

信号処理

~ 第 2 部 確率的な信号の処理 ~

横田 康成

平成 14 年 11 月 27 日

目次

第 1 章	確率の基礎	2
1.1	1 つの確率変数の性質	2
1.2	2 つの確率変数の間の性質	5
1.3	複数の確率変数の間の性質	8
1.4	正規確率変数	9
第 2 章	推定の基礎	12
2.1	推定とその性質	12
2.2	最尤推定	19
2.3	推定量の分散の下限	20
2.4	EM アルゴリズム	23
第 3 章	確率過程	28
3.1	確率過程	28
3.2	定常過程と正規確率過程	29
3.3	エルゴード過程	30
第 4 章	確率過程のパワースペクトル密度の推定	33
4.1	自己相関関数の推定	33
4.2	ウィナー - ヒンチンの定理	33
4.3	ブラックマン・チューキー法によるパワースペクトル密度の推定と窓関数	35
4.4	ペリオドグラム法によるパワースペクトル密度の推定	37
第 5 章	線形予測	43
5.1	線形予測モデル	43
5.2	予測誤差の統計的性質	46
5.3	AR 係数の推定値の統計的性質	49
5.4	線形予測モデルを用いたパワースペクトル密度の推定	52
5.5	最適次数決定基準	55
第 6 章	AR モデルの係数推定法	58
6.1	Levinson-Durbin アルゴリズム	58
6.2	ラティスフィルタによる AR モデルの構成	61
6.3	Burg 法	62

第1章 確率の基礎

「信号処理」では、信号とシステムの表現方法を学んだ。信号は、複素指数関数の和、すなわちフーリエスペクトルで表現でき、線形時不変システムは、インパルスを入力した際の応答、つまりインパルス応答のラプラス変換、あるいは z 変換として与えられる伝達関数で表現できる。しかし、これまでの議論では、信号には雑音が含まれていないとする暗黙の仮定が含まれていた。音声のフーリエ変換を行う場合に気がついたと思うが、実際に観測、計測される信号には、熱雑音、背景雑音、観測に際して含まれる観測雑音などが含まれている。したがって、求められた信号のフーリエスペクトル、システムのインパルス応答、伝達関数、周波数特性には、統計的な変動が含まれおり、これをいかに抑え、本来の信号、システムの性質、特徴を知ることができるかという問題が発生する。そのためには、信号自体を確率的な信号として扱わなければならない。本章ではまず確率の基礎を復習しよう。

1.1 1つの確率変数の性質

ある1回の観測により、1つの観測値 x が得られるものとする。この観測値 x は、観測にともなう観測雑音などの影響により観測毎に異なる値をとり得る。各回の観測で得られたそれらの値を標本 (sample) という。確率的に値が変化する標本を発生させる変数、つまり確率変数 X を考えれば、標本は確率変数 X の実現値とみなせる。確率的に値が変化する確率変数の性質を表現するためには、その変数がとり得る値の確率を記述すればよい。そこで、確率変数 X が $X < x$ である確率を $P_X(x)$ と書くことにする。これは確率分布 (probability distribution) と呼ばれている。確率分布 $P_X(x)$ が x に関し微分可能であるものとして、 $p_X(x) \equiv \frac{d}{dx}P_X(x)$ とおく。 $p_X(x)$ は、 $\int_{x_1}^{x_2} p_X(x)dx$ が $x_1 < X < x_2$ である確率を表すことから、 X の確率密度分布 (probability density distribution) と呼ばれている。 X の実現値が正確に x に一致する確率は0であり、 $p_X(x)$ ではないことに注意しよう。また、確率変数 X は、確率密度分布 $p_X(x)$ に従うともいわれる。

確率変数 X は、その確率密度分布 $p_X(x)$ により、その性質が完全に規定される。しかし、確率密度分布そのものを表現できない、あるいは表現する必要がない場合、確率密度分布形状やその特徴を表すいくつかの統計量で確率変数の性質を表現する場合がある。よく知られた統計量としては、分布の平均 μ や分散 σ^2 があり、それぞれ、次式で与えられる。

$$\mu = \int_{-\infty}^{\infty} xp_X(x)dx \quad (1.1)$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 p_X(x) dx \quad (1.2)$$

$p_X(x)$ を剛体の質量分布とみなすならば、平均 μ 、分散 σ^2 は、それぞれ剛体の重心、重心の周りの回転モーメントを表す。平均、分散は、このように直感的な理解が得られやすいため、確率変数、あるいはその確率密度分布の記述に頻繁に利用される。また、分散の平方根である標準偏差 σ は、その単位が x の単位に一致するため、分散の代わりに用いられることが多い。

式 (1.1),(1.2) において、積分の内部にある $x, (x - \mu)^2$ などを x の関数と見て、一般的に $g(x)$ と書けば、

確率密度分布 $p_X(x)$ を特徴づける統計量を

$$E[g(X)] \equiv \int_{-\infty}^{\infty} g(x)p_X(x)dx \quad (1.3)$$

と表現することができる。これは、 $g(X)$ の期待値 (expectation, expected value) と呼ばれ、確率密度分布 $p_X(x)$ の統計量を一般的に表現したものである。この式と第 2 章で学んだフーリエ変換

$$X(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t}x(t)dt$$

を比較してみよう。フーリエ変換は、信号 $x(t)$ に含まれる複素指数関数 $e^{-i\omega t}$ の成分を表すものであるから、 t と x 、 $x(t)$ と $p_X(x)$ 、 $e^{-i\omega t}$ と $g(x)$ をそれぞれ対応させれば、 $g(X)$ の期待値は $p_X(x)$ の中に含まれる $g(x)$ の成分を表すことになる。

$g(x) = x$ の場合、すなわち X の期待値は平均 μ である。一般に、 $g(x) = x^n$ の場合、つまり X^n の期待値

$$m_n \equiv E[X^n] = \int_{-\infty}^{\infty} x^n p_X(x) dx, \quad n = 1, 2, \dots$$

は、 X の n 次モーメント (n th order moment) と呼ばれる。ただし、 $m_0 = 1, m_1 = \mu$ である。

また、 $g(x) = (x - \mu)^2$ の場合、すなわち、 $(X - \mu)^2$ の期待値は、 X の分散である。一般に、 $g(x) = (x - \mu)^n$ の場合、すなわち $(X - \mu)^n$ の期待値

$$C_m \equiv E[(X - \mu)^n] = \int_{-\infty}^{\infty} (x - \mu)^n p_X(x) dx$$

は、 X の n 次中心モーメント (n th order central moment) と呼ばれる。2 次中心モーメントは、

$$C_2 = E[X^2 - 2X\mu + \mu^2] = E[X^2] - 2E[X]\mu + E[X]^2 = m_2 - m_1^2$$

と書くことができる。同様に、3 次中心モーメントは、

$$C_3 = m_3 - 3m_2m_1 + 2m_1^3$$

と書くことができ、また、分布の非対称性を表すことから、偏り度 (skewness) と呼ばれる。さらに、4 次中心モーメントは、

$$C_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4$$

と書くことができ、また、分布の偏平さを表すことから、偏平度 (flatness) と呼ばれる。 n 次モーメントや n 次中心モーメントを n 次統計量という。

次に、 $g(x) = e^{sx}$ の場合、すなわち $\varphi(s) \equiv E[e^{sX}]$ の性質を考えよう。 e^{sx} を s に関してマクローリン展開すると、

$$e^{sx} = 1 + sx + \frac{1}{2!}(sx)^2 + \frac{1}{3!}(sx)^3 + \dots$$

となるので、

$$\begin{aligned} \varphi(s) &= E\left[1 + sX + \frac{1}{2!}(sX)^2 + \frac{1}{3!}(sX)^3 + \dots\right] \\ &= 1 + sE[X] + \frac{1}{2!}s^2E[X^2] + \frac{1}{3!}s^3E[X^3] + \dots \\ &= \sum_{k=0}^{\infty} \frac{m_k}{k!} s^k \end{aligned} \quad (1.4)$$

となる．上式の両辺を s で n 階微分して， $s = 0$ とおくと

$$\left. \frac{\partial^n}{\partial s^n} \varphi(s) \right|_{s=0} = m_n, \quad n = 0, 1, 2, \dots$$

となり， X の n 次モーメントを表すことになる．こうしたことから， $\varphi(s) = E[e^{sX}]$ は，モーメント母関数 (moment generating function) と呼ばれている． $g(x) = e^{sx}$ の代わりに， $g(x) = e^{isx}$ において，すなわち $E[e^{isX}]$ としてもモーメントを算出することが可能である． $E[e^{isX}]$ は特性関数 (characteristic function) と呼ばれ，確率密度分布 $p_X(x)$ と特性関数 $E[e^{isX}]$ は，互いにフーリエ変換，逆フーリエ変換の関係にある．

モーメント母関数の対数 $\log \varphi(s)$ は，キュムラント母関数 (cumulant generating function) と呼ばれている．キュムラント母関数 $\log \varphi(s)$ を s のべきで展開し，次式で表現するものとする．

$$\log \varphi(s) = \sum_{k=0}^{\infty} \frac{c_k}{k!} s^k \quad (1.5)$$

ここで， c_k ， $k = 0, 1, 2, \dots$ は， k 次キュムラント (cumulant) と呼ばれ，キュムラント母関数を n 階微分し， $s = 0$ とおいたものとして与えられる．

$$c_n \equiv \left. \frac{\partial^n}{\partial s^n} \log \varphi(s) \right|_{s=0}, \quad n = 0, 1, 2, \dots$$

式 (1.5) の両辺を s で偏微分すると

$$\frac{1}{\varphi(s)} \frac{\partial}{\partial s} \varphi(s) = \sum_{k=1}^{\infty} \frac{c_k}{(k-1)!} s^{k-1} \quad (1.6)$$

となり，また，式 (1.4) の両辺を s で偏微分すると

$$\frac{\partial}{\partial s} \varphi(s) = \sum_{k=1}^{\infty} \frac{m_k}{(k-1)!} s^{k-1} \quad (1.7)$$

となる．式 (1.4) と式 (1.7) を式 (1.6) に代入すると，

$$\sum_{k=1}^{\infty} \frac{m_k}{(k-1)!} s^{k-1} = \left(\sum_{k=1}^{\infty} \frac{c_k}{(k-1)!} s^{k-1} \right) \left(\sum_{k=0}^{\infty} \frac{m_k}{k!} s^k \right)$$

となる．この式の s の各べき項の係数を両辺で比較することにより，

$$m_n = c_1 m_{n-1} + \binom{n-1}{1} c_2 m_{n-2} + \dots + \binom{n-1}{n-2} c_{n-1} m_1 + c_n, \quad n = 1, 2, \dots$$

なる関係が得られる．この式より，

$$\mathbf{m} \equiv \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ \vdots \end{pmatrix}, \quad \mathbf{c} \equiv \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ \vdots \end{pmatrix}, \quad \mathbf{A} \equiv \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ m_1 & 1 & 0 & 0 & \dots \\ m_2 & 2m_1 & 1 & 0 & \dots \\ m_3 & 3m_2 & 3m_1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

とおけば，

$$\mathbf{m} = \mathbf{A} \mathbf{c}$$

が成り立つ．したがって，キュムラントは，

$$c = A^{-1}m$$

として求められる．また，

$$B \equiv \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ -c_1 & 1 & 0 & 0 & \cdots \\ -2c_2 & -c_1 & 1 & 0 & \cdots \\ -3c_3 & -3c_2 & -c_1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

とおけば，

$$c = Bm$$

が成り立つ．したがって，モーメントは，

$$m = B^{-1}c$$

として求められる．キュムラントとモーメントの関係は低次については次式となる．

$$\begin{aligned} c_1 &= m_1 \\ c_2 &= m_2 - m_1^2 \\ c_3 &= m_3 - 3m_2m_1 + 2m_1^3 \\ c_4 &= m_4 - 4m_3m_1 - 3m_2^2 + 12m_2m_1^2 - 6m_1^4 \\ &\vdots \end{aligned}$$

$$\begin{aligned} m_1 &= c_1 \\ m_2 &= c_2 + c_1^2 \\ m_3 &= c_3 + 3c_1c_2 + c_1^3 \\ m_4 &= c_4 - 3c_2^2 + 4c_1c_3 + 6c_1^2c_2 + c_1^4 \\ &\vdots \end{aligned}$$

これらの関係式より，1次キュムラントは平均 μ ，2次，3次キュムラントは，それぞれ2次，3次中心モーメントに一致する．

1.2 2つの確率変数の間の性質

観測される量が2個 (x, y) 存在する場合のそれらの間の関係を考えよう．それぞれの変数を確率変数 X, Y で表現し， $X < x$ かつ $Y < y$ である確率を $P_{X,Y}(x, y)$ と書くことにする． $P_{X,Y}(x, y)$ は，2変数の場合の確率分布であり，結合確率分布 (joint probability distribution) と呼ばれる． $P_{X,Y}(x, y)$ が x, y に関して2階微分可能であるとすれば， $p_{X,Y}(x, y) \equiv \frac{d}{dx} \frac{d}{dy} P_{X,Y}(x, y)$ は，2変数の場合の確率密度分布を表し，結合確率密度分布 (joint probability density distribution) と呼ばれる． X の確率密度分布 $p_X(x)$ ， Y の確率密度分布 $p_Y(y)$ は，それぞれ残った変数の実現値を問わないことから，結合確率密度分布 $p_{X,Y}(x, y)$ より，それぞれ

$$p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x, y) dy \quad (1.8)$$

$$p_Y(y) = \int_{-\infty}^{\infty} p_{X,Y}(x, y) dx \quad (1.9)$$

として与えられる．これらは，結合確率密度分布 $p_{X,Y}(x,y)$ の周辺分布 (marginal distribution) と呼ばれる．結合確率密度分布が，それぞれの確率変数の周辺分布の積

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

で与えられるならば，確率変数 X, Y は独立 (independent) であるという．また， Y の実現値が y である条件の下での X の確率密度分布， X の実現値が x である条件の下での Y の確率密度分布は，それぞれ

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \quad (1.11)$$

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)} \quad (1.12)$$

として与えられる． $p_{X|Y}(x|y)$ ， $p_{Y|X}(y|x)$ は，条件付確率密度分布 (conditional probability density distribution) と呼ばれ， X, Y が独立ならば，それぞれの周辺分布 $p_X(x)$ ， $p_Y(y)$ に一致する．

複数の確率変数の確率的な性質は，結合確率密度分布 $p_{X,Y}(x,y)$ で表現されるが，1 変数の場合と同様に，結合確率密度分布 $p_{X,Y}(x,y)$ と適当な 2 変数関数 $g(x,y)$ との内積，つまり $g(X,Y)$ の期待値

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)p_{X,Y}(x,y)dx dy \quad (1.13)$$

でその一部の性質を表現することが可能である．例えば， $g(x,y) = x$ ， $g(x,y) = y$ とした場合は，それぞれ

$$\mu_X \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xp_{X,Y}(x,y)dx dy = \int_{-\infty}^{\infty} xp_X(x)dx \quad (1.14)$$

$$\mu_Y \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yp_{X,Y}(x,y)dx dy = \int_{-\infty}^{\infty} yp_Y(y)dy \quad (1.15)$$

となり， X, Y のそれぞれの平均を表す．また， $g(x,y) = (x - \mu_X)^2$ ， $g(x,y) = (y - \mu_Y)^2$ とした場合は，それぞれ

$$\sigma_X^2 \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^2 p_{X,Y}(x,y)dx dy = \int_{-\infty}^{\infty} (x - \mu_X)^2 p_X(x)dx \quad (1.16)$$

$$\sigma_Y^2 \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_Y)^2 p_{X,Y}(x,y)dx dy = \int_{-\infty}^{\infty} (y - \mu_Y)^2 p_Y(y)dy \quad (1.17)$$

となり， X, Y のそれぞれの分散を表すことになる．

平均，分散は，1 つの確率変数の性質を表す統計量である．一方，複数の確率変数の間の関係，例えば， X が大きい値を実現値にとれば， Y もまた大きい値を実現値にとりやすくなるといった性質を表現するためには， $g(x,y) = xy$ ，あるいは $g(x,y) = (x - \mu_X)(y - \mu_Y)$ が用いられる．それぞれ，

$$R_{X,Y} \equiv E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyp_{X,Y}(x,y)dx dy \quad (1.18)$$

$$C_{X,Y} \equiv E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)p_{X,Y}(x,y)dx dy \quad (1.19)$$

となり， X と Y の間の相関 (correlation)，共分散 (covariance) と呼ばれる．相関 $R_{X,Y} = 0$ の場合， X, Y は直交 (orthogonal) しているという．また， $E[XY] = E[X]E[Y]$ となるならば， X, Y は無相関 (uncorrelated) であるといい，共分散は，

$$C_{X,Y} = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_Y E[X] - \mu_X E[Y] + \mu_X \mu_Y = 0$$

となる． X, Y が独立ならば，

$$R_{X,Y} = E[XY] = \int_{-\infty}^{\infty} xyf_{X,Y}(x,y)dxdy = \int_{-\infty}^{\infty} xf_X(x)dx \cdot \int_{-\infty}^{\infty} yf_Y(y)dy = E[X]E[Y]$$

となるので， X, Y は無相関になる．さて， X, Y の共分散 $C_{X,Y}$ をそれぞれの標準偏差の積で規格化した

$$\rho_{X,Y} \equiv \frac{C_{X,Y}}{\sigma_X \sigma_Y} \quad (1.20)$$

は， $[-1, 1]$ の値をとる相関係数 (correlation coefficient) と呼ばれる統計量になる．一般に，

$$E[X^k Y^{n-k}], \quad k = 0, 1, \dots, n \quad (1.21)$$

$$E[(X - \mu_X)^k (Y - \mu_Y)^{n-k}], \quad k = 0, 1, \dots, n \quad (1.22)$$

は，それぞれ X, Y の n 次結合モーメント (joint moment)， n 次結合中心モーメント (joint central moment) と呼ばれる．1 次結合モーメントには， X, Y のそれぞれの平均，2 次結合中心モーメントには， X, Y のそれぞれの分散とそれらの共分散が含まれる．また， n 次結合モーメントや n 次結合中心モーメントを n 次統計量という．

1 変数の場合と同様に， $g(x, y) = e^{s_x x + s_y y}$ ，すなわち， $\varphi(s_x, s_y) \equiv E[e^{s_x X + s_y Y}]$ は，2 つの確率変数 X, Y に対するモーメント母関数を与える． $e^{s_x x + s_y y}$ をマクローリン展開すると，

$$e^{s_x x + s_y y} = 1 + (s_x x + s_y y) + \frac{1}{2!}(s_x x + s_y y)^2 + \frac{1}{3!}(s_x x + s_y y)^3 + \dots$$

となるので，

$$\begin{aligned} \varphi(s_x, s_y) &= E\left[1 + (s_x X + s_y Y) + \frac{1}{2!}(s_x X + s_y Y)^2 + \frac{1}{3!}(s_x X + s_y Y)^3 + \dots\right] \\ &= 1 + s_x E[X] + s_y E[Y] \\ &\quad + \frac{1}{2!}\left(s_x^2 E[X^2] + s_y^2 E[Y^2] + 2s_x s_y E[XY]\right) \\ &\quad + \frac{1}{3!}\left(s_x^3 E[X^3] + 3s_x^2 s_y E[X^2 Y] + 3s_x s_y^2 E[XY^2] + s_y^3 E[Y^3]\right) \\ &\quad + \dots \end{aligned} \quad (1.23)$$

となる．式 (1.23) の両辺を s_x, s_y で微分して， $s_x = s_y = 0$ とおくと

$$\begin{aligned} E[X] &= \left. \frac{\partial}{\partial s_x} \varphi(s_x, s_y) \right|_{s_x=s_y=0} & E[Y] &= \left. \frac{\partial}{\partial s_y} \varphi(s_x, s_y) \right|_{s_x=s_y=0} \\ E[X^2] &= \left. \frac{\partial^2}{\partial s_x^2} \varphi(s_x, s_y) \right|_{s_x=s_y=0} & E[XY] &= \left. \frac{\partial^2}{\partial s_x \partial s_y} \varphi(s_x, s_y) \right|_{s_x=s_y=0} \\ E[Y^2] &= \left. \frac{\partial^2}{\partial s_y^2} \varphi(s_x, s_y) \right|_{s_x=s_y=0} & \dots & \end{aligned}$$

となり， X, Y の各次数の結合モーメントを表す． $g(x, y) = e^{s_x x + s_y y}$ の代わりに， $g(x, y) = e^{i(s_x x + s_y y)}$ とおいた $E[e^{i(s_x X + s_y Y)}]$ は，2 変数の特性関数と呼ばれる．

1 変数の場合と同様に，モーメント母関数 $\varphi(s_x, s_y)$ の対数 $\log \varphi(s_x, s_y)$ をベキに展開した

$$\begin{aligned} \log \varphi(s_x, s_y) &= c + c_x s_x + c_y s_y \\ &\quad + \frac{1}{2}(c_{xx} s_x^2 + 2c_{xy} s_x s_y + c_{yy} s_y^2) \\ &\quad + \frac{1}{3!}(c_{xxx} s_x^3 + 3c_{xxy} s_x^2 s_y + 3c_{xyy} s_x s_y^2 + c_{yyy} s_y^3) \\ &\quad + \dots \end{aligned} \quad (1.24)$$

の係数 (c_{xy}, c_{xx}, c_{yy} など) は, キュムラントと呼ばれる. 式 (1.23) より, $\varphi(s_x, s_y)|_{s_x=s_y=0} = 1$ なので, その対数は, $\log \varphi(s_x, s_y)|_{s_x=s_y=0} = 0$ となる. したがって, 式 (1.24) より, $c = 0$ であることがわかる.

さて, X, Y が独立である場合には,

$$\begin{aligned}\varphi(s_x, s_y) &= \int_{-\infty}^{\infty} e^{s_x x + s_y y} p_{X,Y}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} e^{s_x x} p_X(x) dx \int_{-\infty}^{\infty} e^{s_y y} p_Y(y) dy \\ &= \varphi(s_x) \varphi(s_y)\end{aligned}\tag{1.25}$$

と個々の確率変数に対するモーメント母関数の積で表現される. 対数をとれば, 積は和で表現されるから, X, Y が独立である場合のキュムラント母関数は,

$$\log \varphi(s_x, s_y) = \log \varphi(s_x) + \log \varphi(s_y)$$

とそれぞれのキュムラント母関数の和で表される. したがって, X, Y が独立である場合

$$\begin{aligned}\log \varphi(s_x, s_y) &= c_x s_x + c_y s_y \\ &\quad + \frac{1}{2}(c_{xx} s_x^2 + c_{yy} s_y^2) \\ &\quad + \frac{1}{3!}(c_{xxx} s_x^3 + c_{yyy} s_y^3) \\ &\quad + \frac{1}{4!}(c_{xxxx} s_x^4 + c_{yyyy} s_y^4) + \dots\end{aligned}\tag{1.26}$$

となり, 複数の確率変数にまたがるキュムラントはゼロになる. 複数の確率変数にまたがるキュムラントはクロスキュムラントと呼ばれ, 確率変数の間の独立性を見るために利用される.

