_k\right)^{-1}\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}v_k$$

となる．また

$$\begin{aligned}
& \min_{\theta}\sum_{k=0}^{t}\Bigl(y_k - H_k\theta\Bigr)R_k^{-1}\Bigl(y_k - H_k\theta\Bigr) \\
&= \sum_{k=0}^{t} y_k^{\mathrm{T}}R_k^{-1}y_k - \sum_{k=0}^{t} y_k^{\mathrm{T}}R_k^{-1}H_k\left(\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}H_k\right)^{-1}\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}y_k \\
&= \sum_{k=0}^{t} y_k^{\mathrm{T}}R_k^{-1}y_k - \sum_{k=0}^{t} y_k^{\mathrm{T}}R_k^{-1}H_k\underline{\hat{\theta}}_t
\end{aligned}$$

と表現できる．

1.8.1 推定値の統計的性質

ここで推定値の統計的性質を調べてみよう．すなわち

$$\begin{aligned}
E\Bigl[\underline{\hat{\theta}}_t\Bigr] &= E\Bigl[\left(\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}H_k\right)^{-1}\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}y_k\Bigr] \\
&= E\Bigl[\left(\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}H_k\right)^{-1}\sum_{k=0}^{t}H_k^{\mathrm{T}}R_k^{-1}\Bigl(H_k\theta + v_k\Bigr)\Bigr] \qquad (1.129)
\end{aligned}$$

となり，$E[v_k] = 0$ あるので

$$E\Bigl[\underline{\hat{\theta}}_t\Bigr] = \theta \tag{1.130}$$

となり, 不偏推定値 (unbiased estimate) となる. また共分散は,

$$
\begin{aligned}
& E\Big\{\big[\hat{\theta}_t-\theta\big]\big[\hat{\theta}_t-\theta\big]^{\mathrm{T}}\Big\} \\
= \;& E\Big\{\Big[(\sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} H_k)^{-1} \sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} v_k\Big]\Big[(\sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} H_k)^{-1} \sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} v_k\Big]^{\mathrm{T}}\Big\} \\
= \;& E\Big\{\Big[(\sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} H_k)^{-1} \sum_{i=0}^{t} H_i^{\mathrm{T}} R_i^{-1} v_i\Big]\Big[\sum_{k=0}^{t} v_k^{\mathrm{T}} R_k^{-1} H_k (\sum_{k=0}^{t} H_k^{T} R_k^{-1} H_k)^{-1}\Big]\Big\} \\
= \;& \big(\sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} H_k\big)^{-1}\big(\sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} R_k R_k^{-1} H_k\big)\big(\sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} H_k\big)^{-1} \\
= \;& \big(\sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} H_k\big)^{-1}
\end{aligned}
$$

となる.

1.8.2 $\hat{\theta}_t$ の逐次式について

$$\hat{\theta}_t = \big(\sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} H_k\big)^{-1}\big(\sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} y_k\big)$$

について逐次式を求めてみよう.

$$\hat{\theta}_{t+1} = \left(\sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} H_k + H_{t+1}^{\mathrm{T}} R_{t+1}^{-1} H_{t+1}\right)^{-1}\left(\sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} y_k + H_{t+1}^{\mathrm{T}} R_{t+1}^{-1} y_{t+1}\right)$$

であるので，上式右辺の最初の項を P_{t+1} とすれば，すなわち

$$
\begin{aligned}
P_{t+1} \;&\equiv\; \left(\sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} H_k + H_{t+1}^{\mathrm{T}} R_{t+1}^{-1} H_{t+1}\right)^{-1} \\
&=\; \left(P_t^{-1} + H_{t+1}^{\mathrm{T}} R_{t+1}^{-1} H_{t+1}\right)^{-1}
\end{aligned}
$$

となる．ここで**逆行列の補題** (この詳細については後に説明する):

$$
\begin{aligned}
(A+BCD)^{-1} \;&=\; A^{-1}(I+BCDA^{-1})^{-1} \\
&=\; A^{-1}[I-B(DA^{-1}B+C^{-1})^{-1}DA^{-1}] \\
&=\; A^{-1}-A^{-1}B(DA^{-1}B+C^{-1})^{-1}DA^{-1}
\end{aligned}
\tag{1.131}
$$

を用いると,

$$(A^{-1}+B^{\mathrm{T}}CB)^{-1} = A - AB^{\mathrm{T}}(BAB^{\mathrm{T}}+C^{-1})^{-1}BA \tag{1.132}$$

が成立する．したがって

$$
\begin{aligned}
P_{t+1} \;&=\; \left(P_t^{-1} + H_{t+1}^{\mathrm{T}} R_{t+1}^{-1} H_{t+1}\right)^{-1} \\
&=\; P_t - P_t H_{t+1}^{\mathrm{T}}\big[H_{t+1} P_t H_{t+1}^{\mathrm{T}} + R_{t+1}\big]^{-1} H_{t+1} P_t
\end{aligned}
$$

となる．したがって

$$
\begin{aligned}
\hat{\theta}_{t+1} &= P_{t+1}\left(\sum_{k=0}^{t} H_k^{-1} R_k^{-1} y_k + H_{t+1} R_{t+1}^{-1} y_{t+1}\right) \\
&= \left(P_t - P_t H_{t+1}^{\mathrm{T}}[H_{t+1} P_t H_{t+1}^{\mathrm{T}} + R_{t+1}]^{-1} H_{t+1} P_t\right)\left(\sum_{k=0}^{t} H_k^{\mathrm{T}} R_k^{-1} y_k + H_{t+1}^{\mathrm{T}} R_{t+1}^{-1} y_{t+1}\right) \\
&= \hat{\theta}_t - P_t H_{t+1}^{\mathrm{T}}[H_{t+1} P_t H_{t+1}^{\mathrm{T}} + R_{t+1}]^{-1} H_{t+1} \hat{\theta}_t \\
&\quad + P_t H_{t+1}^{\mathrm{T}} R_{t+1}^{-1} y_{t+1} - P_t H_{t+1}^{\mathrm{T}}[H_{t+1} P_t H_{t+1}^{\mathrm{T}} + R_{t+1}]^{-1} H_{t+1} P_t H_{t+1}^{\mathrm{T}} R_{t+1}^{-1} y_{t+1}
\end{aligned}
$$

ここで，最後の 2 項は, 以下のように変形できる．

$$
\begin{aligned}
& P_t H_{t+1}^{\mathrm{T}} R_{t+1}^{-1} y_{t+1} - P_t H_{t+1}^{\mathrm{T}}[H_{t+1} P_t H_{t+1}^{\mathrm{T}} + R_{t+1}]^{-1} H_{t+1} P_t H_{t+1}^{\mathrm{T}} R_{t+1}^{-1} y_{t+1} \\
=\ & P_t H_{t+1}^{\mathrm{T}}[H_{t+1} P_t H_{t+1}^{\mathrm{T}} + R_{t+1}]^{-1}[H_{t+1} P_t H_{t+1}^{\mathrm{T}} + R_{t+1}] R_{t+1}^{-1} y_{t+1} \\
& - P_t H_{t+1}^{\mathrm{T}}[H_{t+1} P_t H_{t+1}^{\mathrm{T}} + R_{t+1}]^{-1} H_{t+1} P_t H_{t+1}^{\mathrm{T}} R_{t+1}^{-1} y_{t+1} \\
=\ & P_t H_{t+1}^{\mathrm{T}}[H_{t+1} P_t H_{t+1}^{\mathrm{T}} + R_{t+1}]^{-1}\left[(H_{t+1} P_t H_{t+1}^{\mathrm{T}} + R_{t+1}) R_{t+1}^{-1} y_{t+1} - H_{t+1} P_t H_{t+1}^{\mathrm{T}} R_{t+1}^{-1} y_{t+1}\right] \\
=\ & P_t H_{t+1}^{\mathrm{T}}[H_{t+1} P_t H_{t+1}^{\mathrm{T}} + R_{t+1}]^{-1} y_{t+1}
\end{aligned}
$$

したがって

$$
\hat{\theta}_{t+1} = \hat{\theta}_t + P_t H_{t+1}^{\mathrm{T}}[H_{t+1} P_t H_{t+1}^{\mathrm{T}} + R_{t+1}]^{-1}\left[y_{t+1} - H_{t+1}\hat{\theta}_t\right]
$$

を得る．また P_t は，次の逐次式で計算できる．

$$
\begin{aligned}
P_{t+1} &= P_t - P_t H_{t+1}^{\mathrm{T}}\left[H_{t+1} P_t H_{t+1}^{\mathrm{T}} + R_{t+1}\right]^{-1} H_{t+1} P_t \\
& \quad P_0 = \left[H_0 R_1^{-1} H_0\right]^{-1}
\end{aligned}
$$

(注)　実は，上記の関係式は，カルマンフィルタ理論を学ぶと，簡単に導出できる．

逆行列の補題について

$$
\begin{aligned}
(A + BCD)^{-1} &= A^{-1}(I + BCDA^{-1})^{-1} \\
&= A^{-1}[I - B(DA^{-1}B + C^{-1})^{-1} DA^{-1}] \\
&= A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1} DA^{-1}
\end{aligned}
$$

この証明のために，まず

$$
(I + FG)^{-1}
$$

を求めよう．

$$
(I + FG)(I - X) = I + FG - (I + FG)X \equiv I
$$

より，

$$
X = (I + FG)^{-1} FG
$$

となり，

$$
(I + FG)^{-1} = I - (I + FG)^{-1} FG
$$

と逆行列が求まる．

ここで，F^{-1} が存在すれば，

$$
\begin{aligned}
(I+FG)^{-1} &= I-(I+FG)^{-1}FG \\
&= I-\Big[F(F^{-1}+G)\Big]^{-1}FG \\
&= I-(F^{-1}+G)^{-1}F^{-1}FG \\
&= I-(F^{-1}+G)^{-1}G
\end{aligned}
\tag{1.133}
$$

となる．同様に，G^{-1} が存在すれば，

$$
\begin{aligned}
(I+FG)^{-1} &= I-(I+FG)^{-1}FG \\
&= I-\Big[(G^{-1}+F)G\Big]^{-1}FG \\
&= I-G^{-1}(G^{-1}+F)^{-1}FG
\end{aligned}
$$