キュムラントとモーメントの関係を導いてみよう. キュムラント母関数を s_x で偏微分すれば,

$$\frac{\partial}{\partial s_x} \log \varphi(s_x, s_y) = \frac{1}{\varphi(s_x, s_y)} \frac{\partial}{\partial s_x} \varphi(s_x, s_y)$$

となるので,

$$\varphi(s_x, s_y) \frac{\partial}{\partial s_x} \log \varphi(s_x, s_y) = \frac{\partial}{\partial s_x} \varphi(s_x, s_y)$$

としておいて, 式 (1.23), 式 (1.24) を適用し, s_x, s_y の各べきの項を比較する. これを, s_y での偏微分, s_x での 2 階微分, s_y での 2 階微分, s_x と s_y での 2 階微分などについても同様に行うことにより,

$$\begin{aligned}c_x &= E[X], \quad c_y = E[Y], \\ c_{xx} &= E[X^2] - E[X]^2, \quad c_{yy} = E[Y^2] - E[Y]^2, \quad c_{xy} = E[XY] - E[X]E[Y], \\ c_{xxx} &= E[X^3] - 3E[X^2]E[X] + 2E[X]^3, \quad c_{yyy} = E[Y^3] - 3E[Y^2]E[Y] + 2E[Y]^3, \\ c_{xxy} &= E[X^2Y] - 2E[X]E[XY] - E[X^2]E[Y] + 2E[X]^2E[Y], \\ c_{xyy} &= E[XY^2] - 2E[XY]E[Y] - E[X]E[Y^2] + 2E[X]E[Y]^2, \dots\end{aligned}$$

なる関係が得られる. 上の $c_x, c_y, c_{xx}, c_{yy}, c_{xxx}, c_{yyy}$ などは, 1 変数の場合のキュムラントとモーメントの関係と同じものである.

1.3 複数の確率変数の間の性質

2 個の確率変数の表現は, N 個の確率変数 X_1, \dots, X_N が存在する場合に一般化できる. 例えば, キュムラントとモーメントの関係は, 次式で表現される.

$$c_{x_i} = E[X_i], \quad \forall i \in \{1, 2, \dots, N\}$$

$$\begin{aligned}
c_{x_i x_j} &= E[X_i X_j] - E[X_i]E[X_j], \quad \forall i, j \in \{1, 2, \dots, N\} \\
c_{x_i x_j x_k} &= E[X_i X_j X_k] - E[X_i]E[X_j X_k] - E[X_j]E[X_i X_k] - E[X_k]E[X_i X_j] \\
&\quad + 2E[X_i]E[X_j]E[X_k], \quad \forall i, j, k \in \{1, 2, \dots, N\} \\
&\vdots
\end{aligned} \tag{1.27}$$

i, j, k のうち、2つが等しいとすれば、2変数の場合のキュムラントとモーメントの関係を導くことができる。同様に、3つが等しいものとすれば、1変数の場合の同様の関係が導かれる。また、2変数の場合を拡張すれば、 X_i, X_j, X_k が2つ以上の独立なものに分離できる場合には、対応するキュムラント $c_{x_i x_j x_k}$ はゼロになることがわかる。

複数の確率変数を扱う場合、各確率変数の実現値を列ベクトル $\mathbf{x} \equiv (x_1, \dots, x_N)^T$ 、確率変数を列ベクトル $X \equiv (X_1, X_2, \dots, X_N)^T$ で表現することが多い。また、平均値、2次結合中心モーメントは、ベクトル、行列を用いて、それぞれ

$$\boldsymbol{\mu}_X \equiv \begin{pmatrix} \mu_{X_1} \\ \mu_{X_2} \\ \vdots \\ \mu_{X_N} \end{pmatrix}, \quad \mathbf{C}_X = \begin{pmatrix} \sigma_{X_1}^2 & C_{X_1, X_2} & \cdots & C_{X_1, X_N} \\ C_{X_1, X_2} & \sigma_{X_2}^2 & \cdots & C_{X_2, X_N} \\ \vdots & \vdots & \ddots & \vdots \\ C_{X_1, X_N} & C_{X_2, X_N} & \cdots & \sigma_{X_N}^2 \end{pmatrix}$$

として表現される。行列 \mathbf{C}_X は、分散共分散行列 (variance covariance matrix) と呼ばれている。

1.4 正規確率変数

信号処理で用いられる最も重要な確率密度分布 $p_X(x)$ は、正規分布であり、正規分布に従う確率変数を正規確率変数という。正規分布は、平均、分散をそれぞれ μ, σ^2 とすれば、次式で与えられる。

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{1.28}$$

平均 μ 、分散 σ^2 の正規分布を $N(\mu, \sigma^2)$ と書く。正規分布 $N(\mu, \sigma^2)$ に従う確率変数 X のモーメント母関数 $\varphi(s)$ は、

$$\int_{-\infty}^{\infty} \exp(-(ax^2 + bx + c)) dx = \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a} - c\right)$$

なる公式を利用すれば、

$$\varphi(s) = \exp\left(\mu s + \frac{\sigma^2}{2} s^2\right) \tag{1.29}$$

となり、キュムラント母関数 $\log \varphi(s)$ は、その対数であるから、

$$\log \varphi(s) = \mu s + \frac{\sigma^2}{2} s^2 \tag{1.30}$$

となる。式 (1.5) と式 (1.30) を比較することにより、正規確率変数では、3次以上のキュムラントが0になることがわかる。また、正規確率変数の中心モーメントは、 $\mu = 0$ とおいた式 (1.29) を Taylor 展開し、式 (1.4) と比較することにより、

$$C_n = \begin{cases} 0, & n \text{ is odd} \\ 1 \cdot 3 \cdot 5 \cdots (n-1) \sigma^n, & n \text{ is even} \end{cases} \tag{1.31}$$

となることが分かる．

多次元正規分布は，次式で表現される．

$$p_X(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}_X|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_X)^T \mathbf{C}_X^{-1}(\mathbf{x} - \boldsymbol{\mu}_X)\right) \quad (1.32)$$

N 個の確率変数 X_1, X_2, \dots, X_N が正規分布に従う場合，キュムラント母関数は，

$$\log \varphi(s_1, \dots, s_N) = \sum_{j=1}^N \mu_{X_j} s_j + \frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N C_{X_j, X_k} s_j s_k$$

として表される．したがって，1次キュムラントは，平均に一致し ($c_{x_j} = \mu_{X_j}$)，2次キュムラントは，共分散に一致する ($c_{x_j x_k} = C_{X_j, X_k}$)．また，平均が0である場合，1次キュムラントが0になるから，結合モーメントは，

$$\begin{aligned} E[X_{j_1} \cdots X_{j_k}] &= \left. \frac{\partial}{\partial s_{j_1}} \cdots \frac{\partial}{\partial s_{j_k}} \exp\left(\frac{1}{2} \sum_{i=1}^N \sum_{k=1}^N C_{X_i, X_k} s_i s_k\right) \right|_{s_1 = \cdots = s_N = 0} \\ &= \begin{cases} 0, & k : \text{odd number} \\ \sum \prod_{(l,m)} C_{lm}, & k : \text{even number} \end{cases} \end{aligned} \quad (1.33)$$

となる．上式中， $\sum \prod_{(l,m)}$ は， j_1, \dots, j_k を $k/2$ 個のペア (l, m) に分割したものについて積をとり，そのすべての組み合わせについて総和をとることを意味する．例えば，4変数の場合には，

$$E[X_1 X_2 X_3 X_4] = C_{X_1, X_2} C_{X_3, X_4} + C_{X_1, X_3} C_{X_2, X_4} + C_{X_1, X_4} C_{X_2, X_3}$$

となる．このように，正規確率変数は，3次以上のモーメントは，2次までのモーメントで表現可能であり，2次までの統計的性質で完全に表現される．

平均 μ ，分散 σ^2 の確率変数 X は， $(X - \mu)/\sigma$ なる変換を行うと，平均0，分散1の確率変数になる．平均0，分散1の確率変数 X を考え，その確率密度分布 $p_X(x)$ を，正規分布 $N(0, 1)$ に対して直交性

$$\int_{-\infty}^{\infty} H_k(x) H_n(x) p(x) dx = \begin{cases} k! & \text{for } k = n \\ 0 & \text{for } k \neq n \end{cases} \quad (1.34)$$

を満たす直交関数系 $H_k(x)$ ， $k = 0, 1, \dots$ を用いて

$$p_X(x) = \sum_{n=0}^{\infty} h_n H_n(x) p(x) \quad (1.35)$$

と展開することを考える．ただし， $p(x)$ は，正規分布 $N(0, 1)$ である． $H_n(x)$ を多項式により構成すると，エルミート多項式 (Hermite polynomial) になる．エルミート多項式の漸化式は，

$$H_{n+1}(x) = xH_n(x) - nH_{n-1}(x)$$

となり，最初の数項は

$$\begin{aligned}
 H_0(x) &= 1 \\
 H_1(x) &= x \\
 H_2(x) &= x^2 - 1 \\
 H_3(x) &= x^3 - 3x \\
 H_4(x) &= x^4 - 6x^2 + 3 \\
 H_5(x) &= x^5 - 10x^3 + 15x \\
 H_6(x) &= x^6 - 15x^4 + 45x^2 - 15 \\
 &\vdots
 \end{aligned} \tag{1.36}$$

となる．

次に，展開係数 h_k を求めることを考えよう．式 (1.35) の両辺に $H_k(x)$ を乗じ， x に関して $[-\infty, \infty]$ の範囲で積分すると，

$$\begin{aligned}
 \int_{-\infty}^{\infty} H_k(x)p_X(x)dx &= \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} h_n H_k(x)H_n(x)p(x)dx \\
 &= \sum_{n=0}^{\infty} h_n \int_{-\infty}^{\infty} H_k(x)H_n(x)p(x)dx
 \end{aligned}$$

となり，式 (1.34) の直交性から， $k \neq n$ では，上式の右辺の積分は 0 であり， $k = n$ では $k!$ である．したがって，展開係数 h_k は

$$h_k = \frac{1}{k!} \int_{-\infty}^{\infty} H_k(x)p_X(x)dx$$

となる．上式の $H_k(x)$ に，式 (1.36) を代入することにより，モーメントを用いた表現

$$\begin{aligned}
 h_0 &= 1 \\
 h_1 &= 0 \\
 h_2 &= 0 \\
 h_3 &= \frac{1}{3!}C_3 \\
 h_4 &= \frac{1}{4!}(C_4 - 3) \\
 h_5 &= \frac{1}{5!}(C_5 + 10C_3) \\
 h_6 &= \frac{1}{6!}(C_6 - 15C_4 + 30) \\
 &\vdots
 \end{aligned}$$

が得られる．こうした確率密度関数の正規分布に対する直交展開をグラム - シャリア展開 (Gram-Charlier expansion) という．

第2章 推定の基礎

2.1 推定とその性質

確率的な現象は、確率変数とその(結合)確率密度分布により表現され、確率変数の適当な関数の期待値が、確率密度分布の特定の性質を表す統計量として用いられる。統計量としては、平均、分散、(中心)モーメント、キュムラントなどがしばしば用いられる。さて、期待値により与えられるこうした統計量は、確率密度分布がわからない限り、その真の値を知ることはできない。そこで、観測により得られた確率変数の実現値、つまり標本から、その統計量を推測することが必要になる。こうした統計量の推測を推定(estimation)という。推定された統計量を推定値(estimated value)といい、標本の関数として推定値を定めたものを推定量(estimator)という。ここでは、統計量を θ と書くことにし、推定された統計量、つまり統計量 θ の推定値を $\hat{\theta}$ とハットを付けて書くことにする。ただし、複数の統計量を持つ場合には、統計量は、 $\theta \equiv (\theta_1, \dots, \theta_L)^T$ とベクトルになる。また、推定量は、確率的に変動する標本の関数であるから、推定量そのものが確率変数であることに注意しなければならない。

推定には、上で述べたように、 $\hat{\theta}$ を θ の推定値とする場合と、あらかじめ決められた確率で $\hat{\theta}_l < \theta < \hat{\theta}_u$ の範囲にあるとする場合がある。前者を点推定(point estimation)といい、後者を区間推定(interval estimation)という。以降では、特に断らない限り、推定とは点推定を意味するものとする。

次に、推定量として望まれる性質を考えよう。推定量 $\hat{\theta}$ は、真の統計量 θ を良く表現していることが望ましいのだが、この「良さ」を図る尺度として、次の一致性(consistency)、不偏性(unbiasness)、有効性(efficiency)が利用される。

[一致性] N 個の標本 x_1, x_2, \dots, x_N から推定された推定量を $\hat{\theta}^{(N)}$ と書くことにする。標本数 N を増やしたとき、推定値 $\hat{\theta}^{(N)}$ が真値 θ に近づくならば、この推定量は一致性を持つという。厳密には、任意の $\varepsilon > 0$ に対して、

$$\lim_{N \rightarrow \infty} \text{Prob}[|\hat{\theta}^{(N)} - \theta| < \varepsilon] = 1 \quad (2.1)$$

となる性質である。ただし、 $\text{Prob}[\]$ は確率を表す。一致性を持つ推定量を一致推定量といい、一致推定量となるための十分条件は、

$$\lim_{N \rightarrow \infty} E[\hat{\theta}^{(N)}] = \theta, \quad \lim_{N \rightarrow \infty} \text{Var}[\hat{\theta}^{(N)}] = 0 \quad (2.2)$$

である。ただし、 $\text{Var}[\]$ は分散を表す。

[不偏性] 不偏性とは、推定に用いる標本の選び方によらず、推定値 $\hat{\theta}$ の期待値が真値 θ に等しい、すなわち、いかに標本 x_1, x_2, \dots, x_N を抽出しても、

$$E[\hat{\theta}^{(N)}] = \theta, \quad \forall N \quad (2.3)$$

となる性質である。不偏性を持つ推定量を不偏推定量という。また、

$$\lim_{N \rightarrow \infty} E[\hat{\theta}^{(N)}] = \theta \quad (2.4)$$

となる場合には、漸近的に不偏(asymptotical unbiasedness)であるという。

[有効性] 推定値 $\hat{\theta}$ の分散，複数の推定値がある場合には，それらの分散共分散行列 $\text{Cov}[\hat{\theta}]$ が 2.3 で述べるクラメール・ラオの下界 (Cramér-Rao's lower bound) を達成する不偏推定量を有効推定量という．

[演習 1] 平均が μ ，分散が σ^2 であるような確率変数 X の N 個の標本 x_1, x_2, \dots, x_N が観測されたとし，これらから， X の平均，分散を以下の (a) ~ (d) の推定量で推定するものとしよう．(a) は標本平均を平均の推定量とする場合，(b) は平均 μ が既知である際に，標本分散を分散の推定量とする場合，(c) は，平均 μ が既知でなく，(b) の平均 μ を標本平均として推定された平均 $\hat{\mu}$ で置き換えた場合，(d) は，(c) で，標本数 N の代わりに $N - 1$ で割る場合である．これらの推定量が一致性，不偏性，もしくは漸近的な不偏性を持つかどうかを数値実験により調べてみよう．

(a) 平均の推定量 $\hat{\mu}$

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^N x_n \quad (2.5)$$

(b) 分散の推定量 $\hat{\sigma}_A^2$

$$\hat{\sigma}_A^2 = \frac{1}{N} \sum_{n=1}^N (x_n - \mu)^2 \quad (2.6)$$

(c) 分散の推定量 $\hat{\sigma}_B^2$

$$\hat{\sigma}_B^2 = \frac{1}{N} \sum_{n=1}^N (x_n - \hat{\mu})^2 \quad (2.7)$$

(d) 分散の推定量 $\hat{\sigma}_C^2$

$$\hat{\sigma}_C^2 = \frac{1}{N-1} \sum_{n=1}^N (x_n - \hat{\mu})^2 \quad (2.8)$$

[解答] 計算機により，平均 $\mu = 10$ ，分散 $\sigma^2 = 2$ の正規分布 $N(10, 2)$ に従う標本を N 個生成し，それらを x_1, x_2, \dots, x_N とする．これらから式 (2.5) により平均値を推定する．これを $\hat{\mu}^{(N)}$ とする．推定値の期待値は，推定値の確率密度分布が既知でないと求められないため，十分大きな数の推定値の標本に対する標本平均で，期待値を近似することにしよう．ここでは，1000 個の $\hat{\mu}^{(N)}$ の標本を用意し，それらの標本平均，標本分散を $E[\hat{\mu}^{(N)}], \text{Var}[\hat{\mu}^{(N)}]$ のそれぞれの近似値とする．標本数 N を 2 から 1000 まで変化させた際の $E[\hat{\mu}^{(N)}]$ の変化を図 2.1(a) に記号 \ominus で示す．また， $E[\hat{\mu}^{(N)}] \pm \sqrt{\text{Var}[\hat{\mu}^{(N)}]}$ を同図 (a) に記号 \oplus で示す． $E[\hat{\mu}^{(N)}] = \mu, \forall N$ となっており，式 (2.5) で示した平均の推定量 $\hat{\mu}$ は，不偏性を満たすことが分かる．また，記号 \oplus で示した推定値 \pm 標準偏差の幅が $N \rightarrow \infty$ に対して 0 に近づいて行くことから， $\lim_{N \rightarrow \infty} \text{Var}[\hat{\mu}^{(N)}] = 0$ となることが分かる．すなわち，式 (2.5) で示した平均の推定量 $\hat{\mu}$ は，一致性を満たすことが分かる．分散の推定量 $\hat{\sigma}_A^2, \hat{\sigma}_B^2, \hat{\sigma}_C^2$ についても同様に調べた結果を図 2.1(b) ~ (d) に示す．分散の推定量 $\hat{\sigma}_A^2, \hat{\sigma}_C^2$ は，不偏性，一致性を満たし，分散の推定量 $\hat{\sigma}_B^2$ は，一致性と漸近的な不偏性を満たすが，不偏性を満たさないことが分かる．

演習 1 で得られた結果を理論的に確認してみよう．まず， $\hat{\mu}, \hat{\sigma}_A^2, \hat{\sigma}_B^2, \hat{\sigma}_C^2$ の不偏性について調べてみよう．式 (2.5) の両辺の期待値をとれば，

$$E[\hat{\mu}] = E\left[\frac{1}{N} \sum_{n=1}^N x_n\right] = \frac{1}{N} \sum_{n=1}^N E[x_n] = \frac{1}{N} \sum_{n=1}^N \mu = \mu \quad (2.9)$$

となり，平均の推定量 $\hat{\mu}$ は，不偏性を持つことが分かる．同様に，式 (2.6) の両辺の期待値をとれば，

$$E[\hat{\sigma}_A^2] = \frac{1}{N} \sum_{n=1}^N E[(x_n - \mu)^2] = \frac{1}{N} \sum_{n=1}^N \sigma^2 = \sigma^2 \quad (2.10)$$

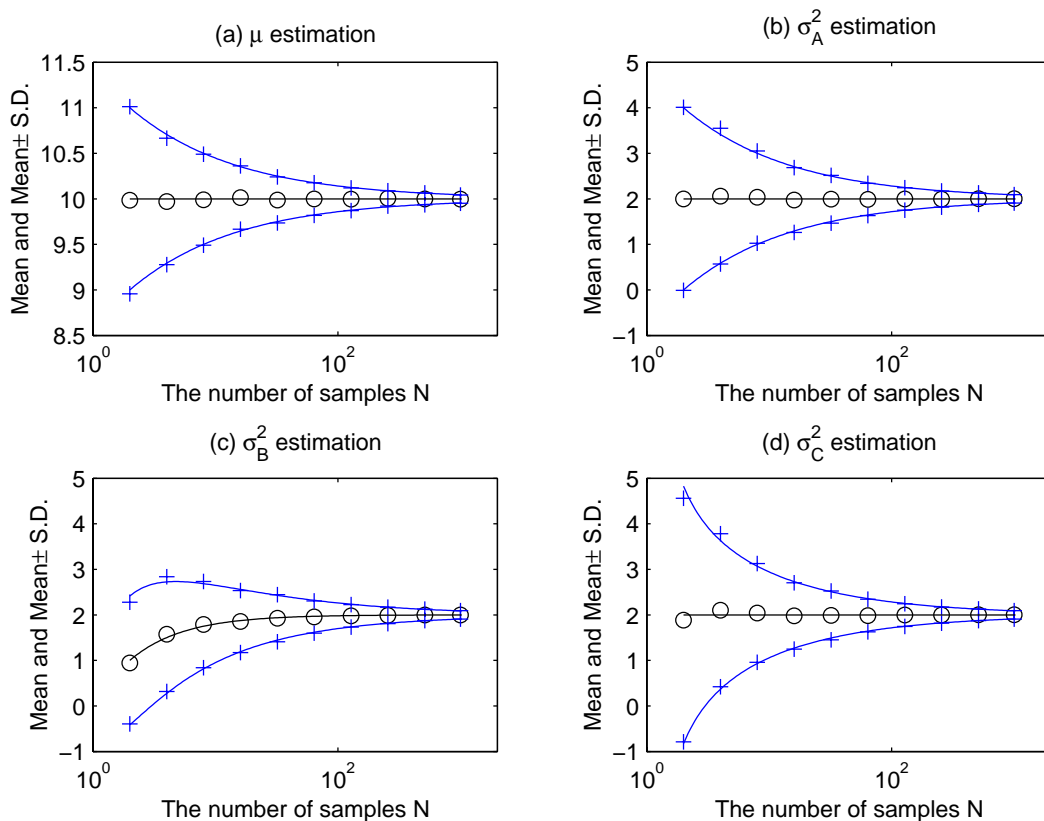


図 2.1: 演習 1 の解答：一致性，不偏性，漸近的な不偏性の検証

となり，分散の推定量 $\hat{\sigma}_A^2$ は不偏性を持つことが分かる．また，平均の真値 μ が未知である場合には，平均の推定値 $\hat{\mu}$ を用いて，分散を推定しなければならない．その場合の分散の推定式 (2.7) が

$$\begin{aligned}\hat{\sigma}_B^2 &= \frac{1}{N} \sum_{n=1}^N (x_n - \hat{\mu})^2 \\ &= \frac{1}{N} \sum_{n=1}^N (x_n - \mu)^2 - (\hat{\mu} - \mu)^2\end{aligned}\quad (2.11)$$

と変形できることを利用すれば，その期待値は，

$$E[\hat{\sigma}_B^2] = \frac{1}{N} \sum_{n=1}^N E[(x_n - \mu)^2] - E[(\hat{\mu} - \mu)^2]\quad (2.12)$$

となる．最後の式の第 1 項は，式 (2.10) より σ^2 である．一方，第 2 項は，

$$\begin{aligned}E[(\hat{\mu} - \mu)^2] &= E\left[\left(\frac{1}{N} \sum_{n=1}^N x_n - \mu\right)^2\right] = E\left[\left(\frac{1}{N} \sum_{n=1}^N (x_n - \mu)\right)^2\right] \\ &= E\left[\frac{1}{N^2} \sum_{n=1}^N \sum_{k=1}^N (x_n - \mu)(x_k - \mu)\right] = \frac{1}{N^2} \sum_{n=1}^N \sum_{k=1}^N E[(x_n - \mu)(x_k - \mu)] \\ &= \frac{1}{N^2} \sum_{n=1}^N \sum_{k=1}^N \delta_{n,k} \sigma^2 = \frac{1}{N} \sigma^2\end{aligned}\quad (2.13)$$

となる．したがって，

$$E[\hat{\sigma}_B^2] = \sigma^2 - \frac{1}{N}\sigma^2 = \frac{N-1}{N}\sigma^2 \quad (2.14)$$

である．したがって，式 (2.7) で与えられる分散の推定値は，不偏性を持たないことが分かる．ただし， $N \rightarrow \infty$ では， $E[\hat{\sigma}_B^2] \rightarrow \sigma^2$ となることから，漸近的には不偏である．式 (2.7) と式 (2.8) から，

$$\hat{\sigma}_C^2 = \frac{N}{N-1}\hat{\sigma}_B^2$$

なので，両辺の期待値をとれば，

$$E[\hat{\sigma}_C^2] = \frac{N}{N-1}E[\hat{\sigma}_B^2] = \frac{N}{N-1}\frac{N-1}{N}\sigma^2 = \sigma^2$$

となる．したがって，式 (2.8) で与えられる分散の推定値は，不偏性を持つ．式 (2.8) を用いた分散の推定値 $\hat{\sigma}_C^2$ を不偏分散と呼ぶことがある．ここで理論的に求めた $E[\hat{\mu}]$, $E[\hat{\sigma}_A^2]$, $E[\hat{\sigma}_B^2]$, $E[\hat{\sigma}_C^2]$ を図 2.1(a) ~ (d) に実線で重ねて示した．

次に， $\hat{\mu}$, $\hat{\sigma}_A^2$, $\hat{\sigma}_B^2$, $\hat{\sigma}_C^2$ の一致性について調べてみよう．まず， $\hat{\mu}$ の分散 $\text{Var}[\hat{\mu}]$ は，式 (2.13) の結果を利用すれば，

$$\begin{aligned} \text{Var}[\hat{\mu}] &= E[(\hat{\mu} - E[\hat{\mu}])^2] = E[(\hat{\mu} - \mu)^2] \\ &= \frac{1}{N^2}N\sigma^2 = \frac{1}{N}\sigma^2 \end{aligned} \quad (2.15)$$

となる． $\lim_{N \rightarrow \infty} \text{Var}[\hat{\mu}] = \lim_{N \rightarrow \infty} \frac{1}{N}\sigma^2 = 0$ であるから， $\hat{\mu}$ は，式 (2.2) の第 2 式を満たし，一致推定量であることになる．同様に， $\hat{\sigma}_A^2$ の分散 $\text{Var}[\hat{\sigma}_A^2]$ は，

$$\begin{aligned} \text{Var}[\hat{\sigma}_A^2] &= E[(\hat{\sigma}_A^2 - E[\hat{\sigma}_A^2])^2] = E[(\hat{\sigma}_A^2 - \sigma^2)^2] \\ &= E\left[\left(\frac{1}{N}\sum_{n=1}^N(x_n - \mu)^2 - \sigma^2\right)^2\right] = E\left[\left(\frac{1}{N}\sum_{n=1}^N((x_n - \mu)^2 - \sigma^2)\right)^2\right] \\ &= E\left[\frac{1}{N^2}\sum_{n=1}^N\sum_{k=1}^N((x_n - \mu)^2 - \sigma^2)((x_k - \mu)^2 - \sigma^2)\right] \\ &= \frac{1}{N^2}\sum_{n=1}^N\sum_{k=1}^N\left(E[(x_n - \mu)^2(x_k - \mu)^2] - \sigma^2E[(x_n - \mu)^2] - \sigma^2E[(x_k - \mu)^2] + \sigma^4\right) \\ &= \frac{1}{N^2}\left(NC_4 + (N^2 - N)\sigma^4 - N^2\sigma^4 - N^2\sigma^4 + N^2\sigma^4\right) \\ &= \frac{1}{N}(C_4 - \sigma^4) \end{aligned} \quad (2.16)$$

となる．ただし，上式において， C_4 は， X の 4 次中心モーメントを表す． $\lim_{N \rightarrow \infty} \text{Var}[\hat{\sigma}_A^2] = \lim_{N \rightarrow \infty} \frac{1}{N}(C_4 - \sigma^4) = 0$ であるから， $\hat{\sigma}_A^2$ は，式 (2.2) の第 2 式を満たし，一致推定量であることになる．上式において，4 次中心モーメント C_4 は， X が正規分布に従うならば，式 (1.31) より， $C_4 = 1 \cdot 3\sigma^4$ となる．したがって，

$$\text{Var}[\hat{\sigma}_A^2] = \frac{1}{N}(3\sigma^4 - \sigma^4) = \frac{2\sigma^4}{N} \quad (2.17)$$

となる．一方， $\hat{\sigma}_B^2$ の分散は，

$$\text{Var}[\hat{\sigma}_B^2] = E[(\hat{\sigma}_B^2 - E[\hat{\sigma}_B^2])^2] = E[(\hat{\sigma}_B^2 - \sigma^2)^2] - (E[\hat{\sigma}_B^2] - \sigma^2)^2 \quad (2.18)$$

と書き直せるので、まず、右辺第1項は、

$$\begin{aligned}
E[(\hat{\sigma}_B^2 - \sigma^2)^2] &= E[(\hat{\sigma}_B^2)^2] - 2E[\hat{\sigma}_B^2]\sigma^2 + \sigma^4 \\
&= E[(\hat{\sigma}_B^2)^2] - 2\frac{N-1}{N}\sigma^2\sigma^2 + \sigma^4 \\
&= E[(\hat{\sigma}_B^2)^2] + (-1 + \frac{2}{N})\sigma^4
\end{aligned} \tag{2.19}$$

となり、右辺第2項は、

$$\begin{aligned}
(E[\hat{\sigma}_B^2] - \sigma^2)^2 &= (\frac{N-1}{N}\sigma^2 - \sigma^2)^2 \\
&= \frac{1}{N^2}\sigma^4
\end{aligned} \tag{2.20}$$

となる。ここで、

$$\begin{aligned}
\hat{\sigma}_B^2 &= \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2 \\
&= \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 - (\hat{\mu} - \mu)^2
\end{aligned} \tag{2.21}$$

であることを利用すれば、 $E[(\hat{\sigma}_B^2)^2]$ は、

$$\begin{aligned}
E[(\hat{\sigma}_B^2)^2] &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E[(x_i - \mu)^2(x_j - \mu)^2] + E[(\hat{\mu} - \mu)^4] - \frac{2}{N} \sum_{i=1}^N E[(x_i - \mu)^2(\hat{\mu} - \mu)^2] \\
&= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E[(x_i - \mu)^2(x_j - \mu)^2] + E[(\frac{1}{N} \sum_{i=1}^N x_i - \mu)^4] \\
&\quad - \frac{2}{N} \sum_{i=1}^N E[(x_i - \mu)^2(\frac{1}{N} \sum_{j=1}^N x_j - \mu)^2] \\
&= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E[(x_i - \mu)^2(x_j - \mu)^2] \\
&\quad + \frac{1}{N^4} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N \sum_{l=1}^N E[(x_i - \mu)(x_j - \mu)(x_k - \mu)(x_l - \mu)] \\
&\quad - \frac{2}{N^3} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N E[(x_i - \mu)^2(x_j - \mu)(x_k - \mu)] \\
&= \frac{1}{N^2}(NC_4 + (N^2 - N)\sigma^4) + \frac{1}{N^4}(NC_4 + 3(N^2 - N)\sigma^4) - \frac{2}{N^3}(NC_4 + (N^2 - N)\sigma^4) \\
&= (\frac{1}{N} - \frac{2}{N^2} + \frac{1}{N^3})C_4 + (1 - \frac{3}{N} + \frac{5}{N^2} - \frac{3}{N^3})\sigma^4
\end{aligned} \tag{2.22}$$

となる。X が正規分布に従うならば、 $C_4 = 3\sigma^4$ なので、上式は、

$$E[(\hat{\sigma}_B^2)^2] = (1 - \frac{1}{N^2})\sigma^4 \tag{2.23}$$

となる。結局、 $\hat{\sigma}_B^2$ の分散は、

$$\text{Var}[\hat{\sigma}_B^2] = \frac{N-1}{N^2}2\sigma^4 \tag{2.24}$$

となる． $\hat{\sigma}_C^2 = \frac{N}{N-1}\hat{\sigma}_B^2$ であることから， $\hat{\sigma}_C^2$ の分散は， X が正規分布に従うものとするれば，

$$\text{Var}[\hat{\sigma}_C^2] = \frac{N^2}{(N-1)^2}\text{Var}[\hat{\sigma}_B^2] = \frac{1}{N-1}2\sigma^4 \quad (2.25)$$

となることがわかる．

$\hat{\sigma}_A^2$, $\hat{\sigma}_B^2$, $\hat{\sigma}_C^2$ は，いずれも， $N \rightarrow \infty$ で期待値は真値に一致し，分散は0になることから，一致推定量である．ここで理論的に求めた $\text{Var}[\hat{\mu}]$, $\text{Var}[\hat{\sigma}_A^2]$, $\text{Var}[\hat{\sigma}_B^2]$, $\text{Var}[\hat{\sigma}_C^2]$ から算出される $E[\hat{\mu}] \pm \sqrt{\text{Var}[\hat{\mu}]}$, $E[\hat{\sigma}_B^2] \pm \sqrt{\text{Var}[\hat{\sigma}_B^2]}$, $E[\hat{\sigma}_B^2] \pm \sqrt{\text{Var}[\hat{\sigma}_B^2]}$, $E[\hat{\sigma}_C^2] \pm \sqrt{\text{Var}[\hat{\sigma}_C^2]}$ を図 2.1(a) ~ (d) に点線で重ねて示した．いずれの推定量についても，数値実験の結果と理論値が良く一致していることがわかる．

[問 1] 2 種類のアルファベット a, b をそれぞれ確率 r , $1-r$ で出力する独立生起情報源を考えよう．この情報源から， N 個の記号からなる記号列を観測し，この記号列に含まれていたアルファベット a の数が N_a であったものとしよう．このとき，この情報源のエントロピーを

$$\hat{H} = \text{ent}(N_a/N) \quad (2.26)$$

として推定するものとする．ただし， $\text{ent}(x)$ はエントロピー関数であり，

$$\text{ent}(x) \equiv -x \log_2 x - (1-x) \log_2 (1-x), \quad 0 \leq x \leq 1$$

として定義される．この情報源の真のエントロピーは， $\text{ent}(r)$ であるが，式 (2.26) で与えられる推定量が不偏性を持つかを調べてみよう．不偏推定量ではない場合，不偏推定量を求めてみよう．

[解答] N_a は 2 項分布 $B_{r,N}(N_a) = \binom{N}{N_a} r^{N_a} (1-r)^{N-N_a}$ に従って観測される一つの実現値であるから， $q \equiv N_a/N$ とおき， $p_Q(q) \equiv B_{r,N}(qN)$, $q = 0, 1/N, 2/N, \dots, 1$ に従う確率変数を Q で表すものとする．2 項分布の性質より，

$$E[Q] = r, \quad E[(Q - E[Q])^2] = E[(Q - r)^2] = \frac{r(1-r)}{N} \quad (2.27)$$

である．式 (2.26) で与えられるエントロピーの推定値の期待値は，

$$E[\hat{H}] = E[\text{ent}(Q)] = \sum_{N_a=0}^N \text{ent}(N_a/N) p_Q(N_a/N) \quad (2.28)$$

となる．上式を直接計算するのは大変であるから，エントロピー関数 $\text{ent}(q)$ を $q = r$ の周りでテイラー展開してみよう．エントロピー関数 $\text{ent}(q)$ の $q = r$ の周りのテイラー展開は，

$$\text{ent}(q) = \text{ent}(r) + \text{ent}'(r)(q-r) + \frac{1}{2}\text{ent}''(r)(q-r)^2 + \frac{1}{3!}\text{ent}'''(r)(q-r)^3 + \dots$$

となる．ただし，

$$\text{ent}'(r) = -\log_2 r + \log_2 (1-r), \quad \text{ent}''(r) = \frac{1}{\log 2} \frac{-1}{r(1-r)}, \dots$$

である．これを式 (2.28) に代入すれば，

$$E[\hat{H}] = \text{ent}(r) + \text{ent}'(r)E[Q-r] + \frac{1}{2}\text{ent}''(r)E[(Q-r)^2] + \frac{1}{3!}\text{ent}'''(r)E[(Q-r)^3] + \dots$$

となり，式 (2.27) を代入し，2 次までのテイラー展開で打ち切るものとするれば，

$$E[\hat{H}] = \text{ent}(r) - \frac{1}{2N \log 2}$$

となる．つまり，式 (2.26) で与えられるエントロピーの推定量は，漸近的な不偏性を持つが不偏推定量ではない．不偏推定量ではないので，もちろん有効推定量でもない．そこで，新たに

$$\hat{H} = \text{ent}(N_a/N) + \frac{1}{2N \log 2} \quad (2.29)$$

としてエントロピーを推定すれば，不偏推定量になる．

[問 2] M 種類のアルファベット $a_i, i = 1, \dots, M$ をそれぞれ確率 $r_i, i = 1, \dots, M$ で出力する独立生起情報源を考えよう．この情報源から， N 個の記号からなる記号列を観測し，この記号列に含まれていたアルファベット $a_i, i = 1, \dots, M$ の個数が $N_i, i = 1, \dots, M$ であったものとしよう．このとき，この情報源のエントロピーを

$$\hat{H} = \text{ent}\left(\frac{N_1}{N}, \dots, \frac{N_M}{N}\right) \quad (2.30)$$

として推定するものとする．ただし， $\text{ent}(x_1, \dots, x_M)$ は多変数に拡張されたエントロピー関数であり，

$$\text{ent}(x_1, \dots, x_M) \equiv - \sum_{j=1}^M x_j \log_2(x_j)$$

として定義される．この情報源の真のエントロピーは， $\text{ent}(r_1, \dots, r_M)$ であるが，式 (2.30) で与えられる推定量が不偏性を持つかを調べてみよう．不偏推定量ではない場合，不偏推定量を求めてみよう．

[解答] 各アルファベット $a_i, i = 1, \dots, M$ が個数 $N_i, i = 1, \dots, M$ ずつ観測される確率 $p(N_1, \dots, N_M)$ は，

$$p(N_1, \dots, N_M) = p(N_1)p(N_2|N_1) \cdots p(N_M|N_1, \dots, N_{M-1})$$

で与えられる． N が十分大きいものと仮定し，

$$p(N_1, \dots, N_M) \simeq p_1(N_1)p_2(N_2) \cdots p_M(N_M)$$

で近似する． $p_i(N_i) = \binom{N}{N_i} r_i^{N_i} (1-r_i)^{N-N_i}$ であるが， N が十分大きい場合には，大数の法則に従い，平均 Nr_i ，分散 $Nr_i(1-r_i)$ の正規分布で表されるようになる．したがって， $p(N_1, \dots, N_M)$ は，平均 $\mu = (Nr_1, \dots, Nr_M)$ ，分散共分散行列 $C = \text{diag}(Nr_1(1-r_1), \dots, Nr_M(1-r_M))$ を持つ多次元正規分布で表されることになる．ここで， $q_i \equiv N_i/N$ とおけば， $p(q_1, \dots, q_M)$ は，平均 $\mu = (r_1, \dots, r_M)$ ，分散共分散行列 $C = \text{diag}\left(\frac{r_1(1-r_1)}{N}, \dots, \frac{r_M(1-r_M)}{N}\right)$ を持つ多次元正規分布で表される．

$\text{ent}(q_1, \dots, q_M)$ の $q_j = r_j, j = 1, \dots, M$ の周りのテイラー展開は，

$$\begin{aligned} \text{ent}(q_1, \dots, q_M) &= \text{ent}(r_1, \dots, r_M) + \sum_{j=1}^M \frac{\partial \text{ent}(r_1, \dots, r_M)}{\partial r_j} (q_j - r_j) \\ &\quad + \sum_{j=1}^M \sum_{k=1}^M \frac{1}{2} \frac{\partial^2 \text{ent}(r_1, \dots, r_M)}{\partial r_j \partial r_k} (q_j - r_j)(q_k - r_k) + \cdots \end{aligned} \quad (2.31)$$

である．ただし，

$$\begin{aligned} \frac{\partial \text{ent}(r_1, \dots, r_M)}{\partial r_j} &= -\log_2 r_j - \frac{1}{\log 2} \\ \frac{\partial^2 \text{ent}(r_1, \dots, r_M)}{\partial r_j \partial r_k} &= \begin{cases} -\frac{1}{r_j \log 2}, & j = k \\ 0, & j \neq k \end{cases} \end{aligned}$$

である．したがって，式 (2.30) で与えられるエントロピーの推定値の期待値は，

$$E[\hat{H}] = \text{ent}(r_1, \dots, r_M) - \frac{M-1}{2N \log 2} \quad (2.32)$$

と近似できる．つまり，式 (2.30) で与えられるエントロピーの推定値は，漸近的に不偏推定量となるが，不偏推定量ではない．不偏推定量ではないので，もちろん有効推定量でもない．そこで，新たに

$$\hat{H} = \text{ent}\left(\frac{N_1}{N}, \dots, \frac{N_M}{N}\right) + \frac{M-1}{2N \log 2} \quad (2.33)$$

としてエントロピーを推定すれば，不偏推定量になる．

2.2 最尤推定

前節では，与えられた推定量の推定の良さを測る 3 種類の尺度を紹介してきた．それでは逆に，ある統計量を推定する場合，どのように推定すれば，すなわちどのように推定量を定めれば，最ももらしい推定が可能になるのだろうか．次に，こうした問題について考えて行こう．

ある確率密度分布 $p(x)$ が統計量 θ により完全に定められるとする．例えば，正規分布の場合，統計量 θ は平均（平均ベクトル）と分散（分散共分散行列）から構成される．こうした確率密度分布 $p(x)$ に従うことを前提とすれば，統計量 θ と標本 x を同列に扱うことが可能になる．すなわち，確率密度分布 $p(x)$ は，統計量が θ である条件の下で， x が生起する条件付確率 $p(x|\theta)$ とみなされる．一方， x が実際に生起した場合，統計量が θ である確率は，条件付き確率 $p(\theta|x)$ とみなすことができる．

統計量 θ を推定する場合，標本 x をすでに観測しているはずであるから，条件付確率 $p(\theta|x)$ を最大にする θ が推定量として尤もらしいと考えられる．このように推定量を

$$\hat{\theta} = \arg \max_{\theta} p(\theta|x)$$

とする推定を最大事後確率推定 (Maximum a posteriori probability estimation) という．条件付き確率 $p(\theta|x)$ は， x が生起する確率 $p(x)$ ，統計量が θ である確率 $p(\theta)$ を用いて，

$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{p(x)}$$

と書ける．上式において， $p(x)$ は θ には依存しないことから無視できる．また， θ に関する事前知識が存在しない場合には $p(\theta)$ を与えることができないので， $p(\theta) = \text{constant}$, $\forall \theta$ と仮定せざるを得ない．つまり，この場合，事後確率 $p(\theta|x)$ を最大にすることは， $p(x|\theta)$ を最大にすることに等しくなる．すなわち， θ を

$$\hat{\theta} = \arg \max_{\theta} p(x|\theta)$$

として推定することになる．こうした推定を最尤推定 (Maximum likelihood estimation)， $\hat{\theta}$ を最尤推定量 (Maximum likelihood estimator) という． $p(x|\theta)$ は， x が具体的に観測された場合，統計量 θ の関数と見ることができるので， $l(\theta)$ と記述し，尤度関数 (Likelihood function) と呼ぶ．対数は単調増加関数なので，尤度関数の最大値を与える θ は，対数をとった尤度関数を最大にする θ に等しい．したがって，尤度関数の代わりに， $L(\theta) \equiv \log l(\theta)$ を用いることも多い．これを対数尤度 (Log likelihood) という．

正規分布に従う互いに独立な N 個の標本 x_1, x_2, \dots, x_N が観測されたものとし，これらの平均と分散の最尤推定量を求めてみよう．各 x_i が観測される確率は，

$$p(x_i|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right), \quad i = 1, \dots, N \quad (2.34)$$

であり，各標本が独立であることから， $\boldsymbol{x} = \{x_1, \dots, x_N\}$ が観測される確率は，

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \quad (2.35)$$

となる．ただし， $\boldsymbol{\theta} = (\mu, \sigma^2)^T$ である．両辺の対数をとれば，

$$L(\boldsymbol{\theta}) = \log p(\boldsymbol{x}|\boldsymbol{\theta}) = -\frac{N}{2} \log 2\pi - \frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \quad (2.36)$$

であり， μ, σ^2 でそれぞれ偏微分して 0 とおけば，

$$\begin{cases} \frac{\partial L(\boldsymbol{\theta})}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \mu) \equiv 0 \\ \frac{\partial L(\boldsymbol{\theta})}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^N (x_i - \mu)^2 \equiv 0 \end{cases} \quad (2.37)$$

となる．この連立方程式を満たす $\boldsymbol{\theta} = (\mu, \sigma^2)^T$ が最尤推定量 $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}^2)^T$ であり，連立方程式を解けば次式となる．

$$\begin{cases} \hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i \\ \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2 \end{cases} \quad (2.38)$$

これより，式 (2.5) で示した平均の推定量 μ は，最尤推定量であり，式 (2.7) で示した分散の推定量 σ_B^2 は，不偏性を持たないが，最尤推定量であることがわかる．

2.3 推定量の分散の下限

最尤推定量が，尤もらしさの意味で最も適切な推定量であった．さて，ある標本数の与えられた標本に対し，推定量の推定精度の限界，つまり，推定量の分散をどこまで下げることが可能なのだろうか．次に，こうした疑問に答えるため，不偏推定量に限定し，推定量の分散の下限を導いてゆこう．

ある確率密度分布が統計量 $\boldsymbol{\theta}$ で完全に定められるとし，その統計量が具体的に $\boldsymbol{\theta}$ であることを知ったとき， \boldsymbol{x} が観測される確率は条件付確率 $p(\boldsymbol{x}|\boldsymbol{\theta})$ で与えられる．確率密度分布の性質より，

$$\int p(\boldsymbol{x}|\boldsymbol{\theta}) d\boldsymbol{x} = 1 \quad (2.39)$$

であるから，上式の両辺を $\boldsymbol{\theta}$ で偏微分することにより，

$$\int \frac{\partial p(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} d\boldsymbol{x} = \mathbf{0} \quad (2.40)$$

となる．上式において， $\mathbf{0}$ は零ベクトルを表す．一方，対数尤度 $\log p(\boldsymbol{x}|\boldsymbol{\theta})$ を $\boldsymbol{\theta}$ で偏微分すれば，

$$\frac{\partial \log p(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})} \frac{\partial p(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad (2.41)$$

となり，その期待値は，

$$E \left[\frac{\partial \log p(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] = E \left[\frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})} \frac{\partial p(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] = \int \frac{1}{p(\boldsymbol{x}|\boldsymbol{\theta})} \frac{\partial p(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} p(\boldsymbol{x}|\boldsymbol{\theta}) d\boldsymbol{x} = \int \frac{\partial p(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} d\boldsymbol{x} = \mathbf{0} \quad (2.42)$$

となる．上式の左から 3,5 番目の式を θ で偏微分すれば，

$$\int \left\{ \left(\frac{1}{p(\mathbf{x}|\boldsymbol{\theta})} \frac{\partial p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T + p(\mathbf{x}|\boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{1}{p(\mathbf{x}|\boldsymbol{\theta})} \frac{\partial p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \right\} dx = \mathbf{O}$$

となる．ただし，上式において， \mathbf{O} は，すべての要素が零の正方行列を表す．上式に，式 (2.41) を代入すれば，

$$\int \left\{ \left(\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T + \frac{\partial^2 \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\} p(\mathbf{x}|\boldsymbol{\theta}) dx = \mathbf{O}$$

となり，期待値を用いて表現すれば，

$$E \left[\left(\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right] = -E \left[\frac{\partial^2 \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] \equiv \mathbf{J}(\mathbf{x}, \boldsymbol{\theta}) \quad (2.43)$$

となる．ここで，行列 $\mathbf{J}(\mathbf{x}, \boldsymbol{\theta})$ は，フィッシャーの情報行列 (Fisher's information matrix) と呼ばれている．標本 \mathbf{x} を用いて推定された $\hat{\boldsymbol{\theta}}$ が統計量 $\boldsymbol{\theta}$ の不偏推定量であれば，

$$E[\hat{\boldsymbol{\theta}}] = \int \hat{\boldsymbol{\theta}} p(\mathbf{x}|\boldsymbol{\theta}) dx = \boldsymbol{\theta}$$