と変形できる．(1.133) 式において，$F=BC, G=DA^{-1}$ とし，かつ A^{-1}, C^{-1} が存在すれば，

$$
\begin{aligned}
(I+BCDA^{-1})^{-1} &= I-\Big[(BC)^{-1}+DA^{-1}\Big]^{-1}DA^{-1} \\
&= I-\Big[C^{-1}B^{-1}+DA^{-1}\Big]^{-1}DA^{-1} \\
&= I-\Big[(DA^{-1}B+C^{-1})B^{-1}\Big]^{-1}DA^{-1} \\
&= I-B\Big[DA^{-1}B+C^{-1}\Big]^{-1}DA^{-1}
\end{aligned}
$$

が成立し，逆行列の補題:

$$
\begin{aligned}
A^{-1}(I+BCDA^{-1})^{-1} &= A(I-B\Big[DA^{-1}B+C^{-1}\Big]^{-1}DA^{-1}) \\
&= A^{-1}-A^{-1}B(DA^{-1}B+C^{-1})^{-1}DA^{-1}
\end{aligned}
\tag{1.134}
$$

が成立する．

1.9 拘束条件付線形回帰式

再び，次の線形回帰式を考える．

$$y = H\theta + v$$

ただし

$$y \in \mathbf{R}^m,\ \theta \in \mathbf{R}^n$$

とする．また v は正規性白色雑音とし，

$$E[v]=0, \qquad E[vv^{\mathrm{T}}]=R$$

とする．ここで以下の拘束条件を考える．

$$c = B\theta, \quad (c : r\times 1, \quad B : r\times n) \tag{1.135}$$

このとき，拘束条件付の最小化問題を考え，

$$\min_{\theta}(y-H\theta)^{\mathrm{T}}R^{-1}(y-H\theta)+\lambda^{\mathrm{T}}(B\theta-c)$$

なる θ を求めることを考える．ただし，λ は，ラグランジュ乗数 (Lagrange multiplier) である．

ここで

$$J(\theta) \quad = \quad (y - H\theta)^{\mathrm{T}} R^{-1} (y - H\theta) + \lambda^{\mathrm{T}} (B\theta - c)$$

とおき，$J(\theta)$ の最小化のため，$J(\theta)$ は θ の平方完成を目指して以下のように変形できる．

$$\begin{aligned}
J(\theta) &= y^{\mathrm{T}} R^{-1} y - \theta^{\mathrm{T}} H^{\mathrm{T}} R^{-1} H\theta - y^{\mathrm{T}} R^{-1} H\theta - \theta^{\mathrm{T}} H^{\mathrm{T}} R^{-1} y + \lambda^{\mathrm{T}} B\theta - \lambda^{\mathrm{T}} c \\
&= \left\{\theta - (H^{\mathrm{T}} R^{-1} H)^{-1} \left(H^{\mathrm{T}} R^{-1} y - \frac{1}{2} B\lambda\right)\right\}^{\mathrm{T}} H^{\mathrm{T}} R^{-1} H \left\{\theta - (H^{\mathrm{T}} R^{-1} H)^{-1} \left(H^{\mathrm{T}} R^{-1} y - \frac{1}{2} B\lambda\right)\right\} \\
&\quad + y^{\mathrm{T}} R^{-1} y - \lambda^{\mathrm{T}} c + \left(H^{\mathrm{T}} R^{-1} y - \frac{1}{2} B^{\mathrm{T}} \lambda\right)^{\mathrm{T}} (H^{\mathrm{T}} R^{-1} H)^{-1} \left(H^{\mathrm{T}} R^{-1} y - \frac{1}{2} B^{\mathrm{T}} \lambda\right)
\end{aligned}$$

したがって

$$\hat{\theta} = \left(H^{\mathrm{T}} R^{-1} H\right)^{-1} \left(H^{\mathrm{T}} R^{-1} y - \frac{1}{2} B^{\mathrm{T}} \lambda\right)$$

とすると，$J(\theta)$ を最小化できる．

また，ここで拘束条件 (1.135) を用いて，

$$B\hat{\theta} = B(H^{\mathrm{T}} R^{-1} H)^{-1} \left(H^{\mathrm{T}} R^{-1} y - \frac{1}{2} B^{\mathrm{T}} \lambda\right) = c$$

より

$$B(H^{\mathrm{T}} R^{-1} H)^{-1} H^{\mathrm{T}} R^{-1} y - \frac{1}{2} B(H^{\mathrm{T}} R^{-1} H)^{-1} B^{\mathrm{T}} \lambda = c$$

または

$$B(H^{\mathrm{T}} R^{-1} H)^{-1} H^{\mathrm{T}} R^{-1} y - c = \frac{1}{2} B(H^{\mathrm{T}} R^{-1} H)^{-1} B^{\mathrm{T}} \lambda$$

となる．したがって，ラグランジュ乗数は

$$\lambda = 2 \left[B(H^{\mathrm{T}} R^{-1} H)^{-1} B^{\mathrm{T}}\right]^{-1} \left[B(H^{\mathrm{T}} R^{-1} H)^{-1} H^{\mathrm{T}} R^{-1} y - c\right]$$

と求まる．ここで，

$$\begin{aligned}
\hat{\theta} &= (H^{\mathrm{T}} R^{-1} H)^{-1} \left(H^{\mathrm{T}} R^{-1} y - \frac{1}{2} B^{\mathrm{T}} \lambda\right) \\
&= (H^{\mathrm{T}} R^{-1} H)^{-1} \Big\{H^{\mathrm{T}} R^{-1} (H\theta + v) \\
&\qquad - \frac{1}{2} B^{\mathrm{T}} 2 \left[B(H^{\mathrm{T}} R^{-1} H)^{-1} B^{\mathrm{T}}\right]^{-1} \left[B(H^{\mathrm{T}} R^{-1} H)^{-1} H^{\mathrm{T}} R^{-1} y - c\right]\Big\} \\
&= (H^{\mathrm{T}} R^{-1} H)^{-1} \left\{H^{\mathrm{T}} R^{-1} y - B^{\mathrm{T}} \left[B(H^{\mathrm{T}} R^{-1} H)^{-1} B^{\mathrm{T}}\right]^{-1} \left[B(H^{\mathrm{T}} R^{-1} H)^{-1} H^{\mathrm{T}} R^{-1} y - c\right]\right\} \\
&= (H^{\mathrm{T}} R^{-1} H)^{-1} \Big\{\left[I - B^{\mathrm{T}} \left[B(H^{\mathrm{T}} R^{-1} H)^{-1} B^{\mathrm{T}}\right]^{-1} B(H^{\mathrm{T}} R^{-1} H)^{-1}\right] H^{\mathrm{T}} R^{-1} y \\
&\qquad + B^{\mathrm{T}} \left[B(H^{\mathrm{T}} R^{-1} H)^{-1} B^{\mathrm{T}}\right]^{-1} c\Big\} \\
&= (H^{\mathrm{T}} R^{-1} H)^{-1} \Big\{\left[I - B^{\mathrm{T}} \left[B(H^{\mathrm{T}} R^{-1} H)^{-1} B^{\mathrm{T}}\right]^{-1} B(H^{\mathrm{T}} R^{-1} H)^{-1}\right] H^{\mathrm{T}} R^{-1} (H\theta + v) \\
&\qquad + B^{\mathrm{T}} \left[B(H^{\mathrm{T}} R^{-1} H)^{-1} B^{\mathrm{T}}\right]^{-1} c\Big\}
\end{aligned}$$

したがって

$$\begin{aligned}
\hat{\theta} - \theta &= (H^{\mathrm{T}} R^{-1} H)^{-1} \Big\{H^{\mathrm{T}} R^{-1} v - B^{\mathrm{T}} \left[B(H^{\mathrm{T}} R^{-1} H)^{-1} B^{\mathrm{T}}\right]^{-1} B(H^{\mathrm{T}} R^{-1} H)^{-1} H^{\mathrm{T}} R^{-1} (H\theta + v) \\
&\qquad + B^{\mathrm{T}} \left[B(H^{\mathrm{T}} R^{-1} H)^{-1} B^{\mathrm{T}}\right]^{-1} c\Big\} \\
&= (H^{\mathrm{T}} R^{-1} H)^{-1} \left\{H^{\mathrm{T}} R^{-1} v - B^{\mathrm{T}} \left[B(H^{\mathrm{T}} R^{-1} H)^{-1} B^{\mathrm{T}}\right]^{-1} B(H^{\mathrm{T}} R^{-1} H)^{-1} H^{\mathrm{T}} R^{-1} v\right\} \\
&= (H^{\mathrm{T}} R^{-1} H)^{-1} \left\{H^{\mathrm{T}} R^{-1} - B^{\mathrm{T}} \left[B(H^{\mathrm{T}} R^{-1} H)^{-1} B^{\mathrm{T}}\right]^{-1} B(H^{\mathrm{T}} R^{-1} H)^{-1} H^{\mathrm{T}} R^{-1}\right\} v
\end{aligned}$$

ここで,

$$
\begin{aligned}
\mathrm{Cov}[\hat{\theta}] &= E\Big[(\hat{\theta}-\theta)(\hat{\theta}-\theta)^{\mathrm{T}}\Big] \\
&= (H^{\mathrm{T}}R^{-1}H)^{-1}\{H^{\mathrm{T}}R^{-1} - B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}]^{-1}B(H^{\mathrm{T}}R^{-1}H)^{-1}H^{\mathrm{T}}R^{-1}\}R \\
&\qquad\quad \{H^{\mathrm{T}}R^{-1} - B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}B(H^{\mathrm{T}}R^{-1}H)^{-1}H^{\mathrm{T}}R^{-1}\}^{\mathrm{T}}(H^{\mathrm{T}}R^{-1}H)^{-1} \\
&= (H^{\mathrm{T}}R^{-1}H)^{-1}\{H^{\mathrm{T}} - B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}]^{-1}B(H^{\mathrm{T}}R^{-1}H)^{-1}H^{\mathrm{T}}\} \\
&\qquad\quad \{R^{-1}H - R^{-1}H(H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}]^{-1}B\}(H^{\mathrm{T}}R^{-1}H)^{-1} \\
&= (H^{\mathrm{T}}R^{-1}H)^{-1}\{(I - B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}]^{-1}B(H^{\mathrm{T}}R^{-1}H)^{-1})H^{\mathrm{T}}R^{-1}H\} \\
&\qquad\quad \times(I - (H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}]^{-1}B\}(H^{\mathrm{T}}R^{-1}H)^{-1}
\end{aligned}
$$

したがって

$$
\begin{aligned}
\mathrm{Cov}[\hat{\theta}] &= (H^{\mathrm{T}}R^{-1}H)^{-1}(H^{\mathrm{T}}R^{-1}H - B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}]^{-1}B) \\
&\qquad\quad [(H^{\mathrm{T}}R^{-1}H)^{-1} - (H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}]B^{\mathrm{T}}]^{-1}B(H^{\mathrm{T}}R^{-1}H)^{-1}] \\
&= (H^{\mathrm{T}}R^{-1}H)^{-1}[I - B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}]^{-1}B(H^{\mathrm{T}}R^{-1}H)^{-1} \\
&\qquad\quad -B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}]^{-1}B(H^{\mathrm{T}}R^{-1}H)^{-1} \\
&\qquad\quad +B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}]^{-1}B(H^{\mathrm{T}}R^{-1}H)B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}]^{-1}B(H^{\mathrm{T}}R^{-1}H)^{-1}] \\
&= (H^{\mathrm{T}}R^{-1}H)^{-1}[I - 2B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}]^{-1}B(H^{\mathrm{T}}R^{-1}H)^{-1} \\
&\qquad\qquad\qquad +B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}]^{-1}B(H^{\mathrm{T}}R^{-1}H)^{-1}] \\
&= (H^{\mathrm{T}}R^{-1}H)^{-1}[I - B^{\mathrm{T}}[B(H^{\mathrm{T}}R^{-1}H)^{-1}B^{\mathrm{T}}]^{-1}B(H^{\mathrm{T}}R^{-1}H)^{-1}]
\end{aligned}
$$

第2章　カルマンフィルタ

カルマンフィルタ (Kalman Filter; ASME,1960) は離散時間システム (Discrete-time systems) についての理論であり，連続時間システム (Continuous-time systems) についてはカルマンービューシィフィルター (Kalman-Bucy Filter; ASME, 1961) といわれる．

ここでは，線形回帰式での未知パラメター θ が時間的に変動する場合について考える．すなわち $x_t = \theta_t$ なる確率過程 x_t を考え，x_t は以下のシステム方程式 (状態方程式: state equation) で与えられているものとする．

2.1　状態空間表現

システム方程式 [(system equation)　　ベクトル差分方程式 (vector difference equation)]:

$$x_{t+1} = Fx_t + Gw_t, \quad t \geq 0$$

$$\text{初期値:} \quad x_0 \quad \text{確率変数 (random variables)}$$

観測方程式 (Observation equations)　　線形回帰式

$$y_t = Hx_t + v_t, \quad t \geq 0$$

ここで，初期値 x_0 についての先験情報として

$$\left.\begin{aligned} E\left[x_0\right] &= \bar{x}_0 \\ \mathrm{Cov}\left[x_0\right] = E[(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T] &= \Sigma_0 \end{aligned}\right\}$$

である正規分布が与えられているものとする．また雑音 w_t, v_t は平均値 0 の正規性白色雑音とし，統計量は

$$E\left[w_t\right] = 0, \quad E\left[v_t\right] = 0$$

$$\left\{\begin{aligned} E\left[w_t w_\tau^T\right] &= Q\delta_{t-\tau} \\ E\left[v_t v_\tau^T\right] &= R\delta_{t-\tau} \quad (\delta_{t-\tau}\text{は Kronecker の}\delta\text{関数}) \end{aligned}\right.$$

あるいは

$$E\left\{\begin{bmatrix} w_t \\ v_t \end{bmatrix} \begin{bmatrix} w_\tau^T & v_\tau^T \end{bmatrix}\right\} = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix}\delta_{t-\tau}$$

ここでは簡単化のため $\{w_t\}$ と $\{v_t\}$ は無相関（独立）であることを仮定する．したがって

$$S = 0$$

となる．また以下の議論は F, G, H, R, Q が時間関数 F_t, G_t, H_t, R_t, Q_t の場合でも同じ結果を示すことができる．

2.2 状態変数の平均値，分散行列について

状態変数の平均値は以下の関係を満たす。

$$\bar{x}_{t+1} \equiv E\{x_{t+1}\} = E\{Fx_t + Gw_t\} = FE\{x_t\} + GE\{w_t\}$$

であるので，

$$\bar{x}_{t+1} = F\bar{x}_t$$

また，分散行列は

$$\begin{aligned}\Sigma_{t+1} &\equiv Cov\{x_{t+1}\} \\ &= E\{[Fx_t + Gw_t - F\bar{\mathbf{x}}_t][Fx_t + Gw_t - F\bar{x}_t]^T\} \\ &= E\{[F(x_t - \bar{x}_t) + Gw_t][F(x_t - \bar{x}_t) + Gw_t]^T\} \\ &= FE\{[x_t - \bar{x}_t][x_t - \bar{x}_t]^T\}F^T + GE\{w_t w_t^T\}G^T\end{aligned}$$

となる．ただし，上式最後の等式では $E\{[x_t - \bar{x}_t]w_t^T\} = 0$, $E\{w_t[x_t - \bar{x}_t]^T\} = 0$ の関係を用いている．

$$\Sigma_{t+1} = F\Sigma_t F^T + GQG^T$$

となる．上式は離散時間システムに対するリアプノフ差分方程式 (Lyapunov differece equation) と呼ばれている．

ここで，考察する問題は，与えられた観測データの列，

$$Y^t \equiv \{y_0, y_1, \cdots, y_t\}$$

から，

$$E\{\|x_t - \hat{x}_{t|t}\|^2|Y^t\}$$

を最小にする推定値 (フィルタ推定値)$\hat{x}_{t|t}$ を求める問題，および

$$E\{\|x_{t+1} - \hat{x}_{t+1|t}\|^2|Y^t\}$$

を最小にする推定値 (1 段予測推定値)$\hat{x}_{t+1|t}$ を求める問題を考察する．そのため，以下，観測値の一段予測誤差，イノベーション過程について述べる．
次に，以下の重要な定理（イノベーション過程について）を示す．

2.3 イノベーション過程

観測ベクトルの列 $\{y_0, y_1, y_2, \cdots, y_t\} \equiv Y_t$ に対して，

$$\begin{aligned}\nu_0 &= y_0 - E[y_0|Y_{-1}] = y_0 - E[y_0] \\ \nu_1 &= y_1 - E[y_1|Y_0] = y_1 - E[y_1] \\ \nu_2 &= y_2 - E[y_2|Y_1] \\ &\cdots\cdots \\ \nu_t &= y_t - E[y_t|Y^{t-1}]\end{aligned}$$

と定義する．このとき $\{y_0, y_1, y_2, \cdots, y_t\}$ が正規性確率ベクトルとすると

$$\begin{aligned} E[y_t|Y^{t-1}] &= \hat{E}[y_t|Y^{t-1}] \\ &= \hat{E}[y_t|y_0, \cdots, y_{t-1}] \\ &= \sum_{i=0}^{t-1} A_{ti} y_i + b_t \end{aligned}$$

と表現でき，任意の $t \quad (t = 1, 2, \cdots)$ に対して，

$$\begin{bmatrix} \nu_0 \\ \nu_1 \\ \vdots \\ \vdots \\ \nu_t \end{bmatrix} = \begin{bmatrix} I & & & & \\ -A_{10} & I & & & \\ \vdots & & & & \\ \vdots & & & & \\ -A_{t1} & \cdots & \cdots & -A_{t,t-1} & I \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ \vdots \\ y_t \end{bmatrix} - \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ \vdots \\ b_t \end{bmatrix}$$

を得る．したがって，ν_t は $\{y_0, y_1, \cdots, y_t, b_t\}$ の１次結合として表され，上式での行列はブロック下三角行列であり，対角ブロックに $m \times m$ の単位行列が並び正則であるので，y_t は $\{\nu_0, \nu_1, \cdots, \nu_t, b_1, \cdots, b_t\}$ の１次結合として表される．また

$$\begin{aligned} \nu_k &= y_k - E[y_k|Y_{k-1}] \\ \nu_t &= y_t - E[y_t|Y^{t-1}] \end{aligned}$$

より，$t > k$ とすると，ν_t と Y^{t-1} は直交しているので

$$E[\nu_t \nu_k^T] = \mathbf{O}_{m \times m}$$

となり，互いに無相関 (正規性の仮定の下では独立) である．$\{\nu_0, \nu_1, \cdots, \nu_t\}$ を**イノベーション過程** (innovation process) と呼ばれる．

定理 2.1 確率ベクトル $X, \nu_0, \nu_1, \cdots, \nu_t$ に対して，$\nu_0, \nu_2, \cdots, \nu_t$ は互いに無相関とする *(mutually uncorrelated)*，すなわち

$$E[\nu_i \nu_j^T] = E[\nu_i] E[\nu_j], \quad i \neq j$$

であるとき，

$$\begin{aligned} \Sigma_{\nu_i \nu_j} &= E[(\nu_i - \mu_i)(\nu_j - \mu_j)^T] \\ &= E[\nu_i \nu_j^T] - \mu_i \mu_j^T \\ &= E[\nu_i] E[\nu_j^T] - \mu_i \mu_j^T = \mathbf{O}_{m \times m} \quad (i \neq j) \end{aligned}$$

となる．このとき，$\nu_0, \nu_1, \cdots, \nu_t$ が与えられたもとでの線形最小分散推定量，$\hat{E}[X|\nu_0, \nu_1, \cdots, \nu_t]$ は（正規性を仮定すると前定理より最小分散推定量でもあり，$\hat{E}[X|\nu_0, \nu_1, \cdots, \nu_t] = E[X|\nu_0, \nu_1, \cdots, \nu_t]$: 条件付期待値，に一致する）

$$\hat{E}[X|\nu_0, \nu_1, \cdots, \nu_t] = \hat{E}[X|\nu_0] + \hat{E}[X|\nu_1] + \cdots + \hat{E}[X|\nu_t] - t\mu_x$$