であることは前節で述べたとおりである．上式を $\boldsymbol{\theta}$ で偏微分し，式 (2.41) を利用すれば，

$$\int \hat{\boldsymbol{\theta}} \left(\frac{\partial p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T dx = \int \hat{\boldsymbol{\theta}} \left(\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T p(\mathbf{x}|\boldsymbol{\theta}) dx = \mathbf{I} \quad (2.44)$$

となる．ただし，上式において， \mathbf{I} は，単位行列である．式 (2.42) より，

$$E \left[\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] = \int \frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} p(\mathbf{x}|\boldsymbol{\theta}) dx = \mathbf{0}$$

であり， \mathbf{x} に依存しない $\boldsymbol{\theta}$ を両辺に乗じたとしても

$$\boldsymbol{\theta} \int \left(\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T p(\mathbf{x}|\boldsymbol{\theta}) dx = \int \boldsymbol{\theta} \left(\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T p(\mathbf{x}|\boldsymbol{\theta}) dx = \mathbf{O} \quad (2.45)$$

となる．式 (2.44) から式 (2.45) を減じることにより，

$$\int (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \left(\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T p(\mathbf{x}|\boldsymbol{\theta}) dx = \mathbf{I}$$

が得られる．これは，期待値を用いて表現すると，

$$E \left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \left(\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right] = \mathbf{I} \quad (2.46)$$

となる．

ところで，平均 $\mathbf{0}$ の確率変数ベクトル \mathbf{x}, \mathbf{y} に対して，

$$\sqrt{E[\mathbf{x}\mathbf{x}^T]E[\mathbf{y}\mathbf{y}^T]} \geq E[\mathbf{x}\mathbf{y}^T]$$

が成立し， $E[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] = \mathbf{0}$ ， $E \left[\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] = \mathbf{0}$ であるから，式 (2.46) から，

$$E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] E \left[\left(\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right] \geq \mathbf{I}$$

が導かれる．ここで，

$$\text{Cov}[\hat{\boldsymbol{\theta}}] = E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T]$$

とおけば，上の不等式は，

$$\text{Cov}[\hat{\boldsymbol{\theta}}] \geq \mathbf{J}(\mathbf{x}, \boldsymbol{\theta})^{-1} \quad (2.47)$$

となる．式 (2.47) 中の不等号は，行列の各要素について不等号が成立することを意味するのではなく，任意のベクトル \mathbf{u} に対し，

$$\mathbf{u}^T \text{Cov}[\hat{\boldsymbol{\theta}}] \mathbf{u} \geq \mathbf{u}^T \mathbf{J}(\mathbf{x}, \boldsymbol{\theta})^{-1} \mathbf{u}$$

となることを意味している． \mathbf{u} を単位ベクトルとすることにより，

$$\text{Cov}[\hat{\boldsymbol{\theta}}]_{k,k} \geq \mathbf{J}(\mathbf{x}, \boldsymbol{\theta})_{k,k}^{-1}, \quad \forall k$$

が導かれる．ただし， $A_{k,k}$ は，行列 A の (k, k) 要素を表す．つまり，式 (2.47) では，行列の対角要素，すなわち推定量の分散については，要素ごとに不等号が成立する．

不等式 (2.47) は，クラメル・ラオの不等式 (Cramér-Rao's inequality) と呼ばれ，不偏推定量 $\hat{\boldsymbol{\theta}}$ の分散，共分散の下界を与えている．等号が成り立つ場合， $\hat{\boldsymbol{\theta}}$ は有効推定量といわれる．

$\hat{\boldsymbol{\theta}}$ が最尤推定量ならば，標本数 $N \rightarrow \infty$ に対し， $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ が平均 $\mathbf{0}$ ， $\mathbf{J}(\mathbf{x}, \boldsymbol{\theta})^{-1}$ を分散共分散行列に持つ正規分布に漸近するという重要な性質があることが証明されている．このことは，最尤推定量が漸近的に不偏であり，なおかつ有効性を持つことを表している．

N 個の標本 x_1, x_2, \dots, x_N が与えられ，これらの平均と分散を推定する際のクラメル・ラオの下界を求めてみよう．各標本は，平均 μ ，分散 σ^2 の正規分布に従い，かつ互いに独立であるものとする．式 (2.36) で与えた対数尤度 $L(\boldsymbol{\theta})$ を μ, σ^2 で 2 階微分すれば，

$$\begin{cases} \frac{\partial^2 L(\boldsymbol{\theta})}{(\partial \mu)^2} = -\frac{N}{\sigma^2} \\ \frac{\partial^2 L(\boldsymbol{\theta})}{(\partial \sigma^2)^2} = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^N (x_i - \mu)^2 \\ \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \mu \partial \sigma^2} = -\frac{1}{\sigma^4} \sum_{i=1}^N (x_i - \mu) \end{cases}$$

となる．さらにそれぞれ期待値をとり， -1 を乗ずることにより，

$$\mathbf{J}(\mathbf{x}, \boldsymbol{\theta}) = \begin{pmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{pmatrix}$$

となる．したがって，Cramér-Rao の下界は，その逆行列

$$\mathbf{J}(\mathbf{x}, \boldsymbol{\theta})^{-1} = \frac{1}{N} \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}$$

となる．つまり，平均 μ ，分散 σ^2 の正規分布に従い，かつ互いに独立である N 個の標本からは，平均と分散をそれぞれ $\frac{\sigma^2}{N}$ ， $\frac{2\sigma^4}{N}$ よりも小さい分散で，不偏性をもって推定することは不可能であることになる．

以上の結果を演習 1 で得られた結果と比較してみよう．不偏推定量 $\hat{\mu}$ は，最尤推定量でもあり，その推定分散は $\frac{\sigma^2}{N}$ であった．クラメル・ラオの下界 $\frac{\sigma^2}{N}$ を達成していることから，不偏推定量 $\hat{\mu}$ は有効推定量であることになる．推定量 $\hat{\sigma}_A^2$ は，平均の真値が既知であるから除外するものとして，推定量 $\hat{\sigma}_B^2$ ， $\hat{\sigma}_C^2$ の推定分散は，それぞれ， $\frac{N-1}{N} \frac{1}{N} 2\sigma^4$ ， $\frac{1}{N-1} 2\sigma^4$ であった．推定量 $\hat{\sigma}_B^2$ の推定分散は，クラメル・ラオの下界 $\frac{2\sigma^4}{N}$ よりも小さいが，これは，クラメル・ラオの下界は，不偏推定量に対する下界であるのに対し，推定量

$\hat{\sigma}_B^2$ が不偏性を持たないからである．こうした意味では，クラメル・ラオの下界は推定分散の下界としては少々あまい下界である．また，不偏推定量 $\hat{\sigma}_C^2$ の推定分散 $\frac{1}{N-1}2\sigma^4$ は，クラメル・ラオの下界 $\frac{2\sigma^4}{N}$ よりも大きくなっている．推定量 $\hat{\sigma}_B^2$ は，最尤推定量であり， $N \rightarrow \infty$ での推定分散は， $\frac{N-1}{N} \frac{1}{N} 2\sigma^4 \simeq \frac{1}{N} 2\sigma^4$ となり，クラメル・ラオの下界に漸近してゆく．

2.4 EM アルゴリズム

統計量 θ で規定される確率密度分布 $p(x|\theta)$ において， x が $x \equiv (y, z)$ と表され， y のみを観測できるものとする．本来，統計量 θ はベクトル，変数 x, y, z は確率変数ベクトルであっても構わないので，太字で書くべきであるが，混同する恐れのない限り，このように書くものとする．最尤推定量は， $p(y, z|\theta)$ の y についての周辺分布 $p(y|\theta) = \int p(y, z|\theta) dz$ に対する対数尤度 $L(y|\theta) = \log p(y|\theta)$ の最大値として与えられる． $L(y|\theta)$ を不完全データに対する対数尤度という．これに対し，結合確率 $p(y, z|\theta)$ に対する対数尤度 $L(y, z|\theta) = \log p(y, z|\theta)$ を完全データに対する対数尤度という．不完全データに対する対数尤度の最大化は一般に困難であるため，こうした最大化を反復的に行う手段として，EM アルゴリズム (Expectation-Maximization アルゴリズム) が考案された． k 回めの反復において得られた θ の推定値を $\hat{\theta}^{(k)}$ とすれば，

$$E_{z|y; \hat{\theta}^{(k)}} [L(y, z|\theta)] \equiv \int L(y, z|\theta) p(z|y; \hat{\theta}^{(k)}) dz$$

は，完全データに対する対数尤度 $L(y, z|\theta)$ の $y, \hat{\theta}^{(k)}$ が与えられた条件の下での条件付期待値を表す．この条件付期待値は， $\theta, \hat{\theta}^{(k)}, y$ の関数であり，そのうち， $\hat{\theta}^{(k)}, y$ は与えられている．そこでこの条件付期待値を最大にする θ を求め，それを $\hat{\theta}^{(k+1)}$ とする．すなわち，

$$\hat{\theta}^{(k+1)} = \arg \max_{\theta} E_{z|y; \hat{\theta}^{(k)}} [L(y, z|\theta)]$$

とする．こうした条件付期待値の計算とその最大化を反復して行う手法が EM アルゴリズムである．

ところで， $r(x)$ が x の関数を要素に持つ列ベクトル， $\psi(\theta)$ が θ の関数を表すものとして，次式で表される確率密度分布を持つものを指数型分布族という．ただし，分布のパラメータ θ は，元々のパラメータから変換して得られるものでも構わない．

$$p(x|\theta) = \exp\left(\theta^T r(x) - \psi(\theta)\right)$$

$p(x|\theta)$ がこうした指数型分布族に属する場合，EM アルゴリズムにより得られる $\hat{\theta}^{(k+1)}$ と $\hat{\theta}^{(k)}$ の間には，近似

$$\hat{\theta}^{(k+1)} \simeq \hat{\theta}^{(k)} + J^{-1}(x, \hat{\theta}^{(k)}) \left. \frac{\partial L(y|\theta)}{\partial \theta} \right|_{\theta = \hat{\theta}^{(k)}}$$

が成立することが知られている．ただし， $J(x, \hat{\theta}^{(k)})$ は，Fisher の情報行列である．こうした反復において，

$$L(y|\hat{\theta}^{(k+1)}) \geq L(y|\hat{\theta}^{(k)}), \quad \forall k$$

となり， $E_{z|y; \hat{\theta}^{(k)}} [L(y, z|\theta)]$ が， θ と $\hat{\theta}^{(k+1)}$ について連続であれば， $\hat{\theta}^{(k)}$ は， $k \rightarrow \infty$ で， $L(y|\hat{\theta}^{(k)})$ の極大値を与える点に収束することが保証されている．

EM アルゴリズムの応用例を一つ示そう．複数の情報源 (確率変数) Y_1, \dots, Y_M があり，そこから，それぞれ確率 ξ_1, \dots, ξ_M で選ばれた一つの情報源から標本 y を出力する情報源 Y を考えよう．個々の情報源 Y_i

が確率密度分布 $p_i(y; \theta_i)$ に従うものとする、情報源 Y の従う確率密度分布 $p(y; \theta)$ は、

$$p(y; \theta) = \sum_{i=1}^M \xi_i p_i(y; \theta_i) \quad (2.48)$$

と表される。ただし、 $\sum_{i=1}^M \xi_i = 1$ である。一般に、こうした分布 $p(y; \theta)$ を (有限) 混合分布という。さて、観測された標本 y_1, \dots, y_N から、この分布のパラメータ $\theta = (\theta_1, \dots, \theta_M, \xi_1, \dots, \xi_M)$ を最尤推定したいのであるが、これは一般には困難な問題である。仮に、観測された標本がそれぞれの情報源から出力されているのが分かれば、それぞれの情報源ごとに分布のパラメータを最尤推定することが可能になる。そこで、観測される変数 y に加え、それがどの情報源からの出力であるかという観測不可能な変数 $z \in \{1, \dots, M\}$ があり、元々の変数が $x = (y, z)$ であると考えよう。これにより、EM アルゴリズムを適用することが可能になる。

分布のパラメータが θ である時、標本 y が観測され、かつそれが情報源 Y_z からのものである確率は、

$$p(y, z; \theta) = \xi_z p_z(y; \theta_z)$$

であり、完全データに対する対数尤度は、

$$L(y, z|\theta) = \log \xi_z p_z(y; \theta_z)$$

である。一方、分布のパラメータが $\hat{\theta}^{(k)}$ 、かつ標本 y を観測した条件の下で、それが情報源 Y_z からのものである条件付確率は、

$$p(z|y; \hat{\theta}^{(k)}) = \frac{\xi_z p_z(y; \hat{\theta}_z^{(k)})}{\sum_{i=1}^M \xi_i p_i(y; \hat{\theta}_i^{(k)})}$$

である。完全データに対する対数尤度 $L(y, z|\theta)$ の $y, \hat{\theta}^{(k)}$ を条件とする z に関する条件付期待値は、

$$E_{z|y; \hat{\theta}^{(k)}} [L(y, z|\theta)] \equiv \sum_{z=1}^M L(y, z|\theta) p(z|y; \hat{\theta}^{(k)})$$

となる。また、複数の標本 y_1, \dots, y_N が存在する場合には、対数尤度 $L(y_j, z|\theta)$ の $y_j, \hat{\theta}^{(k)}$ を条件とする z に関する条件付期待値の各標本についての和

$$E_{z|y; \hat{\theta}^{(k)}} [L(y, z|\theta)] \equiv \sum_{j=1}^N \sum_{z=1}^M L(y_j, z|\theta) p(z|y_j; \hat{\theta}^{(k)})$$

を用いる。

次に、この条件付期待値 $E_{z|y; \hat{\theta}^{(k)}} [L(y, z|\theta)]$ を最大にする分布のパラメータ θ を求めよう。 $\sum_{i=1}^M \xi_i = 1$ なる制約があるから、ラグランジェの未定乗数法を用いる必要がある。また、式の簡単化のため、

$$w_{zj} \equiv p(z|y_j; \hat{\theta}^{(k)})$$

とおく。

$$\begin{aligned} J &= E_{z|y; \hat{\theta}^{(k)}} [L(y, z|\theta)] - \lambda \left(\sum_{i=1}^M \xi_i - 1 \right) \\ &= \sum_{j=1}^N \sum_{z=1}^M \log(\xi_z p_z(y_j; \theta_z)) w_{zj} - \lambda \left(\sum_{i=1}^M \xi_i - 1 \right) \\ &= \sum_{j=1}^N \sum_{z=1}^M (w_{zj} \log \xi_z + w_{zj} \log p_z(y_j; \theta_z)) - \lambda \left(\sum_{i=1}^M \xi_i - 1 \right) \end{aligned} \quad (2.49)$$

とにおいて、この J を ξ_m で偏微分すると、

$$\frac{\partial J}{\partial \xi_m} = \frac{1}{\xi_m} \sum_{j=1}^N w_{mj} - \lambda$$

となり、 $\frac{\partial J}{\partial \xi_m} \equiv 0$ とおけば、

$$\xi_m = \frac{1}{\lambda} \sum_{j=1}^N w_{mj}$$

が得られる。また、 $\sum_{z=1}^M w_{zj} = 1$ なので、 $\sum_{j=1}^N \sum_{z=1}^M w_{zj} = N$ である。これを利用すれば、 $\lambda = N$ であることが分かるので、結局、条件付期待値 $E_{z|y; \hat{\theta}^{(k)}}[L(y, z|\theta)]$ を最大にする ξ_m は、

$$\xi_m = \frac{1}{N} \sum_{j=1}^N w_{mj} \quad (2.50)$$

となる。一方、 J を θ_m で偏微分すると、

$$\frac{\partial J}{\partial \theta_m} = \sum_{j=1}^N w_{mj} \frac{\partial L_m(y_j|\theta_m)}{\partial \theta_m}$$

となる。ただし、 $L_z(y_j|\theta_z) \equiv \log p_z(y_j|\theta_z)$ は、個々の分布の対数尤度を表す。すなわち、 ξ_1, \dots, ξ_M 以外の分布のパラメータ $\theta_1, \dots, \theta_M$ については、分布 $p_m(y|\theta_m)$ ごとに、個々の標本に対する対数尤度を w_{mj} で重み付けた後、最尤推定を行えばよいことが分かる。

ところで、 $p_i(y|\theta_i)$ がいずれも正規分布である場合には、 $p(y|\theta)$ は、混合正規分布 (Gaussian mixture distribution) と呼ばれ、以下のように表される。

$$p(y|\theta) = \sum_{i=1}^M \xi_i \phi(y|\mu_i, \sigma_i^2) \quad (2.51)$$

ただし、 $\phi(y|\mu_i, \sigma_i^2)$ は、平均 μ_i 、分散 σ_i^2 の正規分布を表す。混合正規分布のパラメータは、

$$\theta = (\xi_1, \dots, \xi_M, \mu_1, \dots, \mu_M, \sigma_1^2, \dots, \sigma_M^2)$$

となる。混合正規分布の場合、先ほどの個々の分布に対する最尤推定問題は、正規分布に対する最尤推定量を求める問題であるから、非常に簡単になる。

$$p(y_j, z|\theta) = \xi_z \phi(y_j|\mu_z, \sigma_z^2)$$

であるから、先ほどと同様に

$$\begin{aligned} J &= \sum_{j=1}^N \sum_{z=1}^M \log(\xi_z \phi(y_j|\mu_z, \sigma_z^2)) w_{zj} - \lambda \left(\sum_{i=1}^M \xi_i - 1 \right) \\ &= \sum_{j=1}^N \sum_{z=1}^M \left(\log \xi_z - \frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma_z^2 - \frac{1}{2\sigma_z^2} (y_j - \mu_z)^2 \right) w_{zj} - \lambda \left(\sum_{i=1}^M \xi_i - 1 \right) \end{aligned} \quad (2.52)$$

を μ_m で偏微分すれば、

$$\frac{\partial J}{\partial \mu_m} = \frac{1}{\sigma_m^2} \sum_{j=1}^N (y_j - \mu_m) w_{mj}$$

となり，これを 0 とおき，式 (2.50) を利用すれば，

$$\mu_m = \frac{1}{N\xi_m} \sum_{j=1}^N y_j w_{mj} \quad (2.53)$$

となる．同様に， J を σ_m^2 で偏微分すれば，

$$\frac{\partial J}{\partial \sigma_m^2} = \sum_{j=1}^N \left(-\frac{1}{2\sigma_m^2} + \frac{1}{2\sigma_m^4} (y_j - \mu_m)^2 \right) w_{mj}$$

となり，これを 0 とおき，式 (2.50) を利用すれば，

$$\sigma_m^2 = \frac{1}{N\xi_m} \sum_{j=1}^N (y_j - \mu_m)^2 w_{mj} = \frac{1}{N\xi_m} \sum_{j=1}^N y_j^2 w_{mj} - \mu_m^2 \quad (2.54)$$

となる．

式 (2.50)，式 (2.53)，式 (2.54) をまとめると，多重正規分布のパラメータを最尤推定するための，EM アルゴリズムの各ステップは，以下のように書くことができる． $\hat{\theta}^{(0)}$ を適当な値に初期化した後，各反復 k 毎に，以下の計算を行う．

$$\left\{ \begin{array}{l} w_{zj}^{(k)} = p(z|y_j; \hat{\theta}^{(k)}) = \frac{\hat{\xi}_z^{(k)} \phi(y_j; \hat{\mu}_z^{(k)}, \hat{\sigma}_z^2{}^{(k)})}{\sum_{i=1}^M \hat{\xi}_i^{(k)} \phi(y_j; \hat{\mu}_i^{(k)}, \hat{\sigma}_i^2{}^{(k)})} \\ \hat{\xi}_z^{(k+1)} = \frac{1}{N} \sum_{j=1}^N w_{zj}^{(k)}, \\ \hat{\mu}_z^{(k+1)} = \frac{1}{N\hat{\xi}_z^{(k+1)}} \sum_{j=1}^N w_{zj}^{(k)} y_j, \\ \hat{\sigma}_z^2{}^{(k+1)} = \frac{1}{N\hat{\xi}_z^{(k+1)}} \sum_{j=1}^N w_{zj}^{(k)} y_j^2 - (\hat{\mu}_z^{(k+1)})^2, \\ j = 1, \dots, N, \quad z = 1, \dots, M \end{array} \right. \quad (2.55)$$

[演習 2] 正規分布 $N(0, 0.2)$, $N(3, 1)$, $N(6, 0.4)$ にそれぞれ従う情報源 (確率変数) Y_1, Y_2, Y_3 があり，そこから，それぞれ確率 0.5, 0.3, 0.2 で選ばれた一つの情報源から標本 $y_i, i = 1, \dots, N$ が出力されるものとしよう．標本が $N = 1000$ 個与えられたものとして，分布のパラメータを EM アルゴリズムにより最尤推定してみよう．EM アルゴリズムは，対数尤度の極大値に収束することを保証しても，それが最大値である保証はないので，パラメータの初期値を変えて何回か行い (初期値試行)，最終的に不完全データに対する対数尤度 $L(y|\hat{\theta})$ が最大になったものを選択せよ．

[解答] 作成した標本 $y_i, i = 1, \dots, N$ の頻度分布を図 2.2(a) に示す．問題を易しくするため，3 つの情報源の平均値が互いに離れているように設定したので，これら 3 つの分布を視覚的にも確認できる．この標本に対し，EM アルゴリズムを適用して分布のパラメータ $\theta = (\xi_1, \xi_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ を推定した．反復 $k, k = 1, \dots, 100$ での推定値 $\theta^{(k)}$ ，その際の不完全データに対する対数尤度 $L(y|\theta^{(k)})$ をそれぞれ同図 (b), (d) に示す．対数尤度は EM アルゴリズムの各更新ごとに，単調に増加し，対数尤度，分布のパラメータ共に反復回数 30 回ほどで収束することが分かる．最終的に得られたパラメータ $\hat{\theta}^{(100)}$ を持つ確率密度分布 $\xi_i \phi(y; \hat{\mu}_i, \hat{\sigma}_i^2), i = 1, 2, 3$ を同図 (c) に示す．もともとの情報源の確率密度分布をよく同定できていることがわかる．