証明:

$$W \equiv \begin{bmatrix} \nu_0 \\ \nu_1 \\ \vdots \\ \nu_t \end{bmatrix}$$

とし，線形最小分散推定に関する結果を用いると

$$\hat{E}[X|W] = \mu_x + \Sigma_{XW}\Sigma_{WW}^{-1}(W - \mu_W)$$

であり，Σ_{XW}, Σ_{WW} について以下の関係が成立する．すなわち

$$\begin{aligned}\Sigma_{XW} &= E\left\{(X-\mu_x)\begin{bmatrix} (\nu_0-\mu_{\nu_0})^T & (\nu_1-\mu_{\nu_1})^T & \cdots & (\nu_t-\mu_{\nu_t})^T \end{bmatrix}\right\} \\ &= \begin{bmatrix} \Sigma_{X\nu_0} & \Sigma_{X\nu_1} & \cdots & \Sigma_{X\nu_t} \end{bmatrix}\end{aligned}$$

$$\begin{aligned}\Sigma_{WW} &= E\left\{\begin{bmatrix} \nu_0-\mu_{\nu_0} \\ \nu_1-\mu_{\nu_1} \\ \vdots \\ \nu_t-\mu_{\nu_t} \end{bmatrix}\begin{bmatrix} (\nu_0-\mu_{\nu_0})^T & \cdots & (\nu_t-\mu_{\nu_t})^T \end{bmatrix}\right\} \\ &= \begin{bmatrix} \Sigma_{\nu_0\nu_0} & & & 0 \\ & \Sigma_{\nu_1\nu_1} & & \\ & & \ddots & \\ 0 & & & \Sigma_{\nu_t\nu_t} \end{bmatrix}\end{aligned}$$

ここに，最後の行列はブロック対称行列 (block symmetric matrix) である．また

$$\Sigma_{WW}^{-1} = \begin{bmatrix} \Sigma_{\nu_0\nu_0}^{-1} & & & 0 \\ & \Sigma_{\nu_1\nu_1}^{-1} & & \\ & & \ddots & \\ 0 & & & \Sigma_{\nu_t\nu_t}^{-1} \end{bmatrix}$$

となるので

$$\begin{aligned}&\hat{E}[X|W] = \mu_x + \Sigma_{XW}\Sigma_{WW}^{-1}(W-\mu_W) \\ &= \mu_x + \begin{bmatrix} \Sigma_{X\nu_0} & \Sigma_{X\nu_1} & \cdots & \Sigma_{X\nu_t} \end{bmatrix}\begin{bmatrix} \Sigma_{\nu_0\nu_0}^{-1} & & & 0 \\ & \Sigma_{\nu_1\nu_1}^{-1} & & \\ & & \ddots & \\ 0 & & & \Sigma_{\nu_t\nu_t}^{-1} \end{bmatrix}\begin{bmatrix} \nu_0-\mu_{\nu_0} \\ \nu_1-\mu_{\nu_1} \\ \vdots \\ \nu_t-\mu_{\nu_t} \end{bmatrix} \\ &= \mu_x + \sum_{j=0}^{t}\left[\Sigma_{X\nu_j}\Sigma_{\nu_j\nu_j}^{-1}(\nu_j-\mu_{\nu_j})\right] + t\mu_x - t\mu_x \\ &= \hat{E}[X|\nu_0] + \hat{E}[X|\nu_1] + \cdots + \hat{E}[X|\nu_t] - t\mu_x\end{aligned}$$

となる．

2.4 逐次式

ここで $\hat{x}_{t|t-1}$ が与えられたもとでの $\hat{x}_{t|t}$ に対する逐次式 (recursive equation) を求める．

$\hat{x}_{t|t-1}$ と y_t による $\hat{x}_{t|t}$ の表現

$$\begin{aligned}\hat{E}(x_t|Y^t) &= \hat{E}(x_t|Y^{t-1}, y_t) \\ &= \hat{E}(x_t|Y^{t-1}, \nu_t), \quad \nu_t \perp Y^{t-1}\end{aligned}$$

$$\nu_t \equiv y_t - H\hat{x}_{t|t-1} = H[x_t - \hat{x}_{t|t-1}] + v_t$$

ここで，以下の関係が成立する．

$$\hat{E}[(x_t|Y^{t-1}, y_t] = \hat{E}[x_t|Y^{t-1}, \nu_t]$$
$$\nu_t \perp Y^{t-1}$$

したがって

$$\begin{aligned} \hat{x}_{t|t} &= \hat{E}[x_t|Y^t] \\ &= \hat{E}[x_t|Y_{t-1}] + \hat{E}[x_t|\nu_t] - \mu_{x_t} \\ &= \hat{x}_{t|t-1} + \hat{E}[x_t|\nu_t] \end{aligned} \tag{2.1}$$

また

$$\hat{E}\left(x_t|\nu_t\right) = \mu_{x_t} + E[x_t\nu_t^T]\{E[\nu_t\nu_t^T]\}^{-1}(\nu_t - \mu_{\nu_t}) \tag{2.2}$$

が成立する．ここで，上式 (2.2) の右辺の各項を評価しよう．

$$\begin{aligned} E\left[x_t\nu_t^T\right] &= E\left\{x_t\left[H(x_t - \hat{x}_{t|t-1}) + v_t\right\}^T\right\} \\ &= E\left\{\left[x_t - \hat{x}_{t|t-1} + \hat{x}_{t|t-1}\right]\left[H(x_t - \hat{x}_{t|t-1}) + v_t\right\}^T\right\} \\ &= \Sigma_{t|t-1}H^T \end{aligned}$$

ただし，上式では以下の関係を用いた．

$$\begin{aligned} E\left[x_t\tilde{x}_{t|t-1}^T\right] &= E\left\{\left[\hat{x}_{t|t-1} + \tilde{x}_{t|t-1}\right]\tilde{x}_{t|t-1}^T\right\} \\ &= E\left[\tilde{x}_{t|t-1}\tilde{x}_{t|t-1}^T\right] \\ &\qquad \left(\text{ここでは直交原理} : \hat{x}_{t|t-1} \perp \tilde{x}_{t|t-1}\text{を用いた}\right) \\ &\equiv \Sigma_{t|t-1} \end{aligned}$$

また，(2.2) 式，右辺第 2 項目は

$$\begin{aligned} &E\left[\nu_t\nu_t^T\right] \\ =\ &E\left\{\left[H\left(x_t - \hat{x}_{t|t-1}\right) + v_t\right]\left[H\left(x_t - \hat{x}_{t|t-1}\right) + v_t\right]^T\right\} \\ =\ &E\left\{\left[H\tilde{x}_{t|t-1} + v_t\right]\left[H\tilde{x}_{t|t-1} + v_t\right]^T\right\} \\ =\ &H\Sigma_{t|t-1}H^T + R \end{aligned}$$

となる．ここで，以上の関係式をまとめると，以下の式を得る．

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t[y_t - H\hat{x}_{t|t-1}] \tag{2.3}$$
$$K_t = \Sigma_{t|t-1}H^T\left[H_t\Sigma_{t|t-1}H^T + R\right]^{-1}, \quad \text{((Kalman Gain))} \tag{2.4}$$

また，1 段予測推定式は

$$\begin{aligned} \hat{x}_{t+1|t} &= E\Big\{x_{t+1}|Y^t\Big\} \\ &= E\Big\{Fx_t + Gw_t|Y^t\Big\} \\ &= F\hat{x}_{t|t} \end{aligned} \tag{2.5}$$

となる．つぎに，$\Sigma_{t|t-1}$ を既知として $\Sigma_{t|t}$ についての逐次式を求める．

誤差共分散 $\Sigma_{t|t}$ の逐次式

$$\begin{aligned}\Sigma_{t|t} &\equiv E\left\{\left[x_t-\hat{x}_{t|t}\right]\left[x_t-\hat{x}_{t|t}\right]^T\right\} \\ &= E\left\{\left[x_t-\hat{x}_{t|t-1}-K_t\nu_t\right]\left[x_t-\hat{x}_{t|t-1}-K_t\nu_t\right]^T\right\}\end{aligned} \tag{2.6}$$

ここに，

$$\begin{aligned}E\left\{(x_t-\hat{x}_{t|t-1})\nu_t^{\mathrm{T}}K_t^{\mathrm{T}}\right\} &= E\left\{(x_t-\hat{x}_{t|t-1})[(x_t-\hat{x}_{t|t-1})^{\mathrm{T}}H^{\mathrm{T}}+v_t]K_t^{\mathrm{T}}\right\} \\ &= \Sigma_{t|t-1}H^{\mathrm{T}}K_t^{\mathrm{T}} \\ &= \Sigma_{t|t-1}H_t^T\left[H_t\Sigma_{t|t-1}H_t^T+R\right]^{-1}H_t\Sigma_{t|t-1}\end{aligned} \tag{2.7}$$

であり，

$$\begin{aligned}E\left\{K_t\nu_t(x_t-\hat{x}_{t|t-1})^{\mathrm{T}}\right\} &= E\left\{K_t[H(x_t-\hat{x}_{t|t-1})+v_t](x_t-\hat{x}_{t|t-1})^{\mathrm{T}}\right\} \\ &= K_tH\Sigma_{t|t-1} \\ &= \Sigma_{t|t-1}H_t^T\left[H_t\Sigma_{t|t-1}H_t^T+R\right]^{-1}H_t\Sigma_{t|t-1}\end{aligned} \tag{2.8}$$

である．また

$$\begin{aligned}E\left[K_t\nu_t\nu_t^{\mathrm{T}}K_t^{\mathrm{T}}\right] &= \Sigma_{t|t-1}H_t^T\left[H_t\Sigma_{t|t-1}H_t^T+R\right]^{-1}\left[H_t\Sigma_{t|t-1}H_t^T+R\right]\left[H_t\Sigma_{t|t-1}H_t^T+R\right]^{-1}H_t\Sigma_{t|t-1} \\ &= \Sigma_{t|t-1}H_t^T\left[H_t\Sigma_{t|t-1}H_t^T+R\right]^{-1}H_t\Sigma_{t|t-1}\end{aligned} \tag{2.9}$$

であるので，

$$\Sigma_{t|t}=\Sigma_{t|t-1}-K_tH_t\Sigma_{t|t-1}=\Sigma_{t|t-1}-\Sigma_{t|t-1}H_t^T\left[H_t\Sigma_{t|t-1}H_t^T+R\right]^{-1}H_t\Sigma_{t|t-1} \tag{2.10}$$

となる．

2.5　カルマンフィルタまとめ

$$\begin{aligned}\hat{x}_{t+1|t} &= F\hat{x}_{t|t} \\ \hat{x}_{t|t} &= \hat{x}_{t|t-1}+K_t[y_t-H_t\hat{x}_{t|t-1}] \\ K_t &= \Sigma_{t|t-1}H_t^T\left[H_t\Sigma_{t|t-1}H_t^T+R\right]^{-1},\quad ((\text{Kalman Gain})) \\ \Sigma_{t+1|t} &= F\Sigma_{t|t}F^T+GQG^T \\ & \qquad \text{初期条件}\left(\begin{array}{l}\hat{x}_{0|-1}=\bar{x}_0 \\ \Sigma_{0|-1}=\Sigma_0\end{array}\right. \\ \Sigma_{t|t} &= \Sigma_{t|t-1}-K_tH_t\Sigma_{t|t-1}=\Sigma_{t|t-1}-\Sigma_{t|t-1}H_t^T\left[H_t\Sigma_{t|t-1}H_t^T+R\right]^{-1}H_t\Sigma_{t|t-1}\end{aligned}$$

- $\hat{x}_{t+1|t+1} = \hat{x}_{t+1|t}+\Sigma_{t+1|t}H_t^T\left[H_t\Sigma_{t+1|t}H_t^T+R\right]^{-1}\left[y_{t+1}-H_t\hat{x}_{t+1|t}\right]$

2.6 ベイズの定理によるカルマンフィルタの導出

ここで，(Y. C. Ho and R.C.K. Lee: IEEE Trans. AC, 1964) によるベイズの定理に従ってカルマンフィルタを導出する方法について述べる．

同じく，次の状態空間モデルについて考察する．

$$x_{t+1} = Fx_t + Gw_t \tag{2.11}$$

$$y_t = Hx_t + v_t \tag{2.12}$$

ただし, x_t は n 次元正規性状態ベクトル, y_t は m 次元観測ベクトル, F, G, H はそれぞれ $n \times n$, $n \times r$, $m \times n$ 定数行列である．$\{w_t\}$, $\{v_t\}$ は互いに独立な, r 次元および m 次元正規性白色雑音ベクトルとする．

このとき，

$$p(w_t, v_{t+1}|x_t, Y^t) = p(w_t, v_{t+1}) = p(w_t)p(v_{t+1}) \tag{2.13}$$

$$E(w_t) = E(v_{t+1}) = \mathbf{0} \tag{2.14}$$

$$\mathrm{Cov}(w_t) = Q, \quad \mathrm{Cov}(v_{t+1}) = R \tag{2.15}$$

とする．

ここで, つぎのベイズの定理に着目する．

$$\begin{aligned} p(x_{t+1}|Y^{t+1}) &= p(x_{t+1}|y_{t+1}, Y^t) = \frac{p(x_{t+1}, y_{t+1}, Y^t}{p(y_{t+1}, Y^t)} = \frac{p(y_{t+1}|x_{t+1}, Y^t)p(x_{t+1}, Y^t)}{p(y_{t+1}, Y^t)} \\ &= \frac{p(y_{t+1}|x_{t+1}, Y^t)p(x_{t+1}|Y^t)}{p(y_{t+1}|Y^t)} = \frac{p(y_{t+1}|x_{t+1})p(x_{t+1}|Y^t)}{p(y_{t+1}|Y^t)} \end{aligned} \tag{2.16}$$

と表される．次に (2.16) 式での右辺の各項を評価する．$p(x_{t+1}|Y^t)$ は正規分布に従い, w_t は x_t, Y^t とは独立であるので，

$$E(x_{t+1}|Y^t) = E(Fx_t + Gw_t|Y^t) = F\hat{x}_{t|t} \equiv \hat{x}_{t+1|t} \tag{2.17}$$

$$\begin{aligned} \mathrm{Cov}(x_{t+1}|Y^t) &= E\left[\left\{Fx_t + Gw_t - F\hat{x}_{t|t}\right\}\left\{Fx_t + Gw_t - F\hat{x}_{t|t}\right\}^{\mathrm{T}}|Y^t\right] \\ &= FP_{t|t}F^{\mathrm{T}} + GQG^{\mathrm{T}} \equiv P_{t+1|t} \end{aligned} \tag{2.18}$$

となる．同様に $p(y_{t+1}|Y^t)$ も正規分布に従うので

$$E(y_{t+1}|Y^t) = E(Hx_{t+1} + v_{t+1}|Y^t) = H\hat{x}_{t+1|t} = HF\hat{x}_{t|t} \tag{2.19}$$

$$\begin{aligned} \mathrm{Cov}(y_{t+1}|Y^t) &= E\left[\left\{Hx_{t+1} + v_{t+1} - H\hat{x}_{t+1|t}\right\}\left\{Hx_{t+1} + v_{t+1} - H\hat{x}_{t+1|t}\right\}^{\mathrm{T}}|Y^t\right] \\ &= HP_{t+1|t}H^{\mathrm{T}} + R \end{aligned} \tag{2.20}$$

となる．また，$p(y_{t+1}|x_{t+1})$ も正規分布に従うので，その平均値，共分散行列を求めると，

$$E(y_{t+1}|x_{t+1}) = E(Hx_{t+1} + v_{t+1}|x_{t+1}) = Hx_{t+1} \tag{2.21}$$

$$\mathrm{Cov}(y_{t+1}|x_{t+1}) = E\left[\{Hx_{t+1} + v_{t+1} - Hx_{t+1}\}\{Hx_{t+1} + v_{t+1} - Hx_{t+1}\}^{\mathrm{T}}|x_{t+1}\right] = R \tag{2.22}$$

となる．従って (2.17) ∼ (2.22) 式より (2.16) 式は

$$
\begin{aligned}
p(x_{t+1}|Y^{t+1}) &= \frac{1}{(2\pi)^{\frac{n}{2}}|R|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(y_{t+1}-Hx_{t+1})^{\mathrm{T}}R^{-1}(y_{t+1}-Hx_{t+1})\right] \\
&\quad \cdot\frac{1}{(2\pi)^{\frac{n}{2}}\left|P_{t+1|t}\right|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(x_{t+1}-F\hat{x}_{t|t})^{\mathrm{T}}{P_{t+1|t}}^{-1}(x_{t+1}-F\hat{x}_{t|t})\right] \\
&\quad \cdot\Bigg\{\frac{1}{(2\pi)^{\frac{n}{2}}\left|HP_{t+1|t}H^{\mathrm{T}}+R\right|^{\frac{1}{2}}} \\
&\quad \cdot\exp\Big[-\frac{1}{2}(y_{t+1}-HF\hat{x}_{t|t})^{\mathrm{T}}(HP_{t+1|t}H^{\mathrm{T}}+R)^{-1}(y_{t+1}-HF\hat{x}_{t|t})\Big]\Bigg\}^{-1} \\
&= \frac{\left|HP_{t+1|t}H^{\mathrm{T}}+R\right|^{\frac{1}{2}}}{(2\pi)^{n}|R|^{\frac{1}{2}}\left|P_{t+1|t}\right|^{\frac{1}{2}}}\exp\Big[-\frac{1}{2}\Big\{(y_{t+1}-Hx_{t+1})^{\mathrm{T}}R^{-1}(y_{t+1}-Hx_{t+1}) \\
&\quad +(x_{t+1}-F\hat{x}_{t|t})^{\mathrm{T}}{P_{t+1|t}}^{-1}(x_{t+1}-F\hat{x}_{t|t}) \\
&\quad -(y_{t+1}-HF\hat{x}_{t|t})^{\mathrm{T}}(HP_{t+1|t}H^{\mathrm{T}}+R)^{-1}(y_{t+1}-HF\hat{x}_{t|t})\Big\}\Big] \qquad (2.23)
\end{aligned}
$$

となる．(2.23) 式での，べき乗部分に平方完成を行う．(1.70)-(1.75) 式で，

$$
\begin{aligned}
x &= x_{t+1} \\
y &= y_{t+1} \\
\bar{x} &= F\hat{x}_{t|t} \\
P_0 &= P_{t+1|t} \\
P &= P_{t+1|t+1} \qquad (2.24)
\end{aligned}
$$

と考えると

$$
p(x_{t+1}|Y^{t+1}) = \frac{\left|HP_{t+1|t}H^{\mathrm{T}}+R\right|^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}|R|^{\frac{1}{2}}\left|P_{t+1|t}\right|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(x_{t+1}-\hat{x}_{t+1|t+1})^{\mathrm{T}}{P_{t+1|t+1}}^{-1}(x_{t+1}-\hat{x}_{t+1|t+1})\right]\times K \qquad (2.25)
$$

となる．ただし，

$$
\begin{aligned}
\hat{x}_{t+1|t+1} &= F\hat{x}_{t|t}+P_{t+1|t}H^{\mathrm{T}}R^{-1}(HP_{t+1|t}H^{\mathrm{T}}+R)^{-1}(y_{t+1}-HF\hat{x}_{t|t}) \qquad (2.26) \\
\hat{x}_{t+1|t} &= F\hat{x}_{t|t} \qquad (2.27) \\
P^{-1}_{t+1|t+1} &= P^{-1}_{t+1|t}+H^{\mathrm{T}}R^{-1}H \qquad (2.28)
\end{aligned}
$$

(2.28) 式に，逆行列の補題を適用すると

$$
\begin{aligned}
P_{t+1|t+1} &= P_{t+1|t}-P_{t+1|t}H^{\mathrm{T}}(HP_{t+1|t}H^{\mathrm{T}}+R)^{-1}HP_{t+1|t} \qquad (2.29) \\
P_{t+1|t} &= FP_{t|t}F^{\mathrm{T}}+GQG^{\mathrm{T}} \qquad (2.30)
\end{aligned}
$$

を得る．ここで, x_{t+1} の条件付き期待値 $\hat{x}_{t+1|t+1}$ は (2.26) 式により求まっている．(2.26) ∼ (2.30) 式がカルマンフィルタとなっている．