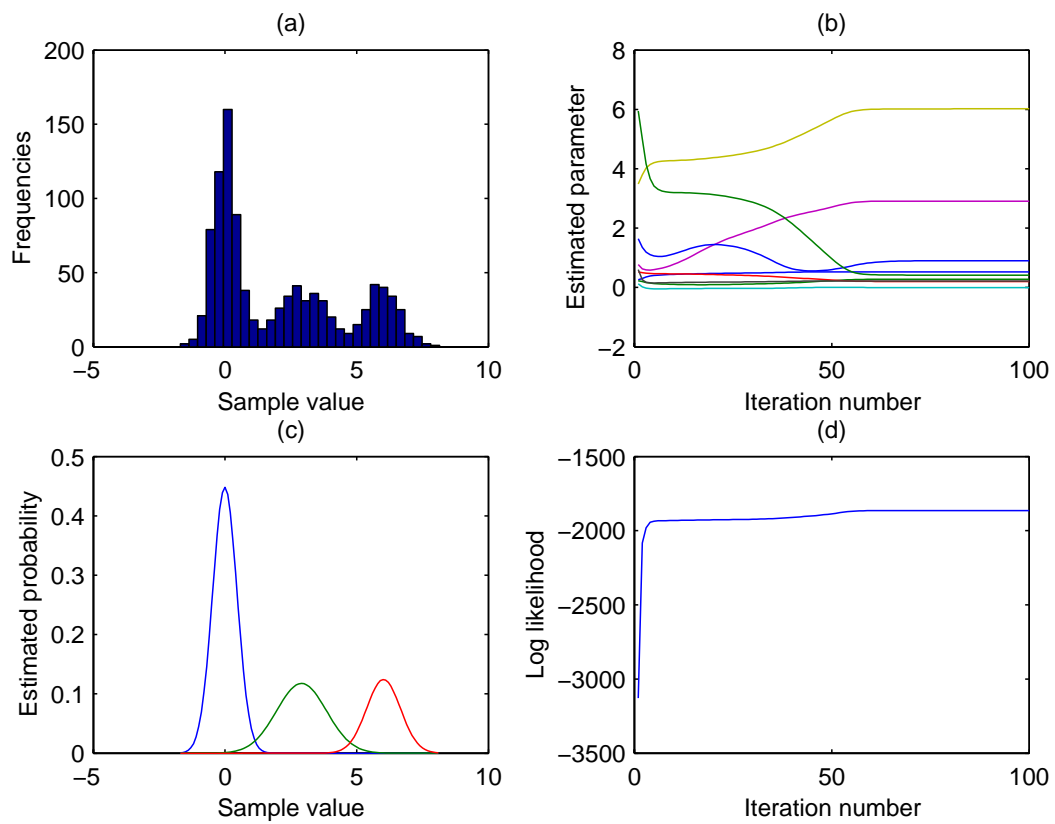


図 2.2: 演習 2 の解答 : EM アルゴリズムによる混合正規分布のモデルパラメータの推定

第3章 確率過程

3.1 確率過程

時刻 t を変数に持つ確率変数 $x(t)$ を確率過程 (stochastic process) という。時刻 t が実数値をとるならば、 $x(t)$ は連続時間確率過程 (continuous time stochastic process)、離散値をとるならば、離散時間確率過程 (discrete time stochastic process)、あるいは確率系列 (stochastic series) といわれる。一般的には、確率系列 $x(t)$ は、 t が整数値をとるものとして定義され、この場合、確率系列 $x(t)$ を x_t と書くことがある。

確率過程 $x(t)$ のすべての統計的性質は、時刻 t_1, t_2, \dots で $x(t)$ がそれぞれ値 x_1, x_2, \dots をとる結合確率密度分布で規定される。確率過程の場合、 t は実数値をとるから、その統計的性質を完全に表現するためには、本来、無限個の変数を持つ結合確率密度分布を示す必要がある。これは、現実的ではないので、以下に示すように、一つ、もしくは二つの時刻での統計的性質に着目することが多い。例えば、確率過程 $x(t)$ が時刻 t で確率密度分布 $p(x; t) \equiv \frac{d}{dx}P(x(t) \leq x)$ に従うとすれば、 $x(t)$ の平均、分散、 n 次モーメント、 n 次中心モーメントは、それぞれ次式で定義されるように、時刻 t を変数に持つことになる。

$$\begin{aligned}\mu_x(t) &= E[x(t)] = \int_{-\infty}^{\infty} xp(x; t)dx \\ \sigma_x^2(t) &= E[(x(t) - \mu(t))^2] = \int_{-\infty}^{\infty} (x - \mu(t))^2 p(x; t)dx \\ m_x^k(t) &= E[x^k(t)] = \int_{-\infty}^{\infty} x^k p(x; t)dx \\ C_x^k(t) &= E[(x(t) - \mu(t))^k] = \int_{-\infty}^{\infty} (x - \mu(t))^k p(x; t)dx\end{aligned}$$

また、異なる2つの時刻 t_1, t_2 での $x(t)$ の統計的性質は、 $x(t_1), x(t_2)$ の結合確率密度分布 $p(x_1, x_2; t_1, t_2) \equiv \frac{d^2}{dx_1 dx_2}P(x(t_1) \leq x_1, x(t_2) \leq x_2)$ で表現される。 $x(t_1), x(t_2)$ の相関、共分散は、それぞれ

$$\begin{aligned}R_x(t_1, t_2) &= E[x(t_1)x(t_2)] = \int_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2; t_1, t_2) dx_1 dx_2 \\ C_x(t_1, t_2) &= E[(x(t_1) - \mu(t_1))(x(t_2) - \mu(t_2))] \\ &= \int_{-\infty}^{\infty} (x_1 - \mu(t_1))(x_2 - \mu(t_2)) p(x_1, x_2; t_1, t_2) dx_1 dx_2\end{aligned}$$

となる。相関 $R_x(t_1, t_2)$ 、共分散 $C_x(t_1, t_2)$ は、それぞれ $x(t)$ の自己相関関数 (autocorrelation function)、自己共分散関数 (autocovariance function) と呼ばれる。

自然界に存在する多くの確率過程の自己共分散関数 $C_x(t_1, t_2)$ は、 $|t_1 - t_2| \rightarrow \infty$ につれて減少する。そこで、 $|t_1 - t_2| > a$ であるすべての t_1, t_2 に対し、 $C_x(t_1, t_2) = 0$ となるならば、 a -従属 (a -subordination) であるという。確率過程 $x(t)$ が 0 -従属の場合、すなわち $t_i \neq t_j, i \neq j$ に対して $C_x(t_i, t_j) = 0$ であるならば、 $x(t)$ は白色雑音 (white noise) と呼ばれる。

2つの確率過程 $x(t), y(t)$ がある場合、時刻 t_1, t_2 で、 $x(t), y(t)$ がそれぞれ値 x, y をとる結合確率密度分

布は , $p_{x,y}(x, y; t_1, t_2) \equiv \frac{d^2}{dx dy} P(x(t_1) \leq x, y(t_2) \leq y)$ と定義される . $x(t_1)$ と $y(t_2)$ の間の相関

$$R_{xy}(t_1, t_2) \equiv E[x(t_1)y(t_2)]$$

共分散

$$C_{xy}(t_1, t_2) \equiv E[(x(t_1) - \mu_x(t_1))(y(t_2) - \mu_y(t_2))]$$

n 次結合モーメント

$$E[x^k(t_1)x^{n-k}(t_2)]$$

n 次結合中心モーメント

$$E[(x(t_1) - \mu_x(t_1))^k(x(t_2) - \mu_y(t_2))^{n-k}]$$

も定義できる . 相関 $R_{xy}(t_1, t_2)$ と共分散 $C_{xy}(t_1, t_2)$ は , それぞれ $x(t)$ と $y(t)$ の相互相関関数 (crosscorrelation function) , 相互共分散関数 (crosscovariance function) と呼ばれる . また , $|t_1 - t_2| > a$ であるすべての t_1, t_2 に対し , $C_{xy}(t_1, t_2) = 0$ ならば , 相関 a -従属 (correlation a -subordination) であるという .

$x(t), y(t)$ が任意の時刻 t_1, t_2 に対して ,

$$R_{xy}(t_1, t_2) = 0$$

ならば , 直交 (orthogonal) ,

$$C_{xy}(t_1, t_2) = 0$$

ならば , 無相関 (uncorrelated) であるといわれる . 更に , 任意の n, m と $t_1, \dots, t_n, t'_1, \dots, t'_m$ に対し

$$p_{x,y}(x_1, \dots, x_n, y_1, \dots, y_m; t_1, \dots, t_n, t'_1, \dots, t'_m) = p(x_1, \dots, x_n; t_1, \dots, t_n)p_y(y_1, \dots, y_m; t'_1, \dots, t'_m)$$

を満たすならば , 2 つの確率過程 $x(t), y(t)$ は , 独立 (independent) であるという .

3.2 定常過程と正規確率過程

確率過程 $x(t)$ は , そのすべての統計的性質が時間の原点の推移により変化しない場合 , すなわち , 任意の n, τ に対し

$$p(x_1, \dots, x_n; t_1, \dots, t_n) = p(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau)$$

を満たす場合 , 強定常過程 (stationary process in strict sense) , あるいは狭義定常過程という . 特定の n に対してのみ , 上式がなりたつ場合は , n 次定常過程 (stationary process of order n) という . 強定常性は非常に強い性質であることから , 実際には $n = 2$ の場合 , つまり 2 次定常である時 , 定常過程ということもある . この場合 , 正確には弱定常過程 (stationary process in weak sense) , あるいは広義定常過程という .

定常過程ではない確率過程 , つまり統計的性質が時刻 t と共に変化する信号を非定常過程 (nonstationary process) という . 非定常過程の顕著な例としては , 順応 , 疲労等の影響を受ける生体からの信号がある . こうした生体信号の解析においては , 一定区間では定常とみせる場合もあり , その際 , 区間定常 (stationary in an interval) と呼ぶこともある .

(弱) 定常過程の平均 $\mu(t)$ は , $\mu(t) = \mu(t + \tau)$, $\forall \tau$ であり , t に依存せず , 単に定数 μ で表される . また , 自己相関関数 $R_x(t_1, t_2)$ は , 2 次の統計的性質であるから , (弱) 定常過程では時間差 $\tau = t_2 - t_1$ のみに依存し , 単に

$$R_x(\tau) \equiv R_x(t_1, t_2) = R_x(t_1, t_1 + \tau) = E[x(t)x(t + \tau)]$$

と表すことができる。(弱)定常過程の自己相関関数 $R_x(t_1, t_2)$ は, $R_x(t_1 + \tau, t_2 + \tau), \forall \tau$ であるから, t_1 と t_2 の時間差のみに依存する偶関数となる。そこで, (弱)定常過程の自己相関関数を $t = t_1 - t_2$ として $R_x(t)$ と書く。自己相関関数 $R_x(t)$ の変数 t をラグ (lag) と呼ぶ。また, $R_x(0) = E[x^2(t)]$ であることから, $R_x(0)$ は $x(t)$ の分散を表す。また,

$$\begin{aligned} E[(x(t) \pm x(t + \tau))^2] &= E[x^2(t)] \pm 2E[x(t)x(t + \tau)] + E[x^2(t + \tau)] \\ &= R_x(0) \pm 2R_x(\tau) + R_x(0) \\ &= 2(R_x(0) \pm R_x(\tau)) \end{aligned}$$

であり, $E[(x(t) \pm x(t + \tau))^2] \geq 0$ であることから

$$|R_x(\tau)| \leq R_x(0)$$

が成り立つ。

確率過程 $x(t)$ の任意の時刻 t_1, \dots, t_n での結合確率密度分布 $p(x_1, \dots, x_n; t_1, \dots, t_n)$ が n 次元正規分布である場合, 確率過程 $x(t)$ は n 次正規確率過程 (Gaussian process) と言われる。任意の n に対してのみ成り立つ場合, 単に正規確率過程という。正規確率過程は, 平均, 分散, 共分散などの 2 次以下の統計量で完全に表現できる。熱雑音などは, 正規確率過程と見なされる。また, 正規確率過程は, いかなる線形変換を施しても必ず正規確率過程になるという重要な性質があるため, 正規確率過程を入力とするシステムの応答が正規性を持つことが, そのシステムが線形であることの一つの必要条件になる。

3.3 エルゴード過程

確率過程における主要な問題は, その統計量の推定である。例えば, 平均 $\mu(t) = E[x(t)]$ は, 確率過程 $x(t)$ の N 個の標本過程 $x_n(t), n = 1, \dots, N$ を観測し, 次式により推定することが考えられる。

$$\hat{\mu}(t) = \frac{1}{N} \sum_{n=1}^N x_n(t)$$

これは, 集合平均 (ensemble average), あるいは標本平均と呼ばれる。しかし, 複数の標本過程が得られない場合には, 1 つの標本過程から平均などの統計量を推定しなければならない。明らかに確率過程 $x(t)$ が非常過程ならば, いかなる統計量も推定することは不可能であるから, $x(t)$ が定常過程であることを限定しよう。定常過程 $x(t)$ において, すべての統計量が任意の 1 つの標本過程 $x(t)$ から, 一致性と不偏性を持って推定できるとき, その確率過程は, エルゴード的 (ergodic), あるいはエルゴード性 (ergodicity) を満たすという。

1 つの標本過程 $x(t)$ から, すべての統計量が推定できるエルゴード性の条件は非常に厳しいから, ここでは, 平均 μ と自己相関関数 $R_x(\tau)$ の 2 つの統計量についてエルゴード性を満たすための条件を考えよう。

1 つの標本過程から, 平均を

$$\hat{\mu} = \frac{1}{T} \int_0^T x(t) dt \quad (3.1)$$

として推定することを考える。これは, 時間平均 (time average) と呼ばれている。こうして推定された時間平均 $\hat{\mu}$ が, 不偏推定量 ($E[\hat{\mu}] = \mu$), かつ一致推定量 ($\lim_{T \rightarrow \infty} E[\hat{\mu}] = \mu, \lim_{T \rightarrow \infty} \text{Var}[\hat{\mu}] = 0$ が十分条件) であるならば, 確率過程 $x(t)$ は平均エルゴード過程 (mean ergodic process) であるといわれる。まず, 不偏性は,

$$E[\hat{\mu}] = E\left[\frac{1}{T} \int_0^T x(t) dt\right] = \frac{1}{T} \int_0^T E[x(t)] dt = \frac{1}{T} \int_0^T \mu dt = \mu$$

であることから満足している．次に，一致性を示すためには，推定値 $\hat{\mu}$ の分散 $\text{Var}[\hat{\mu}] = E[(\hat{\mu} - \mu)^2]$ が $T \rightarrow \infty$ で，0 に収束することを示せば良い．式 (3.1) から

$$\begin{aligned}\text{Var}[\hat{\mu}] &= E[(\hat{\mu} - \mu)^2] \\ &= E\left[\left(\frac{1}{T} \int_0^T x(t) dt - \mu\right)^2\right] \\ &= E\left[\left(\frac{1}{T} \int_0^T (x(t) - \mu) dt\right)^2\right] \\ &= \frac{1}{T^2} \int_0^T \int_0^T E[(x(t_1) - \mu)(x(t_2) - \mu)] dt_1 dt_2\end{aligned}\quad (3.2)$$

となり，上式の最後の式中 $E[\cdot]$ は，定常過程の自己共分散関数 $C_x(t_1 - t_2) = C_x(t_1, t_2)$ である．したがって，平均エルゴード性を満たすための十分条件は，

$$\lim_{T \rightarrow \infty} \text{Var}[\hat{\mu}] = \lim_{T \rightarrow \infty} \frac{1}{T^2} \int_0^T \int_0^T C_x(t_1 - t_2) dt_1 dt_2 = 0$$

となることである．すなわち，確率過程の平均エルゴード性を調べるためには，確率過程 $x(t)$ の自己共分散関数 $C_x(\tau)$ を調べれば良いことになる．

次に，自己相関関数のエルゴード性 (autocorrelation ergodic) を考えよう． $x(t)x(t + \tau)$ の時間平均

$$\hat{R}_x(\tau) = \frac{1}{T} \int_0^T x(t)x(t + \tau) dt \quad (3.3)$$

を自己相関関数の推定値とすれば，この期待値は，

$$E[\hat{R}_x(\tau)] = \frac{1}{T} \int_0^T E[x(t)x(t + \tau)] dt = \frac{1}{T} \int_0^T R_x(\tau) dt = R_x(\tau) \quad (3.4)$$

となる．したがって，式 (3.3) は，自己相関関数の不偏推定量になる．一方，推定量の分散は，

$$\begin{aligned}\text{Var}[\hat{R}_x(\tau)] &= E\left[\left(\hat{R}_x(\tau) - R_x(\tau)\right)^2\right] \\ &= E\left[\hat{R}_x^2(\tau) - 2\hat{R}_x(\tau)R_x(\tau) + R_x^2(\tau)\right] \\ &= \frac{1}{T^2} \int_0^T \int_0^T E\left[x(t_1)x(t_1 + \tau)x(t_2)x(t_2 + \tau)\right] dt_1 dt_2 - R_x^2(\tau)\end{aligned}\quad (3.5)$$

となる．すなわち，自己相関関数のエルゴード性を満たすための十分条件は，上式が $T \rightarrow \infty$ で 0 になることである．このように，自己相関関数のエルゴード性を検証するためには，定常過程 $x(t)$ の 4 次までの統計的性質を調べなければならない．こうしたことから，すべての統計量に対するエルゴード性を検証することは非常に困難であることがわかるであろう．仮に， $x(t)$ が正規確率過程であるとするれば，

$$\begin{aligned}E[x(t_1)x(t_1 + \tau)x(t_2)x(t_2 + \tau)] &= E[x(t_1)x(t_1 + \tau)]E[x(t_2)x(t_2 + \tau)] \\ &\quad + E[x(t_1)x(t_2)]E[x(t_1 + \tau)x(t_2 + \tau)] \\ &\quad + E[x(t_1)x(t_2 + \tau)]E[x(t_1 + \tau)x(t_2)]\end{aligned}\quad (3.7)$$

なので，上式は，

$$\text{Var}[\hat{R}_x(\tau)] = \frac{1}{T^2} \int_0^T \int_0^T \{R_x^2(t_2 - t_1) + R_x(t_2 - t_1 + \tau)R_x(t_2 - t_1 - \tau)\} dt_1 dt_2 \quad (3.8)$$

と変形でき，自己相関関数のエルゴード性は，自己相関関数自身により決まることになる．自己相関関数 $R_x(\tau)$ が $\tau \rightarrow \infty$ で 0 に収束すれば，自己相関関数のエルゴード性を満たすことになる．また，そのためには，確率過程 $x(t)$ の平均 μ が 0 でなければならない．こうしたことから，確率過程を扱う場合には，平均 μ を 0 とおくことが多く，実際の信号処理では，あらかじめ，標本過程 $x(t)$ から，その平均 μ ，つまり直流分を除去しておくことが多い．こうした処理を DC カットと呼ぶことがある．

式 (3.8) は， $t_2 - t_1 = u$ とおき， $\lim_{\tau \rightarrow \infty} R_x(\tau) = 0$ であることを利用すれば， T が十分大きい場合を考えれば，

$$\text{Var}[\hat{R}_x(\tau)] \simeq \frac{1}{T} \int_{-T/2}^{T/2} \{R_x^2(u) + R_x(u + \tau)R_x(u - \tau)\} du \quad (3.9)$$

と近似できる．上式の積分は， T が十分大きければ一定値になるので，自己相関関数の推定量の分散は，データ長 T に反比例することになる．

第4章 確率過程のパワースペクトル密度の推定

4.1 自己相関関数の推定

実際に，エルゴード過程 $x(t)$ の一つの有限長の標本 $x(t), 0 \leq t \leq T$ を観測した場合の自己相関関数の推定は， t の範囲 $0 \leq t \leq T$ を考えれば，

$$\hat{R}_x(\tau) = \begin{cases} \frac{1}{T-\tau} \int_0^{T-\tau} x(t)x(t+\tau)dt, & \tau \geq 0 \\ \frac{1}{T-|\tau|} \int_{|\tau|}^T x(t)x(t+\tau)dt, & \tau < 0 \end{cases} \quad (4.1)$$

として行うことになる．あるいは，1つの式にまとめれば，

$$\hat{R}_x(\tau) = \frac{1}{T-|\tau|} \int_{\max(0, -\tau)}^{\min(T, T-\tau)} x(t)x(t+\tau)dt \quad (4.2)$$

と書ける．この場合，推定量は不偏性を持つが，時間差 τ が T に近い場合には，積分範囲が少なくなることにより推定精度が悪化する．そこで，不偏推定量にはならないが，

$$\hat{R}'_x(\tau) = \frac{1}{T} \int_{\max(0, -\tau)}^{\min(T, T-\tau)} x(t)x(t+\tau)dt \quad (4.3)$$

を自己相関関数の推定量として用いることも多い．これらの2つの推定量の間には，

$$\hat{R}'_x(\tau) = \frac{T-|\tau|}{T} \hat{R}_x(\tau) \quad (4.4)$$

なる関係がある．

4.2 ウィナー - ヒンチンの定理

確率過程 $x(t)$ のパワースペクトルを考えよう．絶対可積分な信号 $x(t)$ のパワースペクトルは，そのフーリエ変換 $X(\omega)$ を用いて， $|X(\omega)|^2$ と定義された．しかしながら， $-\infty \leq t \leq \infty$ の範囲で定義されている確率過程 $x(t)$ は，一般には絶対可積分ではない．そこで，まず，パワースペクトルの概念をこうした絶対可積分ではない信号 $x(t)$ に対して拡張しなければならない．まず，次式で示すように $x(t)$ の有限区間のみを切り出した信号 $x_T(t)$ を考える．

$$x_T(t) = \begin{cases} x(t) & 0 \leq t \leq T \\ 0 & \text{others} \end{cases} \quad (4.5)$$

この有限区間信号 $x_T(t)$ のフーリエ変換は，

$$X_T(\omega) = \int_{-\infty}^{\infty} x_T(t)e^{-i\omega t}dt \quad (4.6)$$

である．単に， $T \rightarrow \infty$ では，パーセバルの等式より， $X_T(\omega) \rightarrow \infty$ なので， $-\infty \leq t \leq \infty$ の範囲で定義されている確率過程に対しては，パワースペクトルの代わりに，単位時間あたりのパワースペクトル，すなわち，パワースペクトル密度 (power spectral density; PSD) を考える．パワースペクトル密度 $P(\omega)$ は，次式で定義される．

$$P(\omega) = \lim_{T \rightarrow \infty} E \left[\frac{|X_T(\omega)|^2}{T} \right] \quad (4.7)$$

式 (4.7) 右辺の $|X_T(\omega)|^2$ を具体的に求めていこう．

$$\begin{aligned} |X_T(\omega)|^2 &= \left| \int_{-\infty}^{\infty} x_T(t) e^{-i\omega t} dt \right|^2 \\ &= \int_{-\infty}^{\infty} x_T(t_1) e^{-i\omega t_1} dt_1 \cdot \int_{-\infty}^{\infty} x_T(t_2) (e^{-i\omega t_2})^* dt_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_T(t_1) x_T^*(t_2) e^{-i\omega(t_1-t_2)} dt_1 dt_2 \end{aligned} \quad (4.8)$$

と表される．ただし， $*$ は， \cdot の複素共役を表す． $x(t)$ が実数であれば， $x^*(t) = x(t)$ なので， $x_T^*(t) = x_T(t)$ である．上式は，変数変換 $t_1 - t_2 = \tau$ ， $t_2 = t$ を行えば，

$$|X_T(\omega)|^2 = \int_{-\infty}^{\infty} e^{-i\omega\tau} \int_{-\infty}^{\infty} x_T(t) x_T(t+\tau) dt d\tau \quad (4.9)$$

となるが， $x_T(t)$ の範囲を考えて，積分範囲を明記すれば，

$$|X_T(\omega)|^2 = \int_{-T}^T e^{-i\omega\tau} \int_{\max(0, -\tau)}^{\min(T, T-\tau)} x_T(t) x_T(t+\tau) dt d\tau \quad (4.10)$$

と書ける．これを式 (4.7) に代入すれば

$$P(\omega) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^T e^{-i\omega\tau} \int_{\max(0, -\tau)}^{\min(T, T-\tau)} E[x_T(t) x_T(t+\tau)] dt d\tau \quad (4.11)$$

となる．ここで，

$$\lim_{T \rightarrow \infty} E[x_T(t) x_T(t+\tau)] = E[x(t) x(t+\tau)] = R_x(\tau)$$

であることを考えれば，

$$\begin{aligned} P(\omega) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^T e^{-i\omega\tau} R_x(\tau) \int_{\max(0, -\tau)}^{\min(T, T-\tau)} dt d\tau \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^T e^{-i\omega\tau} R_x(\tau) (T - |\tau|) d\tau \\ &= \lim_{T \rightarrow \infty} \int_{-T}^T e^{-i\omega\tau} R_x(\tau) \left(1 - \frac{|\tau|}{T}\right) d\tau \\ &= \int_{-\infty}^{\infty} R_x(\tau) e^{-i\omega\tau} d\tau \end{aligned} \quad (4.12)$$

となる．最後の式は，自己相関関数 $R_x(\tau)$ のフーリエ変換であるから，結局，定常過程 $x(t)$ の自己相関関数 $R_x(\tau)$ とパワースペクトル密度 $P(\omega)$ は，互いにフーリエ変換の関係にあることになる．この関係をウィナー - ヒンチンの定理 (Wiener-Khinchine's theorem) という．また，逆変換は，

$$R_x(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega) e^{i\omega\tau} d\omega \quad (4.13)$$

となる．自己相関関数 $R_x(\tau)$ は実偶関数となるため，パワースペクトル密度 $P(\omega)$ もまた実偶関数になる．
2つの確率過程 $x(t)$, $y(t)$ の相互相関関数 $R_{xy}(\tau)$ のフーリエ変換

$$P_{xy}(\omega) = \int_{-\infty}^{\infty} R_{xy}(\tau) e^{-i\omega\tau} d\tau \quad (4.14)$$

は，相互パワースペクトル密度 (cross-power spectral density) と呼ばれる．相互相関関数 $R_{xy}(\tau)$ は，一般に偶関数にはならないので，相互パワースペクトル密度 $P_{xy}(\omega)$ が実関数になるとは限らない．

次に，2つの確率過程の和 $z(t) = x(t) + y(t)$ のパワースペクトル密度を調べてみよう． $z(t)$ の自己相関関数 $R_z(\tau)$ は

$$\begin{aligned} R_z(\tau) &= E[z(t)z(t+\tau)] = E[(x(t) + y(t))(x(t+\tau) + y(t+\tau))] \\ &= E[x(t)x(t+\tau) + y(t)y(t+\tau) + x(t)y(t+\tau) + y(t)x(t+\tau)] \\ &= R_x(\tau) + R_y(\tau) + R_{xy}(\tau) + R_{yx}(\tau) \end{aligned} \quad (4.15)$$

であるから，両辺をフーリエ変換することにより，次の関係が得られる．

$$P_z(\omega) = P_x(\omega) + P_y(\omega) + P_{xy}(\omega) + P_{yx}(\omega) \quad (4.16)$$

仮に，2つの確率過程が直交している ($R_{xy}(\tau) = R_{yx}(\tau) = 0$) ならば， $z(t)$ のパワースペクトル密度は，それを構成する確率過程のパワースペクトル密度の和

$$P_z(\omega) = P_x(\omega) + P_y(\omega) \quad (4.17)$$

として表される．

4.3 ブラックマン・チューキー法によるパワースペクトル密度の推定と窓関数

確率過程 $x(t)$ の自己相関関数 $R_x(\tau)$ を推定し，そのフーリエ変換として，パワースペクトル密度を求めるパワースペクトル密度の推定法をブラックマン・チューキー法 (Blackman-Tukey method) という．4.1で述べたように，自己相関関数 $R_x(\tau)$ の推定では，時間差 τ が大きくなると推定精度が悪化する．したがって，自己相関関数のフーリエ変換により得られたパワースペクトル密度もまた，推定精度が悪化することになる．そこで，自己相関関数において，推定精度の低い，すなわち時間差 τ の大きいところを切り捨てることにより，パワースペクトル密度の推定精度を改善を試みることがある．例えば，時間差 $|\tau| > M$ を切り捨てる場合には，

$$w_R(\tau) = \begin{cases} 1, & |\tau| \leq M \\ 0, & \text{others} \end{cases} \quad (4.18)$$

なる関数を用いて，

$$\hat{R}_x(\tau) w_R(\tau)$$

とすることにより，自己相関関数の推定精度の高いところのみを利用することができる．一般に，こうした関数 $w_R(\tau)$ は，自己相関関数を特定の範囲の窓を通して観測することに相当することから，窓関数 (window function) と呼ばれている．また，その窓の形は，矩形であることから，矩形窓 (Rectangle window) と呼ばれている．この窓関数を用いて推定されたパワースペクトル密度 $\hat{P}'(\omega)$ は，

$$\hat{P}'(\omega) = \mathcal{F}[\hat{R}_x(\tau) w_R(\tau)] \quad (4.19)$$

となる。

次に、この窓関数がパワースペクトル密度の推定にどのような影響を与えるかを考えてみよう。時間軸における積は、周波数軸ではそれぞれのフーリエ変換の畳み込み積分である。矩形窓関数 $w_R(\tau)$ のフーリエ変換は、

$$W_R(\omega) = \mathcal{F}[w_R(\tau)] = 2M \text{sinc}(M\omega)$$

であり、窓関数を用いない場合のパワースペクトル密度の推定値を $\hat{P}(\omega) = \mathcal{F}[\hat{R}_x(\tau)]$ とすれば、窓関数を用いた場合のパワースペクトル密度の推定値は、

$$\hat{P}'(\omega) = \frac{1}{2\pi} \hat{P}(\omega) * W_R(\omega) \quad (4.20)$$

となる。つまり、 $W_R(\omega)$ の広がりが多いほど、パワースペクトル密度の推定値は、周波数に関しなまることになる。一般に、時間軸における窓関数をラグウィンドウ (lag window) といい、そのフーリエ変換である周波数軸における窓をスペクトルウィンドウ (spectrum window) と呼ぶ。ラグウィンドウの幅 M が大きくなると、そのスペクトルウィンドウはデルタ関数に近づき、スペクトル分解能が高くなる。したがって、窓関数を用いるということは、パワースペクトル密度推定において、不偏性を犠牲にして、推定量の分散を小さくする、すなわち推定精度を改善することを意味する。不偏性と推定精度は、互いにトレードオフの関係にある。

さて、矩形窓を用いたパワースペクトル密度推定の別の問題を考えてみよう。上で述べたように、矩形窓のフーリエ変換は、sinc 関数であることから負の値を持つ。そこで、鋭いピークを持つスペクトル $\hat{P}_x(\omega)$ がある場合、そのピークが sinc 関数の負の位置に入った場合には、畳み込み積分により、パワースペクトル密度 $\hat{P}'_x(\omega)$ は、実際にはありえない負のパワーレベルを持つことがある。これは、矩形窓を用いた場合にしばしば生ずる現象であり注意を要する。

矩形窓を用いた場合の負のスペクトルの発生を避けるために、三角窓 (Bartlett window)

$$w_B(\tau) = \begin{cases} 1 - \frac{|\tau|}{M}, & \text{for } |\tau| \leq M \\ 0, & \text{others} \end{cases} \quad (4.21)$$

が用いられることも多い。式 (4.3) で述べた自己相関関数の推定量 $\hat{R}'_x(\tau)$ は、上式において、 $M = T$ とした三角窓に他ならないことがわかる。三角窓のフーリエ変換、つまりスペクトルウィンドウは、

$$W_B(\omega) = M \text{sinc}^2\left(\frac{\omega M}{2}\right) \quad (4.22)$$

であり、すべての ω に対して非負となる大きな特徴がある。三角窓のラグウィンドウとスペクトルウィンドウを図 4.1 上から 2 段目に示す。

その他、

$$w(\tau) = \begin{cases} a + b \cos\left(\frac{\pi\tau}{M}\right) + c \cos\left(\frac{2\pi\tau}{M}\right), & \text{for } |\tau| \leq M \\ 0, & \text{others} \end{cases} \quad (4.23)$$

とし、 $a = 0.54, b = 0.46, c = 0$ として与えられるハミング窓 (Hamming window)、 $a = b = 0.5, c = 0$ として与えられるハンニング窓 (Hanning window)、 $a = 0.42, b = 0.5, c = 0.08$ として与えられるブラックマン窓 (Blackman window) がある。

矩形窓、三角窓、ハミング窓、ハンニング窓、ブラックマン窓のラグウィンドウとスペクトルウィンドウを図 4.1 に示した。図中、左の列はラグウィンドウ、右の列は対するスペクトルウィンドウを表している。いずれも、窓幅 $M = 64[\text{ms}]$ とした際の結果である。スペクトルウィンドウにおいて、中心にある最も大き

な山をメインローブ、それ以外の山をサイドローブという。メインローブの幅が小さく、サイドローブの高さが小さいほど、窓関数を用いることによるスペクトルの歪みが小さく、望ましいことになる。矩形窓は、メインローブの幅を最小にする窓関数であり、その代償としてサイドローブの振幅がかなり大きく、負値をとるといふ問題もある。三角窓は、スペクトルウィンドウで負値をとることはないが、ラグウィンドウで微分不可能な点が3箇所もあるため、比較的サイドローブの振幅が大きくなっている。ブラックマン窓は、ここに示した窓関数の中で、最も滑らかなラグウィンドウであり、スペクトルウィンドウのサイドローブの高さを最小にする窓関数である。その代償としてメインローブの幅は最大である。ハミング窓は、矩形窓についてスペクトルウィンドウのメインローブの幅が小さいが、ラグウィンドウの両端で不連続になるため、スペクトルウィンドウのサイドローブの収まりが悪い。また、ハニング窓は、サイドローブの収まりは良いが、最大の高さをもつサイドローブの高さはハミング窓に比べ大きい。これ以外にも、チェビシェフ窓、カイザー窓などがある。スペクトル密度を推定する際には、これらの窓関数の特徴を理解し、目的により使い分けることが必要である。

[演習 3] 平均 0, 分散 1 の正規分布 $N(0, 1)$ に従う白色雑音系列を, 伝達関数

$$H(z) = \frac{1}{1 - 0.98z^{-1} + 0.5z^{-2} + 0.2z^{-3}}$$

で表現されるシステムに入力した際の出力を $x_t, t = 1, \dots, T$ とする。この正規確率系列 $x_t, t = 1, \dots, T$ の自己相関関数を式 (4.1) を用いて推定し, そのフーリエ変換として, つまりブラックマン・チューキー法によりパワースペクトル密度を推定するものとする。その際, さまざまな窓幅の矩形窓, パートレット窓, ブラックマン窓を用いて推定されたパワースペクトル密度の推定精度を比較せよ。

[解答] 系列長 $T = 1024$ [s] として, 上記のように正規確率系列 $x_t, t = 1, \dots, T$ を作成し, 自己相関関数 $R_x(\tau), \tau = -511, -510, \dots, -1, 0, 1, \dots, 510, 511$ [s] を推定した。推定された自己相関関数に, 窓幅 $M = 32, 64, 256$ [s] の矩形窓, パートレット窓, ブラックマン窓をそれぞれ乗じ, フーリエ変換 (FFT) を行うことによりパワースペクトル密度 (PSD) を推定した。その結果を図 4.2(a) ~ (i) に示す。図中, 緑色の実線は PSD の推定値, 黒色の実線はその理論値を表す。窓幅 M が小さい場合には, 自己相関関数を打ち切ることによるスペクトルのひずみが生じ, 逆に窓幅 M が大きい場合には, 自己相関関数の精度の悪い時間差の大きい部分を使うことによる精度の悪化が見られる。また, 不連続な矩形窓, 連続であるが微分不可能なパートレット窓, 微分可能なブラックマン窓の順で, パワースペクトル密度の推定精度が改善されてゆくことがわかる。

4.4 ペリオドグラム法によるパワースペクトル密度の推定

ブラックマン・チューキー法では, 確率過程の自己相関関数の推定を通して, そのフーリエ変換によりパワースペクトル密度を求めた。しかしながら, 高速離散フーリエ変換 (FFT) の発見により, フーリエ変換が容易にかつ高速に実現できるようになってからは, 確率過程を直接フーリエ変換する手法が用いられるようになった。こうしたパワースペクトル密度の推定法は, ペリオドグラム法 (periodogram method), 直接法 (direct method), FFT 法 (FFT method) などと呼ばれている。フーリエ変換として FFT が用いられること以外は, 時刻を連続時間で定義しても, 離散時間で定義しても同じであるため, ここでは連続時間として議論を進める。

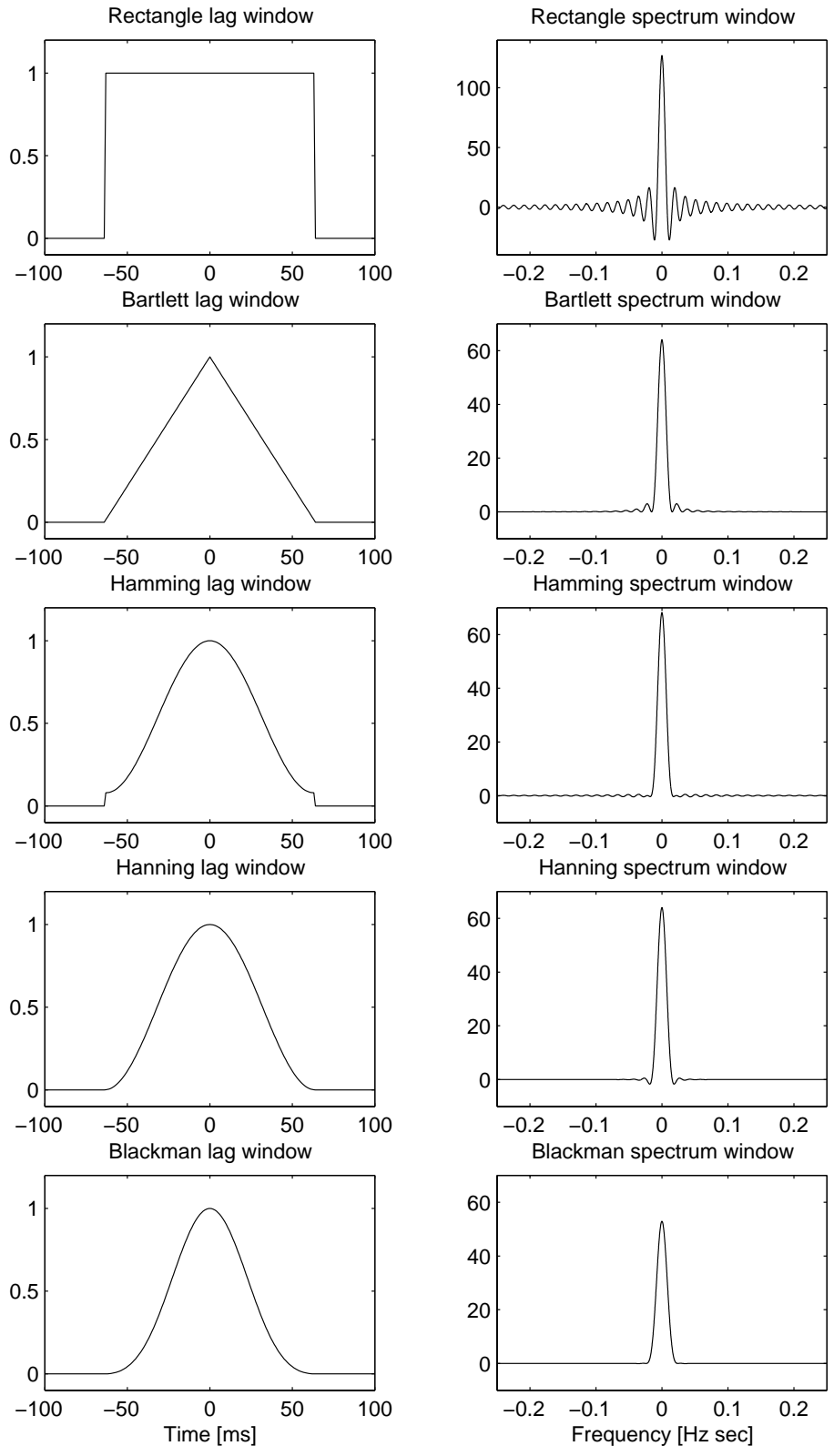


図 4.1: 代表的な窓関数のラグウィンドウとスペクトルウィンドウ

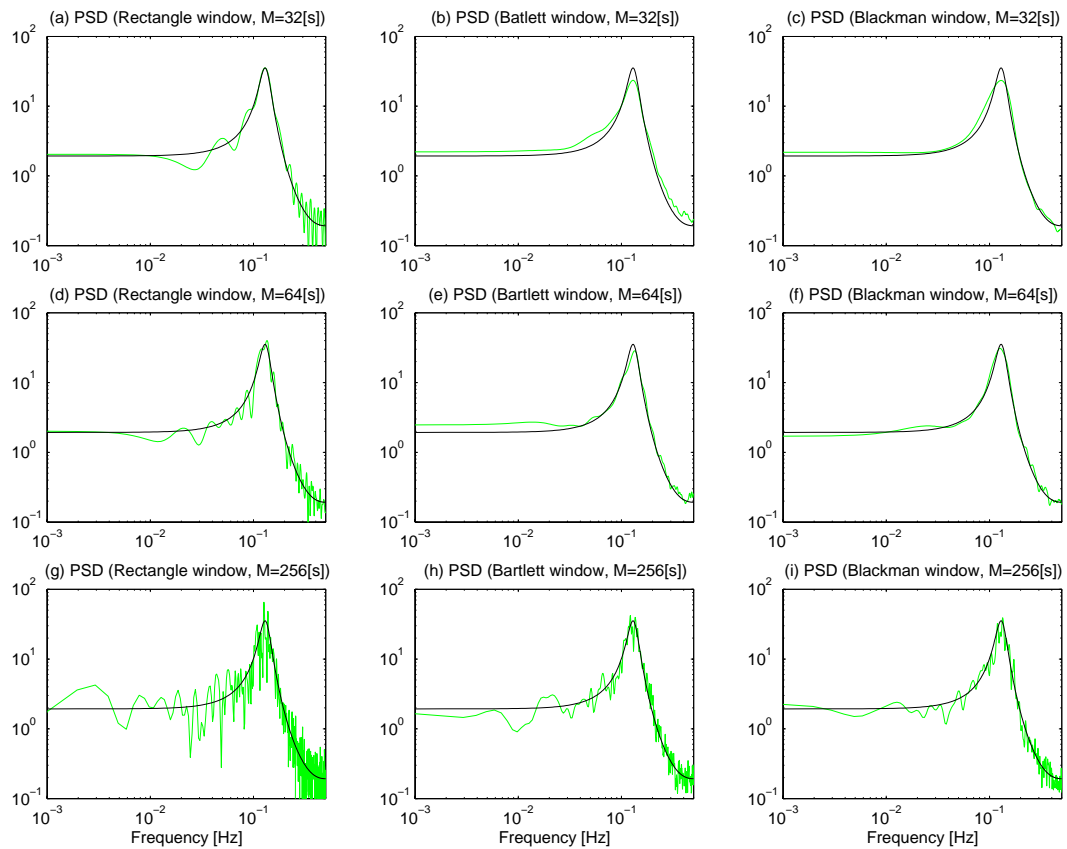


図 4.2: 演習 3 の解答 : ブラックマン・チューキー法により推定されたパワースペクトル密度

有限長の確率過程

$$x_T(t) = \begin{cases} x(t) & 0 \leq t \leq T \\ 0 & \text{others} \end{cases}$$

に対するフーリエ変換の絶対値の2乗を T で除した

$$P_{\text{per}}(\omega) \equiv \frac{1}{T} |\mathcal{F}[x_T(t)]|^2$$

を考える．上式は更に次のように変形できる．

$$\begin{aligned} P_{\text{per}}(\omega) &= \frac{1}{T} \mathcal{F}[x_T(t)] \mathcal{F}[x_T(t)]^* \\ &= \frac{1}{T} \int_{-\infty}^{\infty} x_T(t) e^{-i\omega t} dt \int_{-\infty}^{\infty} x_T(t) e^{i\omega t} dt \\ &= \frac{1}{T} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_T(t_1) x_T(t_2) e^{-i\omega(t_1-t_2)} dt_1 dt_2 \\ &= \int_{-\infty}^{\infty} \frac{1}{T} \int_{-\infty}^{\infty} x_T(t) x_T(t-\tau) dt e^{-i\omega\tau} d\tau \end{aligned} \quad (4.24)$$

ここで， $x_T(t)$ の範囲を考えれば，

$$\frac{1}{T} \int_{-\infty}^{\infty} x_T(t) x_T(t-\tau) dt$$

は， $x(t)$ のデータ長 T の一つの標本 $x(t)$ から推定された自己相関関数 $\hat{R}_x(\tau)$ に等しいから，式 (4.24) は，

$$P_{\text{per}}(\omega) = \int_{-T}^T \hat{R}_x(\tau) e^{-i\omega\tau} d\tau \quad (4.25)$$

となる． $P_{\text{per}}(\omega)$ は，ペリオドグラム (periodogram) と呼ばれ，データ長 T に相当する時間差 τ までの自己相関関数の推定値を用いた Blackman-Tukey 法によるパワースペクトル密度の推定値と等価である．したがって，前に述べたとおり，その推定精度は悪い． $x(t)$ を正規確率過程とすれば，ペリオドグラムの分散は，

$$\begin{aligned} \text{Var}[P_{\text{per}}(\omega)] &= E \left[\left(P_{\text{per}}(\omega) - E[P_{\text{per}}(\omega)] \right)^2 \right] \\ &= P^2(\omega) \left\{ 1 + \left(\frac{\sin(\omega T)}{T \sin(\omega)} \right)^2 \right\} \end{aligned} \quad (4.26)$$

となることが知られている．つまり，推定値の分散は，データ長 T を長くしても減少せず，真のスペクトル $P(\omega)$ の2乗に比例することになる．したがって，ペリオドグラムそのものは，一致推定量とはならず，スペクトルの推定値としては不適切である．したがって，推定精度を向上させるためには，複数の標本に対してペリオドグラムを推定し，それらの集合平均をとらなければならない．与えられているデータが $x(t)$ ， $0 \leq t \leq T$ のみである場合には，エルゴード性を満たすことを条件として，この $x(t)$ を K 分割し，各部分列のペリオドグラム $P_{\text{per}}^{(k)}$ ， $k = 1, \dots, K$ を求め，それらを集合平均することにより，

$$\hat{P}(\omega) = \frac{1}{K} \sum_{k=1}^K P_{\text{per}}^{(k)}(\omega) \quad (4.27)$$

と推定すればよい．この場合のパワースペクトル密度の推定分散は，

$$\text{Var}[\hat{P}(\omega)] \propto \frac{1}{K}$$

となる． $K \rightarrow \infty$ となるように $T \rightarrow \infty$ とすれば， $\text{Var}[\hat{P}(\omega)] \rightarrow 0$ となり一致推定量になり得る．ただし，離間時間で定義された確率過程に対し，FFT を用いて行う場合には，分割により系列長が少なくなると，2

章で述べたように、周波数軸でのサンプリング間隔が広がるため、周波数分解性能が劣るという問題がある。周波数分解性能と推定精度は互いにトレードオフの関係にある。また、FFT を用いる場合には、部分列の系列長が 2 のべき乗であることが計算速度の意味で望ましい。もとの確率系列の系列長が 2 のべき乗の整数倍とはならない場合には、オーバーラップを許可して部分列に分割することも可能である。しかし、この場合、オーバーラップの幅を増やして分割数を増やしても、推定精度がいくらかでも改善されることはない。