第3章　スムーザ(Smoother)

3.1　スムージング問題

前章では，観測データに基づいて状態変数 x_t あるいは x_{t+1} の最小分散推定値を逐次的に計算するカルマンフィルタを導出した．本章では Y^t に基づき，現在時刻 t より以前の状態 x_{t-L}， $L=1,2,\cdots$ を推定する平滑化問題 (smoothed problem) について述べる．前章の議論から，Y^t に基づく x_{t-L} の最小分散推定値は, 条件付期待値

$$\hat{x}_{t-L|t} = E\{x_{t-L}|Y_t\} \tag{3.1}$$

で与えられる．この $\hat{x}_{t-L|t}$ を x_{t-L} の**平滑化推定値** (smoothed estimate)，平滑化推定値を計算するアルゴリズムを**スムーザ** (smoother) という．

当然 Y^t の方が Y^{t-L} よりも多くの情報を含んでいるので，x_{t-L} の推定値としては，$\hat{x}_{t-L|t-L}$ より $x_{t-L|t}$ の方が最小分散の意味で優れている．スムージング問題では次の 3 つ型に分類される．

(a) 固定点 (fixed-point) スムージング
(b) 固定遅れ (fixed-lag) スムージング
(c) 固定区間 (fixed-interval) スムージング

(a) の固定点スムージングは，k をある固定された時点とするとき，$\{Y_t,\ \ t=k+1,k+2,\cdots\}$ に基づいて，過去 (k 時点) の状態 x_k の最適推定値を求める問題である．すなわち，時刻 t の経過とともに逐次的に $\hat{x}_{k|t}$ ， $t=k+1,k+2,\cdots$ を求めるアルゴリズムを**固定点**スムーザという．固定点スムーザはシステムの初期状態 x_0 を観測データ $\{Y^t,\ t>0\}$ に基づいて推定する場合などによく用いられる．

(b) の固定遅れ (fixed-lag) スムージングは，与えられた $L\geq 1$ に対して，観測データ Y^t に基づいて，常に L だけ遅れた状態 x_{t-L} の最小分散推定値を求める問題である．逐次的に $\hat{x}_{t-L|t}$ ， $t=L+1,L+2,\cdots$ を計算するアルゴリズムを**固定遅れ**スムーザという．

(c) の固定区間 (fixed-interval) スムージングは，時間区間 $[0,N]$ 観測データ $y_0,y_1,\cdots,y_N$ が与えられたとき，区間 $[0,N]$ のすべての時点 t での状態 x_t の最小分散推定値 $\hat{x}_{t|N}$， $t=0,1,\cdots,N$ を求める問題である．$\hat{x}_{t|N}$， $t=N,N-1,\cdots,0$ を逐次的 (後向き) に計算するアルゴリズムを**固定区間**スムーザという．この場合の計算はオフラインで行われる．

3.2　固定点スムーザ (fixed-point smoother)

以下の線形離散システムについて考察する．

$$x_{t+1} = F_t x_t + G_t w_t \tag{3.2}$$

$$y_t = H_t x_t + v_t \ \ ,\ \ t=0,1,\cdots \tag{3.3}$$

ここでは簡単のため，$S=E\{w_t v_s^T\}=0$ と仮定する．ある特定の時点を k とする．ここでの問題は増大する観測値の集合 Y^t， $t=k+1,k+2,\cdots$ に基づいて，状態 x_k の最小分散推定値 $\hat{x}_{k|t}$ を逐次

的に計算するアルゴリズムを導くことである．以下では，推定誤差を

$$\tilde{x}_{k|t} = x_k - \hat{x}_{k|t}$$

とし，推定誤差共分散行列を

$$P_{k|t} = E\{\tilde{x}_{k|t}\tilde{x}_{k|t}^T\}$$

とする．

ここで n 次元ベクトル θ_t を

$$\theta_{t+1} = \theta_t, \quad t = k, k+1, \cdots ; \quad \theta_k = x_k \tag{3.4}$$

により定義する．このとき，$\theta_t = x_k, \quad t = k, k+1, \cdots$ の関係が成立する．したがって Y^t に基づく θ_t のフィルタ推定値を計算すれば，平滑化推定値 $\hat{x}_{k|t} = \hat{\theta}_{t|t}$ を得る．さらに $2n \times 2n$ 推定誤差共分散行列

$$\begin{bmatrix} P_{t|s} & \Omega_{t|s}^T \\ \Omega_{t|s} & \Sigma_{t|s} \end{bmatrix} = E\left\{ \begin{bmatrix} \tilde{x}_{t|s} \\ \tilde{\theta}_{t|s} \end{bmatrix} \begin{bmatrix} \tilde{x}_{t|s}^T & \tilde{\theta}_{t|s}^T \end{bmatrix} \right\} \tag{3.5}$$

を導入する．ただし，$\tilde{\theta}_{t|s} := \theta_t - \hat{\theta}_{t|s} = x_k - \hat{x}_{k|s} = \tilde{x}_{k|s}$ である．

定理 3.1 *(固定点スムーザ)*　固定点スムージングアルゴリズムは，次の *(i)-(iii)* で与えられる．

(i) 平滑化推定値

$$\hat{x}_{k|t} = \hat{x}_{k|t-1} + A_t(k)[y_t - H_t\hat{x}_{t|t-1}], \quad t = k, k+1, \cdots \tag{3.6}$$

(ii) 平滑化ゲイン行列

$$A_t(k) = \Omega_{t|t-1}H_t^T[H_tP_{t|t-1}H_t^T + R_t]^{-1} \tag{3.7}$$

(iii) 推定誤差共分散行列

$$\Omega_{t+1|t} = \Omega_{t|t-1}[I - K_tH_t]^TF_t^T, \quad t = k, k+1, \cdots \tag{3.8}$$

$$P_{k|t} = P_{k|t-1} - \Omega_{t|t-1}H_t^T[H_tP_{t|t-1}H_t^T + R_t]^{-1}H_t\Omega_{t|t-1}^T \tag{3.9}$$

ただし，境界条件は $\Omega_{k|k-1} = P_{k|k-1}$ であり，$\hat{x}_{t|t-1}, P_{t|t-1}, K_t$ はカルマンフィルタにおける推定値，推定誤差共分散およびカルマンゲインである．

証明 次の拡大システムを得る．

$$\begin{bmatrix} x_{t+1} \\ \theta_{t+1} \end{bmatrix} = \begin{bmatrix} F_t & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} x_t \\ \theta_t \end{bmatrix} + \begin{bmatrix} G_t \\ 0 \end{bmatrix} w_t \tag{3.10}$$

$$y_t = \begin{bmatrix} H_t & 0 \end{bmatrix}\begin{bmatrix} x_t \\ \theta_t \end{bmatrix} + v_t \tag{3.11}$$

上式 (3.10), (3.11) にカルマンフィルタの理論を適用すると，(3.6)-(3.9) が得られる．すなわち $t = k, k+1, \cdots$ に対して次式を得る．

$$\begin{bmatrix} \hat{x}_{t+1|t} \\ \hat{\theta}_{t+1|t} \end{bmatrix} = \begin{bmatrix} F_t & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} \hat{x}_{t|t} \\ \hat{\theta}_{t|t} \end{bmatrix} \tag{3.12}$$

$$\begin{bmatrix} \hat{x}_{t|t} \\ \hat{\theta}_{t|t} \end{bmatrix} = \begin{bmatrix} \hat{x}_{t|t-1} \\ \hat{\theta}_{t|t-1} \end{bmatrix} + \begin{bmatrix} K_t \\ \triangle_t \end{bmatrix}\begin{bmatrix} y_t - H_t\hat{x}_{t|t-1} \end{bmatrix} \tag{3.13}$$

ただし、境界条件は $\hat{\theta}_{k|k-1} = \hat{x}_{k|k-1}$ である．また，ゲインは

$$\begin{bmatrix} K_t \\ \triangle_t \end{bmatrix} = \begin{bmatrix} P_{t|t-1} & \Omega_{t|t-1}^T \\ \Omega_{t|t-1} & \Sigma_{t|t-1} \end{bmatrix} \begin{bmatrix} H_t^T \\ 0 \end{bmatrix} \begin{bmatrix} H_t P_{t|t-1} H_t^T + R_t \end{bmatrix}^{-1}$$

となる．ここで，$\triangle_t = A_k(t)$ とおくと，(3.7) 式を得る．また $\hat{\theta}_{t|t} = \hat{x}_{k|t}, \hat{\theta}_{t|t-1} = \hat{x}_{k|t-1}$ であるから，式 (3.13) から式 (3.6) が導かれる．次に拡大システムに対するリカッチ方程式は，以下で与えられる．

$$\begin{bmatrix} P_{t+1|t} & \Omega_{t+1|t}^T \\ \Omega_{t+1|t} & \Sigma_{t+1|t} \end{bmatrix} = \begin{bmatrix} F_t & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} P_{t|t} & \Omega_{t|t}^T \\ \Omega_{t|t} & \Sigma_{t|t} \end{bmatrix} \begin{bmatrix} F_t^T & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} G_t Q_t G_t^T & 0 \\ 0 & 0 \end{bmatrix} \tag{3.14}$$

$$\begin{bmatrix} P_{t|t} & \Omega_{t|t}^T \\ \Omega_{t|t} & \Sigma_{t|t} \end{bmatrix} = \begin{bmatrix} I - K_t H_t & 0 \\ -\triangle_t H_t & I \end{bmatrix} \begin{bmatrix} P_{t|t-1} & \Omega_{t|t-1}^T \\ \Omega_{t|t-1} & \Sigma_{t|t-1} \end{bmatrix} \tag{3.15}$$

となる．ただし $t = k$ においては，$\Omega_{k|k-1} = P_{k|k-1} = \Sigma_{k|k-1}$ が成立する．式 (3.14)，(3.15) を成分に分解すると，$\Omega_{t+1|t}$ は式 (3.8) を満足する．さらに $\Sigma_{t+1|t} = P_{k|t}, \Sigma_{t|t-1} = P_{k|t-1}$ に注意すると，式 (3.9) を得る．本定理のアルゴリズムはカルマンフィルタの推定値 $\hat{x}_{t|t-1}$，共分散行列 $P_{t|t-1}$，カルマンゲイン K_t を必要とするので，固定点スムーザは，カルマンフィルタと共に解き，連動させる．

Remark 1

$$I - K_t H_t = P_{t|t}(P_{t|t-1})^{-1}$$

の関係を用いると，

$$\Omega_{t+1|t} = \Omega_{t|t-1}(P_{t|t-1})^{-1} P_{t|t} F_t^T, \quad \Omega_{k|k-1} = P_{k|k-1} \tag{3.16}$$

の関係式を示せる．

Remark 2

上記 **Remark 1** と $A_t(k) = \Omega_{t|t-1}(P_{t|t-1})^{-1} K_t, K_t[y_t - H_t \hat{x}_{t|t-1}] = K_t \nu_t = \hat{x}_{t|t} - \hat{x}_{t|t-1}$ の関係を利用すると，式 (3.8) は次式のように表せる．

$$\hat{x}_{k|t} = \hat{x}_{k|t-1} + B_t(k)[\hat{x}_{t|t} - \hat{x}_{t|t-1}], \quad t = k+1, \cdots \tag{3.17}$$

$$B_t(k) = M_k M_{k+1} \cdots M_{t-1}, \quad M_i = P_{i|i} F_i^T [P_{i+1|i}]^{-1} \tag{3.18}$$

Remark 3

$\tilde{F}_t = F_t(I - K_t H_t)$ および

$$\Psi(t, \tau) = \begin{cases} \tilde{F}_{t-1} \cdots \tilde{F}_\tau, & t > \tau \\ I, & t = \tau \end{cases} \tag{3.19}$$

とおくと，$\Omega_{i|i-1} = P_{k|k-1} \Psi^T(i, k)$ の関係より，平滑化推定値 (3.6) 式は

$$\begin{aligned} \hat{x}_{k|t} &= \hat{x}_{k|k+1} + \sum_{i=k}^{t} A_i(k) \nu_i \\ &= \hat{x}_{k|k-1} + P_{k|k-1} \sum_{i=k}^{t} \Psi^T(i, k) H_i^T [H_i P_{i|i-1} H_i^T + R_i]^{-1} \nu_i \end{aligned} \tag{3.20}$$

と表現できる．

次に，スムージングによる推定精度の改善について考察する．式 (3.9) を繰り返し用いることにより，

$$P_{k|k-1} - P_{k|t} = \sum_{i=k}^{t} \Omega_{i|i-1} H_i^T [H_i P_{i|i-1} H_i^T + R_i]^{-1} H_i \Omega_{i|i-1}^T \tag{3.21}$$

を得る．上式右辺は半正定値であるので，$P_{k|k-1} - P_{k|t}$ は t について，単調非減少である．ここで，議論の簡単化のため，システムは定常であると仮定し，$F_t = F, G_t = G, H_t = H, Q_t = I, R_t = R$ と時不変行列，および $P_{t|t-1} = P = P_{k|k-1}, K_t = K$ とする．さらに (F, G) は可安定，(H, F) は可検出であるとする．このとき $\tilde{F}_t = \tilde{F}$ を用いると，$\Omega_{i|i-1} = P_{k|k-1}\Psi^T(i, k) = P(\tilde{F}^T)^{i-k}$ となるので，式 (3.21) は

$$P - P_{k|t} = P\left(\sum_{i=k}^{t}(\tilde{F}^T)^{i-k}H^T[HPH^T + R]^{-1}H\tilde{F}^{i-k}\right)P \tag{3.22}$$

となる．(F, G) が可安定，(H, F) が可検出であれば，ARE(Algebraic Riccati Equation) の性質より，$\tilde{F} = F(I - KH)$ は漸近安定である．したがって，$t \to \infty$ とすると，

$$P - P_{k|\infty} = P\left(\sum_{i=0}^{\infty}(\tilde{F}^T)^{i}H^T[HPH^T + R]^{-1}H\tilde{F}^{i}\right)P \tag{3.23}$$

を得る．この等式は，スムージングによる推定精度の改善の上限を与えている．

3.3 固定遅れスムーザ

L を固定ラグとするとき，固定遅れ平滑化推定値 $\hat{x}_{t-L|t}$ および推定誤差共分散行列 $P_{t-L|t}$ に対する逐次アルゴリズムを以下のように導出する．まず x_t，x_{t-1}，$\cdots$，x_{t-L} を縦に並べた拡大状態ベクトルを

$$z_t = \begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-L} \end{bmatrix} \quad [n \times (L+1)\ \text{次元ベクトル}] \tag{3.24}$$

とおき，以下の拡大システムを考える．

$$z_{t+1} = \begin{bmatrix} F_t & 0 & \cdots & 0 \\ I & & & \\ & \ddots & & \\ 0 & & I & 0 \end{bmatrix} z_t + \begin{bmatrix} G_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} w_t \tag{3.25}$$

$$y_t = \begin{bmatrix} H_t & 0 & \cdots & 0 \end{bmatrix} z_t + v_t \tag{3.26}$$

このとき，z_t の条件付き期待値は

$$\hat{z}_{t|t} = E\{z_t|Y_t\} = \begin{bmatrix} \hat{x}_{t|t} \\ \hat{x}_{t-1|t} \\ \vdots \\ \hat{x}_{t-L|t} \end{bmatrix}$$

となるので，拡大システムの状態ベクトルのフィルタ推定値は遅れ (ラグ) $1 \sim L$ のすべての平滑化推定値を与える．したがって，(3.25), (3.26) 式の拡大システムにカルマンフィルタ理論を適用すれば，固定遅れスムーザを得る．

ここで，推定誤差共分散を

$$P_{t-j,t-l|t} = E\{[x_{t-j} - \hat{x}_{t-j|t}][x_{t-l} - \hat{x}_{t-l|t}]^T\} \tag{3.27}$$

と表す．明らかに，$P_{t-j,t-j|t} = P_{t-j|t}$ が成り立つ．

定理 3.2 *(固定遅れスムーザ)*
(i) 平滑化推定値

$$\hat{x}_{t-j|t} = \hat{x}_{t-j|t-1} + K_t(j)[y_t - H_t\hat{x}_{t|t-1}], \quad (j = 0, 1, \cdots, L) \tag{3.28}$$

(ii) 予測推定値

$$\hat{x}_{t+1|t} = F_t\hat{x}_{t|t} \tag{3.29}$$

(iii) 平滑化ゲイン

$$K_t(j) = P_{t-j,t|t-1}H_t^T[H_tP_{t|t-1}H_t^T + R_t]^{-1} \quad (j = 0, 1, \cdots, L) \tag{3.30}$$

(iv) 推定誤差共分散

$$P_{t+1,t+1|t} = F_tP_{t,t|t}F_t^T + G_tQ_tG_t^T, \tag{3.31}$$
$$P_{t+1,t-j|t} = F_tP_{t,t-j|t}; \quad P_{t-j,t+1|t} = (P_{t+1,t-j|t})^T, \tag{3.32}$$
$$P_{t-j,t-l|t} = P_{t-j,t-l|t-1} - K_t(j)H_tP_{t,t-l|t-1}, \quad (j, l = 0, 1, \cdots, L) \tag{3.33}$$

(v) 初期条件

$$\hat{x}_{0|-1} = \bar{x}_0, \quad \hat{x}_{-j|-1} = 0 \quad (j = 1, \cdots, L) \tag{3.34}$$
$$P_{0,0|-1} = P_0, \quad P_{-j,-l|-1} = 0 \quad (j, l = 0, 1, \cdots, L;\ j^2 + l^2 \neq 0) \tag{3.35}$$

Remark 4
$(P_{t-j,t-l|t})^T = P_{t-l,t-j|t}$ および $P_{t-j,t-j|t}$ の対称性を利用すると，式 (3.28)-(3.33) のアルゴリズムは簡単化できる．また初期値 $\hat{x}_{-j|-1} = 0$ $(j = 1,\ \cdots, L)$ としているが，$x_{-j|-1},\ j = 1,\ \cdots, L$ の平均値および共分散行列が既知であれば，それらを初期条件として用いてもよい．

次の関係を示すことができる．

$$K_t(j) = P_{t-j,t|t-1}(P_{t|t-1})^{-1}K_t \tag{3.36}$$
$$P_{t-j,t+1|t} = P_{t-j,t|t-1}\tilde{F}^T \tag{3.37}$$

3.4 固定区間スムーザ (fixed-lag smoother)

区間 $[0, N]$ において与えられた観測データを $Y^N = [y_0, \cdots, y_N]$ とする．Y^N に基づいて，状態 x_t を推定する固定区間スムーザについて考察する．固定区間スムーザのアルゴリズムは，次の 2 つのステップに分解できる．
(1) カルマンフィルタにより，推定値 $\hat{x}_{t|t-1}$，$\hat{x}_{t|t}$ および $P_{t|t-1}, P_{t|t}$，$t = 0,\ 1,\ \cdots, N$ を計算し，$\hat{x}_{t|t-1}$，$\hat{x}_{t|t}$(あるいはイノベーション $\nu_t = y_t - H_t\hat{x}_{t|t-1}$ および $P_{t|t-1}$，$P_{t,t}$，$t = 0,\ 1,\ \cdots,\ N$ を記憶する．
(2) 境界値 $\hat{x}_{N|N}$ から出発して逆向き $\hat{x}_{N-1|N}$，$\cdots$，$\hat{x}_{0|N}$ 計算する．

以下，固定点スムージングに対するアルゴリズムを利用して，固定区間スムーザを導出し，Fraser のアルゴリズムについて述べる．

固定区間スムーザのアルゴリズムは以下の (i)-(iii) で与えられる．

定理 3.3
(i) 平滑化推定値

$$\hat{x}_{t|N} = \hat{x}_{t|t} + C_t[\hat{x}_{t+1|N} - \hat{x}_{t+1|t}], \quad t = N-1,\ \cdots,\ 0 \tag{3.38}$$

(ii) 平滑化ゲイン

$$C_t = P_{t|t} F_t^T (P_{t+1|t})^{-1} \tag{3.39}$$

(iii) 平滑化推定誤差共分散行列

$$P_{t|N} = P_{t|t} + C_t [P_{t+1|N} - P_{t+1|t}] C_t^T, \quad t = N-1, \cdots, 0 \tag{3.40}$$