離間時間で定義された確率過程に対し FFT を用いてパワースペクトル密度を推定する際には、新たな別の問題が生ずる。FFT は、信号を周期関数と見なしたフーリエ変換であるから、 $x(t)$ から切り出された部分列に対し FFT を行うことは、各部分列が周期関数であるものとしてフーリエ変換を求めることに相当し、真の値に対して歪みを持つことになる。こうしたスペクトルの歪みの発生は、ギブス現象 (Gibb's phenominon) と呼ばれ、特に、部分列の始めと終わりの値が大きくかけ離れている場合に顕著である。そこで、ギブス現象を避けるため適当な窓関数を用いて、確率過程 $x(t)$ を部分列に切り出すことがしばしば行われる。この場合、用いる窓関数により、切り出された部分列のパワーが変化するため、窓関数のパワーを 1 に正規化しておく必要がある。つまり、窓関数を $w(\tau), |\tau| \leq M$ とすれば、

$$\frac{w(\tau)}{\sqrt{\frac{1}{2M} \int_{-M}^M w^2(\tau) d\tau}}$$

を新たな窓関数として用いることになる。

[演習 4] 平均 0, 分散 1 の正規分布 $N(0, 1)$ に従う白色雑音系列を、伝達関数

$$H(z) = \frac{1 + 1.5z^{-1} + 0.5z^{-2}}{1 - 0.98z^{-1}}$$

で表現されるシステムに入力した際の出力を $x_t, t = 1, \dots, T$ とする。こうした確率系列に対し、ブラックマン・チューキー法とペリオドグラム法により推定されたパワースペクトル密度の推定精度を比較せよ。

[解答] 系列長 $T = 2048[s]$ として、上記のように正規確率系列 $x_t, t = 1, \dots, T$ を作成した。式 (4.1) を使用して推定された自己相関関数 $R_x(\tau), \tau = -2047, -2046, \dots, -1, 0, 1, \dots, 2046, 2047[s]$ に窓幅 $M = 2048[s]$ でブラックマン窓をかけ、そのフーリエ変換によりパワースペクトル密度を推定 (ブラックマン・チューキー法) した結果を図 4.3(a) に示す。また、確率系列 $x_t, t = 1, \dots, T$ から推定されたペリオドグラム P_{per} を同図 (b) に示す。確率系列 $x_t, t = 1, \dots, T$ をオーバーラップすることなく分割数 $K = 2, 4, 8$ として部分列に切り出し、それぞれのペリオドグラムの集合平均として推定されたパワースペクトル密度を同図 (c), (e), (g) に示す。 $K = 2, 4, 8$ として、各部分列にブラックマン窓をかけた場合の結果を同図 (d), (f), (h) に示す。(a) と (b) は、式 (4.25) に示したように、ほとんど等しい結果となっていることがわかる。また、分割数 K を増やすと、パワースペクトル密度の推定精度が改善されてゆくことがわかる。更に、窓関数を用いることにより、高い周波数でスペクトル密度が持ち上がるというスペクトルの歪み、つまりギブス現象が抑えられることがわかる。

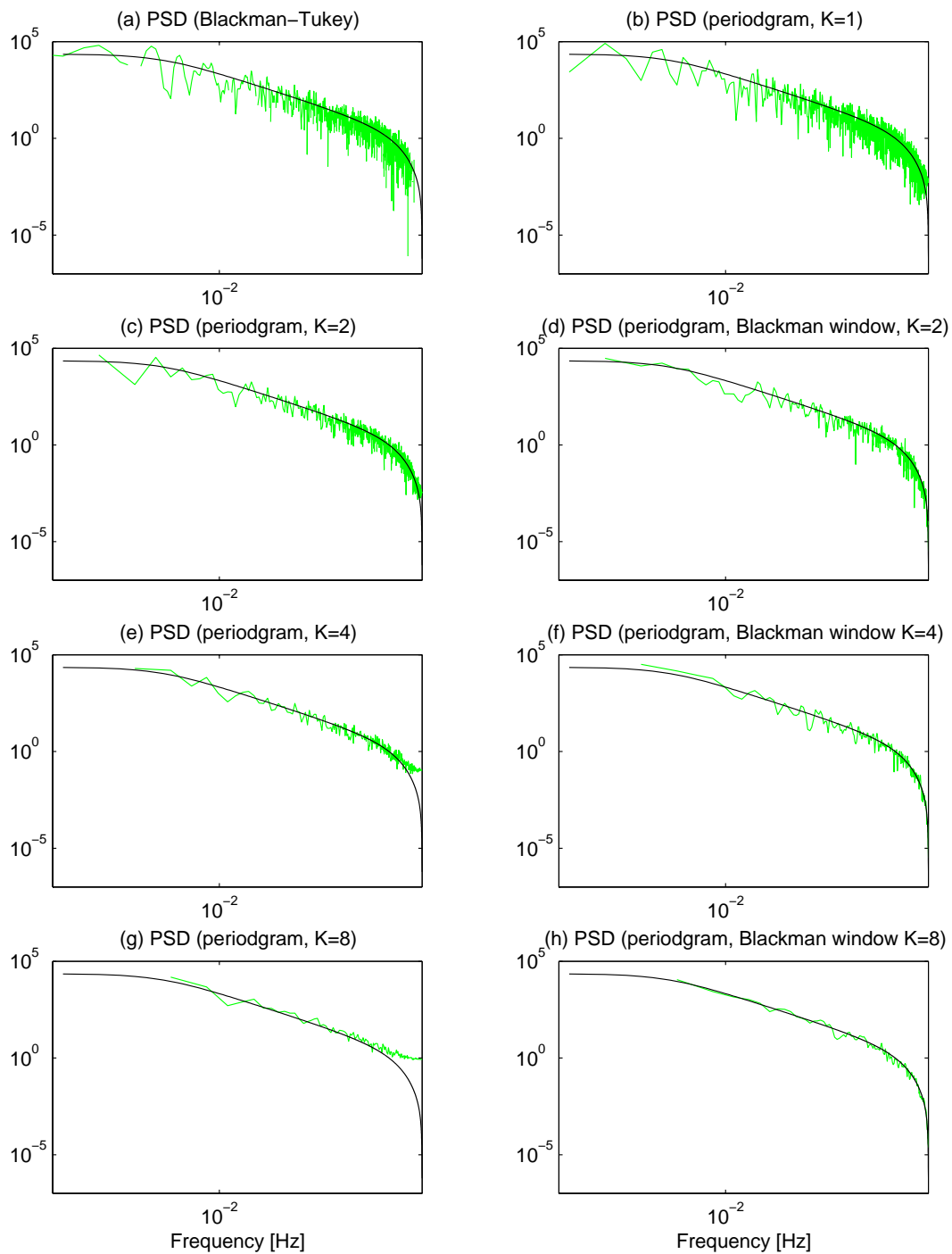


図 4.3: 演習 4 の解答：ブラックマン・チューキー法とペリオドグラム法により推定されたパワースペクトル密度の比較（推定値：緑色の実線，理論値：黒色の実線）

第5章 線形予測

5.1 線形予測モデル

定常確率過程とみなせる時系列 $x_n, n = 1, 2, \dots$ があるものとしよう。ある時刻 n での時系列の値 x_n をそれよりも過去 N 点の時系列の値 x_{n-1}, \dots, x_{n-N} から予測，すなわち推定することを考えよう。これを

$$\hat{x}_n = f(x_{n-1}, \dots, x_{n-N}) \quad (5.1)$$

と書くことにする。 $f(\cdot)$ は予測値，すなわち推定量を与える関数である。一般に，観測された時系列から未来の値を予測する問題を予測問題 (prediction problem) という。推定値 \hat{x}_n と真値 x_n の差 $u_n \equiv x_n - \hat{x}_n$ を予測誤差 (prediction error) と呼ぶ。式 (5.1) の代わりに，予測誤差 u_n を明示して，

$$x_n = f(x_{n-1}, \dots, x_{n-N}) + u_n \quad (5.2)$$

と書くこともある。式 (5.1) や式 (5.2) を予測モデルと呼ぶ。特に，予測関数 $f(\cdot)$ を線形関数で実現する場合，式 (5.1) や式 (5.2) は，線形予測モデル (linear prediction model) と呼ばれ，次式で表現される。

$$x_n = - \sum_{k=1}^N a_k x_{n-k} + u_n \quad (5.3)$$

右辺の第1項につけられている負記号は，便宜上つけたものであり，本質的な意味はない。式 (5.3) で示した線形予測モデルは， N 次自己回帰モデル (autoregressive model)，あるいは N 次 AR モデルと呼ばれ，AR(N) モデルと書かれる。AR モデルのパラメータ $a_k, k = 1, \dots, N$ を自己回帰係数，もしくは AR 係数という。

x_n の予測に， x_{n-1}, \dots, x_{n-N} だけでなく，時刻 n より以前の予測誤差 u_{n-1}, \dots, u_{n-M} を用いる場合には，線形予測モデルは，

$$x_n = \sum_{k=0}^M b_k u_{n-k} - \sum_{k=1}^N a_k x_{n-k} \quad (5.4)$$

と書ける。式 (5.4) で示した線形予測モデルは，自己回帰・移動平均モデル (autoregressive-moving average model)，あるいは (N, M) 次 ARMA モデルといわれ，ARMA(N, M) モデルと書く。式 (5.4) において，右辺第2項がない場合，つまり

$$x_n = \sum_{k=0}^M b_k u_{n-k} \quad (5.5)$$

は， M 次移動平均モデル (moving average model)，あるいは M 次 MA モデルといい，MA(M) モデルと書く。式 (5.4)，式 (5.5) における $b_k, k = 1, \dots, M$ を移動平均係数，もしくは MA 係数という。

定常確率過程とみなせる時系列 $x_n, n = 1, 2, \dots$ の平均が0，つまり $\mu_x = E[x_n] = 0, n = 1, 2, \dots$ であることを仮定し，AR モデルにより予測することを考えよう。平均が0でない場合には，時系列 $x_n, n = 1, 2, \dots$

から平均値を引いた $x_n - \mu_x$, $n = 1, 2, \dots$ を改めて x_n , $n = 1, 2, \dots$ とみなすものとする。式 (5.3) は、ベクトル

$$\mathbf{x} \equiv (x_{n-1}, x_{n-2}, \dots, x_{n-N})^T, \quad \mathbf{a} \equiv (a_1, a_2, \dots, a_N)^T$$

を用いれば、

$$x_n = -\mathbf{x}^T \mathbf{a} + u_n \quad (5.6)$$

と書きなおせる。

予測問題では、予測誤差 u_n が小さい方がより望ましい予測であると思えるから、予測誤差の2乗の期待値を最小にするように、AR 係数 \mathbf{a} を決めるものとしよう。時系列 x_n , $n = 1, 2, \dots$ の平均が0なので、その線形結合で与えられる u_n , $n = 1, 2, \dots$ もまた平均が0である。したがって、予測誤差の2乗の期待値は、予測誤差の分散に等しい。つまり、

$$\sigma_u^2 = E[(u_n - E[u_n])^2] = E[u_n^2] \quad (5.7)$$

を最小にするように AR 係数 \mathbf{a} を決めよう。式 (5.7) は、式 (5.6) を代入すれば、

$$\begin{aligned} \sigma_u^2 &= E[(x_n + \mathbf{x}^T \mathbf{a})^2] \\ &= E[(x_n + \mathbf{x}^T \mathbf{a})^T (x_n + \mathbf{x}^T \mathbf{a})] \\ &= E[x_n^2 + 2x_n \mathbf{x}^T \mathbf{a} + (\mathbf{x}^T \mathbf{a})^T (\mathbf{x}^T \mathbf{a})] \\ &= E[x_n^2] + 2\mathbf{a}^T E[x_n \mathbf{x}] + \mathbf{a}^T E[\mathbf{x} \mathbf{x}^T] \mathbf{a} \end{aligned} \quad (5.8)$$

と変形でき、 $E[x_n \mathbf{x}]$, $E[\mathbf{x} \mathbf{x}^T]$ をそれぞれ

$$\mathbf{r} \equiv E[x_n \mathbf{x}] = \begin{pmatrix} R_x(1) \\ R_x(2) \\ \vdots \\ R_x(N) \end{pmatrix}, \quad \mathbf{R} \equiv E[\mathbf{x} \mathbf{x}^T] = \begin{pmatrix} R_x(0) & R_x(1) & \cdots & R_x(N-1) \\ R_x(1) & R_x(0) & \cdots & R_x(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_x(N-1) & R_x(N-2) & \cdots & R_x(0) \end{pmatrix}$$

とおけば、

$$\sigma_u^2 = R_x(0) + 2\mathbf{a}^T \mathbf{r} + \mathbf{a}^T \mathbf{R} \mathbf{a} \quad (5.9)$$

と書ける。これを \mathbf{a} で偏微分すれば、

$$\frac{\partial \sigma_u^2}{\partial \mathbf{a}} = 2\mathbf{r} + 2\mathbf{R} \mathbf{a} \quad (5.10)$$

となるので、 $\frac{\partial \sigma_u^2}{\partial \mathbf{a}} \equiv \mathbf{0}$ の解は、

$$\mathbf{R} \mathbf{a} = -\mathbf{r} \quad (5.11)$$

を解くことにより与えられる。式 (5.11) は、Yule-Walker 方程式と呼ばれ、この式を用いた AR 係数の推定は、Yule-Walker 推定といわれる。

式 (5.11) を解くことにより得られる AR 係数の推定値を $\hat{\mathbf{a}}$ 、予測誤差の分散 σ_u^2 の最小値を推定されたものと見て $\hat{\sigma}_u^2$ と書くことにする。式 (5.11) を式 (5.9) に代入すれば、

$$\hat{\sigma}_u^2 = R_x(0) + \hat{\mathbf{a}}^T \mathbf{r}$$

となるので、 u_n の分散の最小値は、

$$\hat{\sigma}_u^2 = \sum_{l=0}^N \hat{a}_l R_x(l) \quad (5.12)$$

であることがわかる．ただし， $\hat{a}_0 \equiv 1$ とおいた．式 (5.11) は，

$$\sum_{l=0}^N \hat{a}_l R_x(l-k) = 0, \quad k = 1, 2, \dots, N \quad (5.13)$$

と変形できるから，式 (5.12) と式 (5.13) をまとめれば，

$$\sum_{l=0}^N \hat{a}_l R_x(l-k) = \begin{cases} \hat{\sigma}_u^2, & k = 0 \\ 0, & k = 1, \dots, N \end{cases} \quad (5.14)$$

なる表現が得られる．これを行列を用いて表現すれば，

$$\begin{pmatrix} R_x(0) & R_x(1) & R_x(2) & \cdots & R_x(N) \\ R_x(1) & R_x(0) & R_x(1) & \cdots & R_x(N-1) \\ R_x(2) & R_x(1) & R_x(0) & \cdots & R_x(N-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_x(N) & R_x(N-1) & R_x(N-2) & \cdots & R_x(0) \end{pmatrix} \begin{pmatrix} 1 \\ \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_N \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_u^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (5.15)$$

となり，さらに変形すれば，

$$\begin{pmatrix} 1 & \hat{a}_1 & \hat{a}_2 & \cdots & \hat{a}_{N-1} & \hat{a}_N & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \hat{a}_1 & \hat{a}_2 & \cdots & \hat{a}_{N-1} & \hat{a}_N & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \hat{a}_1 & \hat{a}_2 & \cdots & \hat{a}_{N-1} & \hat{a}_N & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \cdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & \hat{a}_1 & \hat{a}_2 & \cdots & \hat{a}_{N-1} & \hat{a}_N & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & \hat{a}_1 & \hat{a}_2 & \cdots & \hat{a}_{N-1} & \hat{a}_N \end{pmatrix} \begin{pmatrix} R_x(N) \\ R_x(N-1) \\ \vdots \\ R_x(0) \\ \vdots \\ R_x(N-1) \\ R_x(N) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \hat{\sigma}_u^2 \end{pmatrix}$$

となる．AR 係数 $\hat{a}_k, k = 1, \dots, N$ と予測誤差の分散 $\hat{\sigma}_u^2$ から，次式のように， x_n の自己相関関数を知ることができる．

$$\begin{pmatrix} R_x(0) \\ R_x(1) \\ R_x(2) \\ \vdots \\ R_x(N) \end{pmatrix} = \left\{ \begin{pmatrix} 0 & \hat{a}_{N-1} & \hat{a}_{N-2} & \cdots & \hat{a}_1 & 1 \\ 0 & \hat{a}_{N-2} & \cdots & \hat{a}_1 & 1 & 0 \\ 0 & \hat{a}_{N-3} & \hat{a}_1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} \hat{a}_N & 0 & 0 & 0 & \cdots & 0 \\ \hat{a}_{N-1} & \hat{a}_N & 0 & 0 & \cdots & 0 \\ \cdots & \hat{a}_{N-1} & \hat{a}_N & 0 & \cdots & 0 \\ \ddots & \cdots & \ddots & \ddots & \ddots & \ddots \\ \hat{a}_1 & \hat{a}_2 & \cdots & \hat{a}_{N-1} & \hat{a}_N & 0 \\ 1 & \hat{a}_1 & \hat{a}_2 & \cdots & \hat{a}_{N-1} & \hat{a}_N \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \hat{\sigma}_u^2 \end{pmatrix} \quad (5.16)$$

5.2 予測誤差の統計的性質

予測誤差 u_n の統計的性質を調べてみよう．まず，予測誤差 u_n と時系列 x_{n-k} , $k = 0, 1, \dots$ の間の相互相関関数

$$E[u_n x_{n-k}] = E[(x_n + \hat{\mathbf{a}}^T \hat{\mathbf{x}}) x_{n-k}] = E[x_n x_{n-k}] + E[\hat{\mathbf{a}}^T \hat{\mathbf{x}} x_{n-k}]$$

を調べてみよう．AR 係数の推定値 $\hat{\mathbf{a}}$ は，確率過程 x_n , $n = 1, 2, \dots$ から推定されたものであるから，本来，確率変数として扱わなければならない．しかし，確率過程 x_n , $n = 1, 2, \dots$ が十分な長さの系列長を持つ場合には，そこから推定された $\hat{\mathbf{a}}$ の統計的変動は，確率過程 x_n , $n = 1, 2, \dots$ そのものと比べると相対的に小さいと考えられる．そこで，議論を簡単にするため，本節では AR 係数の推定値 $\hat{\mathbf{a}}$ を x_n , $n = 1, 2, \dots$ に対して確定的な変数であるとみなすことにする．したがって，上式は，

$$\begin{aligned} E[u_n x_{n-k}] &= E[x_n x_{n-k}] + \hat{\mathbf{a}}^T E[\mathbf{x} x_{n-k}] \\ &= R_x(k) + \sum_{l=1}^N \hat{a}_l R_x(l-k) \\ &= \sum_{l=0}^N \hat{a}_l R_x(l-k) \end{aligned} \quad (5.17)$$

となり，式 (5.14) に等しいことがわかる．したがって，予測誤差 u_n と時系列 x_{n-k} , $k = 0, 1, \dots$ の間の相互相関関数は，

$$E[u_n x_{n-k}] = \begin{cases} \hat{\sigma}_u^2, & k = 0 \\ 0, & k = 1, 2, \dots, N \end{cases} \quad (5.18)$$

となる．予測誤差 u_n は $x_{n-1}, x_{n-2}, \dots, x_{n-N}$ と無相関であり，このことから，予測誤差 u_n は， x_n から $x_{n-1}, x_{n-2}, \dots, x_{n-N}$ と相関のある成分を差し引いた残りの成分であると考えられる．

次に，予測誤差 u_n , $n = 1, 2, \dots$ の自己相関関数 $E[u_n u_{n-k}]$ を調べてみよう．予測誤差 u_n , $n = 1, 2, \dots$ の自己相関関数 $E[u_n u_{n-k}]$ は，

$$\begin{aligned} E[u_n u_{n-k}] &= E\left[\left(\sum_{l=0}^N \hat{a}_l x_{n-l}\right)\left(\sum_{m=0}^N \hat{a}_m x_{n-k-m}\right)\right] \\ &= \sum_{l=0}^N \sum_{m=0}^N \hat{a}_l \hat{a}_m E[x_{n-l} x_{n-k-m}] \\ &= \sum_{l=0}^N \sum_{m=0}^N \hat{a}_l \hat{a}_m R_x(k+m-l) \\ &= \sum_{m=0}^N \hat{a}_m \sum_{l=0}^N \hat{a}_l R_x(k+m-l) \end{aligned} \quad (5.19)$$

と表現できる．式 (5.14) より， $1 \leq k+m \leq N$ では，

$$\sum_{l=0}^N \hat{a}_l R_x(k+m-l) = 0$$

となるので， $k = 0$ では，

$$E[u_n u_n] = \sum_{m=0}^N \hat{a}_m \sum_{l=0}^N \hat{a}_l R_x(m-l)$$

$$\begin{aligned}
&= \hat{a}_0 \sum_{l=0}^N \hat{a}_l R_x(l) \\
&= \sum_{l=0}^N \hat{a}_l R_x(l) \\
&= \hat{\sigma}_u^2
\end{aligned} \tag{5.20}$$

というあたりまえの結果が得られる．一方， $k > 0$ では，

$$E[u_n u_{n-k}] = \sum_{m=N-k+1}^N \hat{a}_m \sum_{l=0}^N \hat{a}_l R_x(k+m-l) \tag{5.21}$$

となる．ここで， $\hat{a}_k = 0$ ， $k > L$ ， $N > L$ とすれば， $\hat{a}_m = 0$ ， $m = L+1, \dots, N$ は 0 となるので，

$$E[u_n u_{n-k}] = 0, \quad k = 1, 2, \dots, N-L$$

となる． $N \rightarrow \infty$ では，

$$E[u_n u_{n-k}] = \begin{cases} \hat{\sigma}_u^2, & k = 0 \\ 0, & k \neq 0 \end{cases} \tag{5.22}$$

となるから，予測誤差 u_n は白色雑音になる．

[演習 5] 平均 0，分散 1 の正規分布 $N(0, 1)$ に従う白色雑音系列を，伝達関数

$$H(z) = \frac{1}{1 - 0.98z^{-1} + 0.5z^{-2} + 0.2z^{-3}}$$

で表現されるシステムに入力した際の出力を $x_n, n = 1, \dots, T$ とする．こうした確率系列 x_n に対し，AR(N) モデル， $N = 1, 2, 3, 4$ でそれぞれ AR 係数を推定し，推定された AR 係数を持つそれぞれの AR モデルで x_n を予測した際の予測誤差 u_n が式 (5.18) を満たすことを確認せよ．

[解答] 上記のように作成された正規確率系列 $x_n, n = 1, \dots, 10000$ [s] に対し，AR(N) モデル， $N = 1, 2, 3, 4$ により，それぞれ AR 係数を推定した．推定された AR 係数をもつ AR モデルを用いて予測を行った際の予測誤差 u_n と x_n の間の相互相関関数 $E[u_n x_{n-k}]$ ， $k = 0, 1, \dots, 14, 15$ [s] を u_n, x_n のエルゴード性を仮定して時間平均を用いることにより推定した．その結果を図 5.1 に示す．推定された相互相関関数は，それ自身，確率系列であるため，合計 20 標本の u_n と x_n の組に対して同様の推定を行った．図にはそれらの結果を重ねて示してある．この結果より，AR(N) モデルを用いた場合には， $E[u_n x_{n-k}] = 0, k = 1, 2, \dots, N$ となっており，式 (5.18) が成立することがわかる．

[演習 6] 平均 0，分散 1 の正規分布 $N(0, 1)$ に従う白色雑音系列を，伝達関数

$$H(z) = \frac{1}{1 - 0.98z^{-1} + 0.5z^{-2} + 0.2z^{-3}}$$

で表現されるシステムに入力した際の出力を $x_n, n = 1, \dots, T$ とする．こうした確率系列 x_n に対し，AR(N) モデル， $N = 2, 3, 4, 5$ でそれぞれ AR 係数を推定し，推定された AR 係数を持つそれぞれの AR モデルで x_n を予測した際の誤差 u_n が式 (5.22) を満たすことを確認せよ．