ただし式 (3.38), (3.40) で $t = N$ における境界条件はフィルタ推定値 $\hat{x}_{N|N}$ および推定誤差共分散行列 $P_{N|N}$ で与えられる．

証明 式 (3.20) において，$t = N,\ k = t+1$ とおくと

$$\begin{aligned} \hat{x}_{t+1|N} &= \hat{x}_{t+1|t} + P_{t+1|t} \sum_{i=t+1}^{N} \Psi^T(i, t+1) \tilde{\nu}_i & (3.41) \\ &= \hat{x}_{t+1|t+1} + P_{t+1|t} \sum_{i=t+2}^{N} \Psi^T(i, t+1) \tilde{\nu}_i & (3.42) \end{aligned}$$

を得る．ただし $\tilde{\nu}_i = H_t^T [H_i P_{i|i-1} H_i^T + R_i]^{-1} \nu_i$ とおいた．上式で $t := t-1$ とおくと

$$\hat{x}_{t|N} = \hat{x}_{t|t} + P_{t|t-1} \sum_{i=t+1}^{N} \Psi^T(i, t) \tilde{\nu}_i \tag{3.43}$$

を得る．$\Psi(i,t) = \tilde{F}_{i-1} \cdots \tilde{F}_{t+1} \tilde{F}_t = \Psi(i, t+1) \tilde{F}_t$ が成立するので，式 (3.41), (3.42) を用いると

$$\begin{aligned} \hat{x}_{t|N} &= \hat{x}_{t|t} + P_{t|t-1} \tilde{F}_t^T \sum_{i=t+1}^{N} \Psi^T(i, t+1) \tilde{\nu}_i \\ &= \hat{x}_{t|t} + P_{t|t-1} \tilde{F}_t^T (P_{t+1|t})^{-1} [\hat{x}_{t+1|N} - \hat{x}_{t+1|t}] & (3.44) \end{aligned}$$

ここで $P_{t|t-1} \tilde{F}_t^T = P_{t|t} F_t^T$ が成立すること，および式 (**??**) の C_t の定義に注意すると，式 (3.44) から式 (3.38) が示される．

次に式 (3.40) を示そう．式 (3.38) の両辺に $-x_t$ を加えて変形すると

$$\tilde{x}_{t|N} + C_t \hat{x}_{t+1|N} = \tilde{x}_{t|t} + C_t \hat{x}_{t+1|t} \tag{3.45}$$

となる．$\tilde{x}_{t|N} = x_t - \hat{x}_{t|N}$ はデータ空間 Y^N に直交し，$\hat{x}_{t+1|N}$ は Y^N に属するので $\tilde{x}_{t|N} \perp \hat{x}_{t+1|N}$ となる．したがって $E\{\tilde{x}_{t|N} (\hat{x}_{t+1|N})^T\} = 0$ が成立する．同じく $\tilde{x}_{t|t} \perp \hat{x}_{t+1|t}$ であるから $E\{\tilde{x}_{t|t} (\hat{x}_{t+1|t})^T\} = 0$ を得る．したがって式 (3.45) の両辺の共分散行列を計算すると

$$P_{t|N} + C_t \hat{P}_{t+1|N} C_t^T = P_{t|t} + C_t \hat{P}_{t+1|t} C_t^T \tag{3.46}$$

となる．ここに $\hat{P}_{t|k} = E\{\hat{x}_{t|k} \hat{x}_{t|k}^T\}$ である．また

$$\begin{aligned} x_{t+1} &= \hat{x}_{t+1|t} + \tilde{x}_{t+1|t}, \quad \hat{x}_{t+1|t} \perp \tilde{x}_{t+1|t} \\ &= \hat{x}_{t+1|N} + \tilde{x}_{t+1|N}, \quad \hat{x}_{t+1|N} \perp \tilde{x}_{t+1|N} & (3.47) \end{aligned}$$

が成立するので

$$E\{x_{t+1} x_{t+1}^T\} = P_{t+1|t} + \hat{P}_{t+1|t} = P_{t+1|N} + \hat{P}_{t+1|N}$$

を得る．したがって $\hat{P}_{t+1|t} - \hat{P}_{t+1|N} = P_{t+1|N} - P_{t+1|t}$ となり，この関係を式 (3.46) に用いれば式 (3.40) を得る．

Remark 5

$G_t=0$(あるいは $Q_t=0$) の場合，F_t^{-1} の存在を仮定すると，固定区間スムーザは

$$\hat{x}_{t|N}=F_t^{-1}\hat{x}_{t+1|N},\quad P_{t|N}=F_t^{-1}P_{t+1|N}F_t^{-T}$$

となることを示せる．これは後ろ向きの予測フィルタである．したがって $F_t=I$ とすると $\hat{x}_{t|N}=\hat{x}_{N|N}$, $t=N-1,\ \cdots,\ 0$ となる．このことから定数の平滑化推定値はフィルタ推定値と同じであることが分かる．この証明には，$G_t=0$ とすると，$P_{t+1|t}=F_tP_{t|t}F_t^T$ となるので，$C_t=F_t^{-1}$ を得る．また $\hat{x}_{t+1|t}=F_t\hat{x}_{t|t}$ であることを利用する．

以上の固定区間スムーザのアルゴリズムの欠点は平滑ゲイン C_t が $P_{t+1|t}$ の逆行列を含むために，式 (3.38), (3.40) において $(P_{t+1|t})^{-1}[\hat{x}_{t+1|N}-\hat{x}_{t+1|t}]$ および $(P_{t+1|t})^{-1}[P_{t+1|N}-P_{t+1|t}](P_{t+1|t})^{-1}$ という項が現れることである．各ステップで $P_{t+1|t}^{-1}$ を計算するのは不便であり，また $P_{t+1|t}^{-1}$ が大きな値をとり，$\hat{x}_{t+1|N}-\hat{x}_{t+1|t}$ あるいは $P_{t+1|N}-P_{t+1|t}$ が小さな値をとる場合には，計算機内部での桁落ちが生じ，数値的に不安定になる可能性がある．次にこの欠点を避ける Fraser の固定区間スムーザについて述べる．

n 次元ベクトル λ_{t+1} を

$$\lambda_{t+1}=(P_{t+1|t})^{-1}[\hat{x}_{t+1|N}-\hat{x}_{t+1|t}] \tag{3.48}$$

とおくと，式 (3.38) は

$$\hat{x}_{t|N}=\hat{x}_{t|t}+P_{t|t}F_t^T\lambda_{t+1} \tag{3.49}$$

と表される．式 (3.48) で $t:=t-1$ とおき，式 (3.49) を用いると

$$\begin{aligned}\lambda_t &= (P_{t|t-1})^{-1}[\hat{x}_{t|N}-\hat{x}_{t|t-1}] &(3.50)\\ &= (P_{t|t-1})^{-1}[\hat{x}_{t|t}+P_{t|t}F_t^T\lambda_{t+1}-\hat{x}_{t|t-1}] &(3.51)\end{aligned}$$

を得る．ここで $\hat{x}_{t|t}-\hat{x}_{t|t-1}=K_t\nu_t, P_{t|t-1}^{-1}P_{t|t}=I-H_t^TK_t^T$ が成立するので，式 (3.51) は

$$\lambda_t=\tilde{F}_t^T\lambda_{t+1}+H_t^T[H_tP_{t|t-1}H_t^T+R_t]^{-1}\nu_t \tag{3.52}$$

となる．ここで，ν_t はイノベーションである．以上をまとめると，次の結果を得る．

定理 3.4 Fraser のアルゴリズム
固定区間スムーザは

$$\begin{aligned}\hat{x}_{t|N} &= \hat{x}_{t|t}+P_{t|t}F_t^T\lambda_{t+1} &(3.53)\\ \lambda_t &= \tilde{F}^T\lambda_{t+1}+H_t^T[H_tP_{t|t-1}H_t^T+R_t]^{-1}\nu_t &(3.54)\end{aligned}$$

で与えられる．ただし，$\lambda_{N+1}=0$ である．また推定誤差共分散行列は次式で表される．

$$P_{t|N}=(I-C_tF_t)P_{t|t}(I-C_tF_t)^T+C_t(P_{t+1|N}+G_tQ_tG_t^T)C_t^T \tag{3.55}$$

以上の関係を証明する．式 (3.48) で $t=N$ とおくと $\lambda_{N+1}=0$ となるので，式 (3.53), (3.54) は明らかである．式 (3.55) を示そう．式 (3.39), (3.40) から

$$\begin{aligned}P_{t|N} &= P_{t|t}-C_tP_{t+1|t}C_t^T+C_tP_{t+1|N}C_t^T\\ &= P_{t|t}-C_tF_tP_{t|t}+C_tP_{t+1|N}C_t^T\\ &= P_{t|t}-C_tF_tP_{t|t}-P_{t|t}F_t^TC_t^T+C_tP_{t+1|t}C_t^T+C_tP_{t+1|N}C_t^T\end{aligned}$$

を得る．ここでカルマンフィルタの関係式を用いて $P_{t+1|t}$ を消去し，上式を因数分解すれば式 (3.55) を得る．

式 (3.53), (3.54) は確かに逆行列演算としては $[H_tP_{t|t-1}H_t^T+R_t]^{-1}$ しか含まないので，数値的に安定なアルゴリズムといえる．しかし式 (3.55) は式 (3.39) の C_t を含むので，やはり逆行列 $(P_{t+1|t})^{-1}$ が必要である．

関連図書

[1] A. Papoulis: *Probability, Random Variables and Stochastic Processes, 2nd Edition*, McGraw Hill, New York (1984).

[2] H. L. Van Trees: *Detection, Estimation, and Modulation Theory, Part 1. Detection, Estimation, and Linear modulation Theory*, Wiley Inter-Science, New York (2001).

[3] 杉山 博: 新編　確率統計要論, 養賢堂 (1980).

[4] 片山 徹: 新版 応用カルマンフィルタ, 朝倉書店 (2000).

[5] B. D. O. Anderson and J. B. Moore: *Optimal Filtering*, McGraw Hill, New York (1965).

[6] M. R. Spiegel, J. Schiller and R. Alu Srinivasan: *Probability and Statistics, 2nd Edition*, McGraw-Hill, New York (2000).