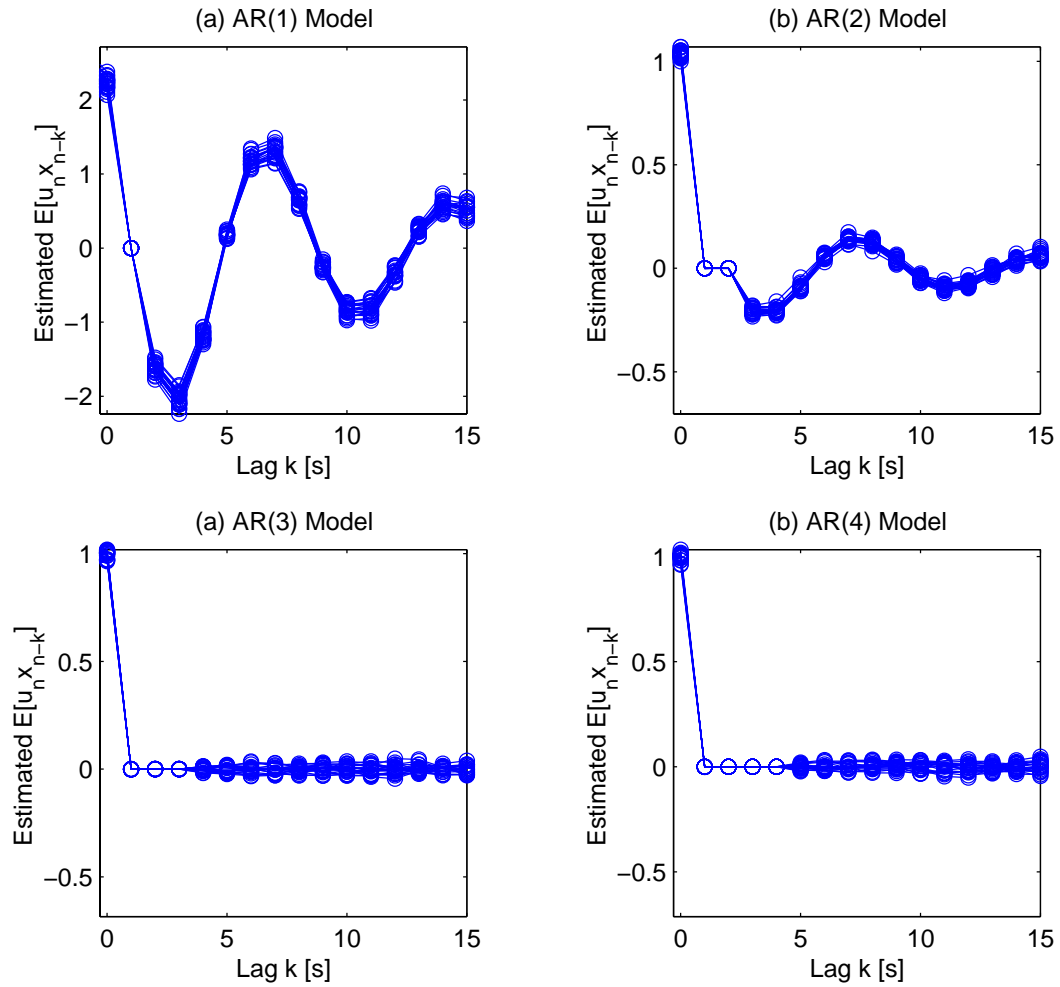


図 5.1: 演習 5 の解答 : $E[u_n x_{n-k}] = 0, k = 1, 2, \dots, N$ の検証 (真のシステム : AR(3) , $a_1 = -0.98, a_2 = 0.5, a_3 = 0.2$)

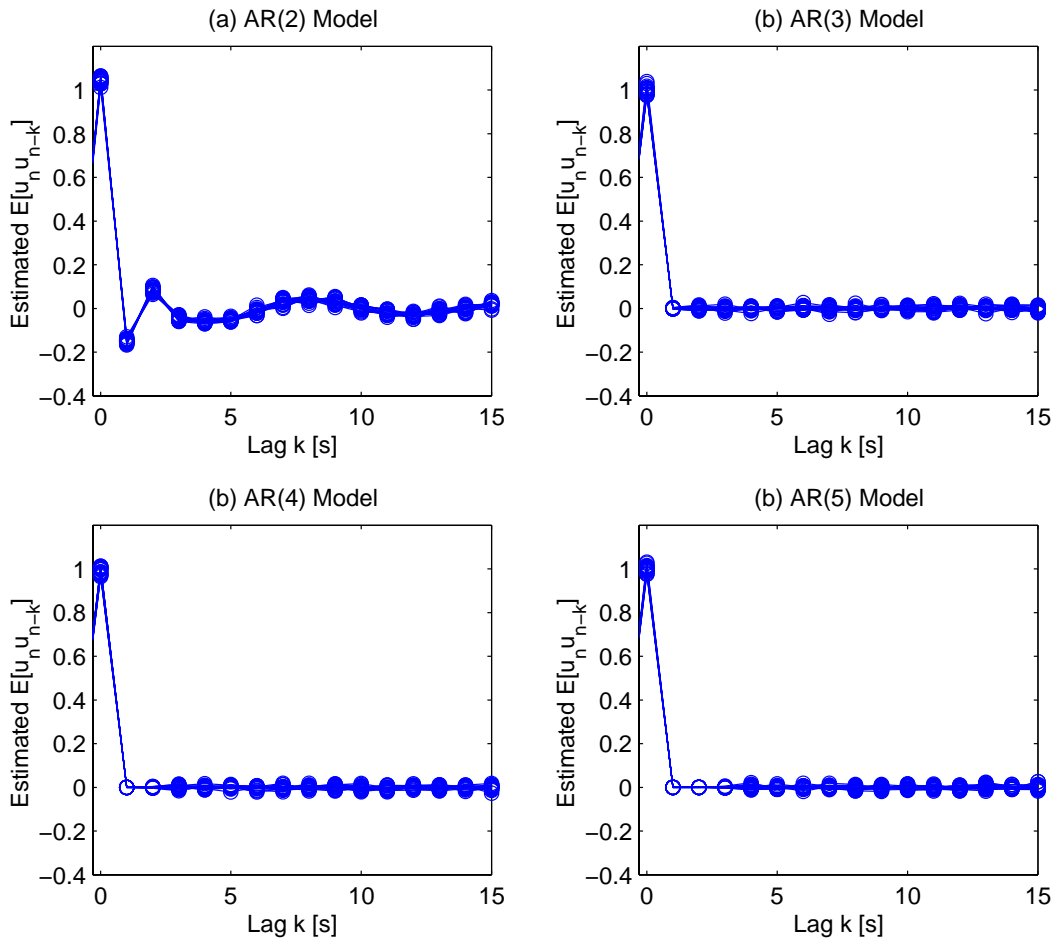


図 5.2: 演習 6 の解答 : $E[u_n u_{n-k}] = \hat{\sigma}_u^2 \delta_k$, $N \rightarrow \infty$ の検証 (真のシステム : AR(3), $a_1 = -0.98, a_2 = 0.5, a_3 = 0.2$)

[解答] 上記のように作成された正規確率系列 $x_n, n = 1, \dots, 10000$ [s] に対し, AR(N) モデル, $N = 2, 3, 4, 5$ により, それぞれ AR 係数を推定した. その際の予測誤差 u_n の自己相関関数 $E[u_n u_{n-k}]$, $k = 0, 1, \dots, 14, 15$ [s] を u_n のエルゴード性を仮定して時間平均を用いることにより推定した. その結果を図 5.2 に示す. 推定された自己相関関数は, それ自身, 確率系列であるため, 合計 20 標本の u_n に対して同様の推定を行った. 図にはそれらの結果を重ねて示してある. この結果より, AR(4) モデルを用いた場合には, $E[u_n u_{n-k}] = 0, k = 1$, AR(5) モデルを用いた場合には, $E[u_n u_{n-k}] = 0, k = 1, 2$ となっており, $N \rightarrow \infty$ で u_n が白色雑音になることが予想される.

5.3 AR 係数の推定値の統計的性質

AR 係数の推定値 \hat{a}_k の統計的性質を調べてみよう. ただし, 予測誤差 u_n は, 平均 0 の正規白色雑音系列であるものとする.

式 (5.18) より ,

$$\sum_{i=0}^N \hat{a}_i R_x(k-i) = 0 \quad \text{for } k = 1, 2, \dots, N \quad (5.23)$$

であるが , 真の AR 係数 a_k に対しては ,

$$\sum_{i=0}^N a_i R_x(k-i) = \varepsilon_k \quad \text{for } k = 1, 2, \dots, N \quad (5.24)$$

と誤差 ε_k を持つ . ただし , $\hat{a}_0 = 1$ とおいた . これら 2 つの式の差分をとれば ,

$$\sum_{i=1}^N (\hat{a}_i - a_i) R_x(k-i) = -\varepsilon_k \quad \text{for } k = 1, 2, \dots, N \quad (5.25)$$

となり , $\varepsilon \equiv (\varepsilon(1), \varepsilon(2), \dots, \varepsilon(N))^T$ とおいて , 行列で表せば

$$\mathbf{R}(\hat{\mathbf{a}} - \mathbf{a}) = -\varepsilon$$

あるいは

$$\hat{\mathbf{a}} - \mathbf{a} = -\mathbf{R}^{-1}\varepsilon \quad (5.26)$$

となる . AR 係数の推定誤差の期待値は ,

$$E[\hat{\mathbf{a}} - \mathbf{a}] = E[-\hat{\mathbf{R}}^{-1}\varepsilon] \simeq -\mathbf{R}^{-1}E[\varepsilon] \quad (5.27)$$

となり , AR 係数の推定値の真値の周りの分散共分散行列は ,

$$\text{Cov}[\hat{\mathbf{a}}] = E[(\hat{\mathbf{a}} - \mathbf{a})(\hat{\mathbf{a}} - \mathbf{a})^T] = E[\mathbf{R}^{-1}\varepsilon\varepsilon^T\mathbf{R}^{-1}] = \mathbf{R}^{-1}E[\varepsilon\varepsilon^T]\mathbf{R}^{-1} \quad (5.28)$$

となる .

$E[\varepsilon]$, $E[\varepsilon\varepsilon^T]$ を求めよう . 式 (5.24) より , $m \equiv n+k$ と変数変換し , 変形してゆくと ,

$$\begin{aligned} \varepsilon_k &= \sum_{i=0}^N a_i \frac{1}{T} \sum_{n=1}^{T-k+i} x_n x_{n+k-i} \\ &= \sum_{i=0}^N a_i \frac{1}{T} \sum_{m=k+1}^{T+i+1} x_{m-k} x_{m-i} \\ &= \frac{1}{T} \sum_{m=k+1}^{T+k+1} x_{m-k} \sum_{i=0}^N a_i x_{m-i} \\ &= \frac{1}{T} \sum_{m=k+1}^{T+k+1} x_{m-k} u_m \end{aligned} \quad (5.29)$$

となる . 上式の両辺の期待値をとると , 式 (5.18) で示した x_n と u_n が無相関になる性質より ,

$$E[\varepsilon_k] = \frac{1}{T} \sum_{m=k+1}^{T+k+1} E[x_{m-k} u_m] = 0, \quad \text{for } k = 1, 2, \dots, N \quad (5.30)$$

となる . 一方 , ε_k の自己相関関数は , 式 (5.29) より ,

$$E[\varepsilon_k \varepsilon_l] = \frac{1}{T^2} \sum_{n=k+1}^{T+k+1} \sum_{m=l+1}^{T+l+1} E[x_{n-k} u_n x_{m-l} u_m] \quad (5.31)$$

となる．上式における 4 次モーメントは， x_n, u_n が正規確率過程ならば，pair-wise rule を利用できるので，

$$\begin{aligned} E[x_{n-k}u_n x_{m-l}u_m] &= E[x_{n-k}u_n]E[x_{m-l}u_m] \\ &\quad + E[x_{n-k}x_{m-l}]E[u_n u_m] \\ &\quad + E[x_{n-k}u_m]E[x_{m-l}u_n] \end{aligned} \quad (5.32)$$

となる．右辺第 1,3 項は，式 (5.18) で示したように， $N \rightarrow \infty$ で 0 になる．一方，右辺第 2 項は，式 (5.22) で示したように， $N \rightarrow \infty$ で u_n が平均 0 の正規白色雑音系列になるから， $m = n$ の時のみ非ゼロであり， $R_x(k-l)\hat{\sigma}_u^2$ となる．したがって，式 (5.31) は

$$\begin{aligned} E[\varepsilon_k \varepsilon_l] &= \frac{1}{T^2} \sum_{n=k+1}^{T+k+1} \sum_{m=l+1}^{T+l+1} R_x(k-l)\hat{\sigma}_u^2 \delta(n-m) \\ &= \frac{\hat{\sigma}_u^2}{T} R_x(k-l), \quad \text{for } k, l = 1, 2, \dots, N \end{aligned} \quad (5.33)$$

となる．式 (5.30)，式 (5.33) を行列で表せば， $N \rightarrow \infty, N \ll T$ に対し

$$E[\varepsilon] = \mathbf{0} \quad (5.34)$$

$$E[\varepsilon \varepsilon^T] = \frac{\hat{\sigma}_u^2}{T} \mathbf{R} \quad (5.35)$$

となるので，式 (5.27)，式 (5.28) にそれぞれ代入し，また，予測誤差 u_n が平均 0 の正規白色雑音系列となるための条件が $N \rightarrow \infty$ であることを含めて考えれば， $N \ll T, N, T \rightarrow \infty$ に対し，

$$E[\hat{\mathbf{a}} - \mathbf{a}] = \mathbf{0} \quad (5.36)$$

$$\text{Cov}[\hat{\mathbf{a}}] = \mathbf{R}^{-1} \frac{\hat{\sigma}_u^2}{T} \mathbf{R} \mathbf{R}^{-1} = \mathbf{R}^{-1} \frac{\hat{\sigma}_u^2}{T} \quad (5.37)$$

となる．十分大きな次数 N の AR モデルで Yule-Walker 推定された AR 係数 \hat{a}_k は，漸近的に不偏，かつ一致推定量となっている．

[演習 7] 平均 0，分散 1 の正規分布 $N(0, 1)$ に従う白色雑音系列を，伝達関数

$$H(z) = \frac{1}{1 - 0.98z^{-1} + 0.5z^{-2} + 0.2z^{-3}}$$

で表現されるシステムに入力した際の出力を $x_n, n = 1, \dots, T$ とする．こうした確率系列 x_n を AR(3) モデルで同定した際の AR 係数の推定値の不偏性と一致性を検証せよ．

[解答] 上記のように作成された正規確率系列 $x_n, n = 1, \dots, T$ [s] に対し，AR(3) モデルを用いて AR 係数 a_1, a_2, a_3 を推定した．1000 標本の x_n に対してそれぞれ推定された AR 係数の推定値の標本平均を $T = 25, 50, 100, 200, 400, 800, 1600, 3200, 6400$ [s] に変化させて調べた．その結果を図 5.3(a) に記号 \circ で，AR 係数の真値を同図に実線で重ねて示した．これより，自己相関関数の推定を通して得られた AR 係数の推定値は漸近的な不偏性を持つことがわかる．また，1000 標本の x_n に対してそれぞれ推定された AR 係数の推定値の標本分散を同図 (b) に示す．これより，自己相関関数の推定を通して得られた AR 係数の推定値は一致性を持つことがわかる．

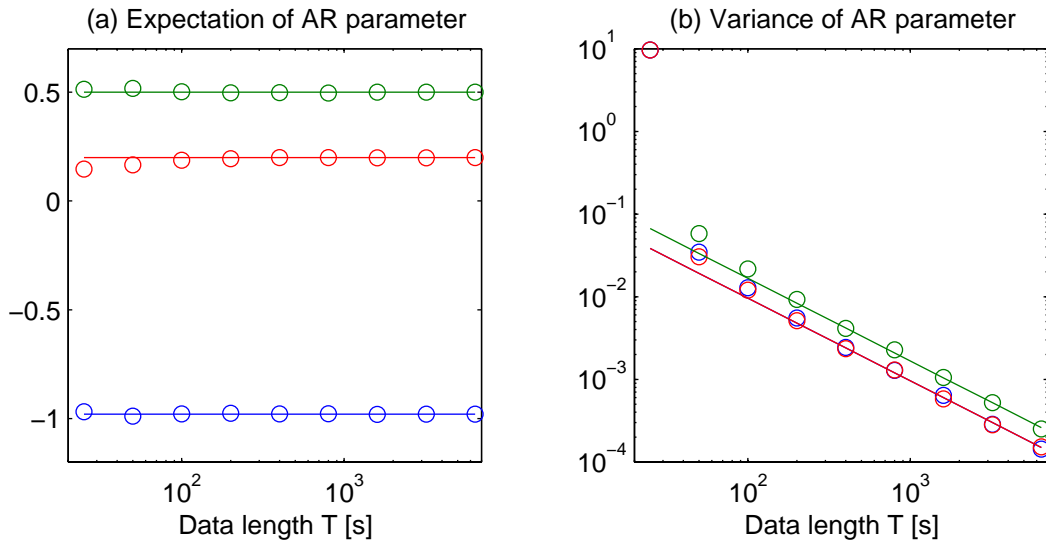


図 5.3: 演習 7 の解答: AR 係数推定値の不偏性, 一致性の検証 (真のシステム: AR(3), $a_1 = -0.98, a_2 = 0.5, a_3 = 0.2$, 青の \circ : $E[\hat{a}_1], E[\hat{a}_1^2]$ の推定値, 赤の \circ : $E[\hat{a}_2], E[\hat{a}_2^2]$ の推定値, 緑の \circ : $E[\hat{a}_3], E[\hat{a}_3^2]$ の推定値, 実線: それぞれの理論値)

5.4 線形予測モデルを用いたパワースペクトル密度の推定

離散時間で定義されたある線形時不変システムを考え, そのシステムへの入力を u_n , システムの伝達関数を $H(z)$, システムからの出力を x_n とする. u_n, x_n の z 変換をそれぞれ $U(z), X(z)$ とすれば, これらの間には,

$$x(z) = H(z)U(z)$$

もしくは, サンプリング間隔を τ として $z = e^{i\omega\tau}$ なる置換を行えば

$$x(\omega) = H(\omega)U(\omega)$$

なる関係がある. $X(\omega)$ は x_n のフーリエ変換であるから, $x_n, n = 1, 2, \dots, T$ のパワースペクトル密度 $P_x(\omega)$ は, $u_n, n = 1, 2, \dots, T$ のパワースペクトル密度を $P_U(\omega) = \frac{1}{T}|U(\omega)|^2$ とすれば,

$$P_x(\omega) = \frac{1}{T}|x(\omega)|^2 = |H(\omega)|^2 \frac{1}{T}|U(\omega)|^2 = |H(\omega)|^2 P_U(\omega)$$

として与えられることになる. 入力 u_n を平均 0, 分散 σ_u^2 の白色雑音系列とすれば, 自己相関関数は $R_U(\tau) = \delta(\tau)\sigma_u^2$ であり, そのフーリエ変換である u_n のパワースペクトル密度は, 角周波数 ω に依存せず, $P_U(\omega) = \sigma_u^2$ となる. したがって, x_n のパワースペクトル密度 $P_x(\omega)$ は,

$$P_x(\omega) = |H(\omega)|^2 \sigma_u^2 \quad (5.38)$$

として与えられる. 式 (5.3) で与えられる AR モデルは, u_n を入力として, x_n を出力する線形時不変システムとみることができる. 十分に大きな次数 N の AR モデルを用いれば, u_n を白色雑音にすることができるので, x_n のパワースペクトル密度 $P_x(\omega)$ は, 結局, 式 (5.11) で推定された AR 係数 $\hat{a}_k, k = 1, \dots, N$ を

用いて次式により推定されることになる．

$$\begin{aligned}
 P_x(\omega) &= |H(\omega)|^2 \hat{\sigma}_u^2 \\
 &= \frac{1}{\left| \sum_{k=0}^N \hat{a}_k e^{-i\omega\tau k} \right|^2} \sum_{k=0}^N \hat{a}_k R_x(k)
 \end{aligned} \tag{5.39}$$

ただし， τ はサンプリング周波数， $\hat{a}_0 = 1$ である．

このように，線形予測モデルなど，あらかじめ特定のモデルを仮定したパワースペクトル密度の推定は，AR モデルを仮定し，そのモデルパラメータである AR 係数を推定することに帰着されることから，パラメトリック推定 (parametric estimation) と呼ばれている．一方，ブラックマン・チュウキー法，ペリオドグラム法などは，特定のモデルを仮定しないことから，ノンパラメトリック推定 (non-parametric estimation) と呼ばれている．

[演習 8] 平均 0，分散 1 の正規分布 $N(0, 1)$ に従う白色雑音系列を，伝達関数

$$H(z) = \frac{1}{1 - 0.98z^{-1} + 0.5z^{-2} + 0.2z^{-3}}$$

で表現されるシステムに入力した際の出力を $x_n, n = 1, \dots, T$ とする．こうした確率系列 x_n を AR(N) モデルを用いてパワースペクトル密度を推定せよ．そして， N, T をさまざまな値に設定して推定されたそれぞれのパワースペクトル密度，ペリオドグラム法により推定されたパワースペクトル密度を比較せよ．

[解答] 上記のように作成された正規確率系列 $x_n, n = 1, \dots, T[s]$ に対し，AR(N) モデルを用いてパワースペクトル密度を推定した．($N = 3, T = 512[s]$)，($N = 3, T = 8192[s]$)，($N = 2, T = 8192[s]$)，($N = 12, T = 8192[s]$) にそれぞれ設定した際して推定されたパワースペクトル密度をそれぞれ図 5.4(a) ~ (d) に緑色の実線で示す．(e), (f) には，系列長 $T = 8192[s]$ ，および $T = 2^{18}[s]$ の x_n に対し，分割数 $K = T/1024$ としてブラックマン窓を用いてペリオドグラム法により推定されたパワースペクトルを示した．パワースペクトル密度そのものは確率変数であるため，10 標本の x_n に対してそれぞれ推定されたパワースペクトル密度を重ねて示した．また，パワースペクトル密度の真値を黒色の実線で重ねて示した．これらの結果より，系列長 T が長くなると AR 係数の推定精度が上がるため，パワースペクトル密度の推定精度も高くなることがわかる．また，推定に用いた AR モデルの次数 N が真のシステムの次数 3 よりも小さい場合には，推定された AR 係数が不偏推定量にならないため，パワースペクトル密度の推定も偏りが生じることがわかる．さらに，推定に用いた AR モデルの次数 N が真のシステムの次数よりも大きくなりすぎると，逆にパワースペクトル密度の推定の推定精度が悪くなることがわかる．真のシステムが AR モデルで表現できる場合には，ペリオドグラム法に比べ，AR モデルを用いた方がパワースペクトル密度の推定精度が高いことがわかる．

[演習 9] 母音「あ」「い」「う」「え」「お」を発音した際の音声のそれぞれのパワースペクトル密度をペリオドグラム法，AR モデルを用いて推定し，それらを比較することにより，それぞれの母音の特徴を調べよ．

[解答] 母音「あ」「い」「う」「え」「お」のパワースペクトル密度を AR(14) モデルを用いて推定した結果を図 5.5 の左列に，ペリオドグラム法により推定した結果を図 5.5 の右列に示す．左列のスペクトルの

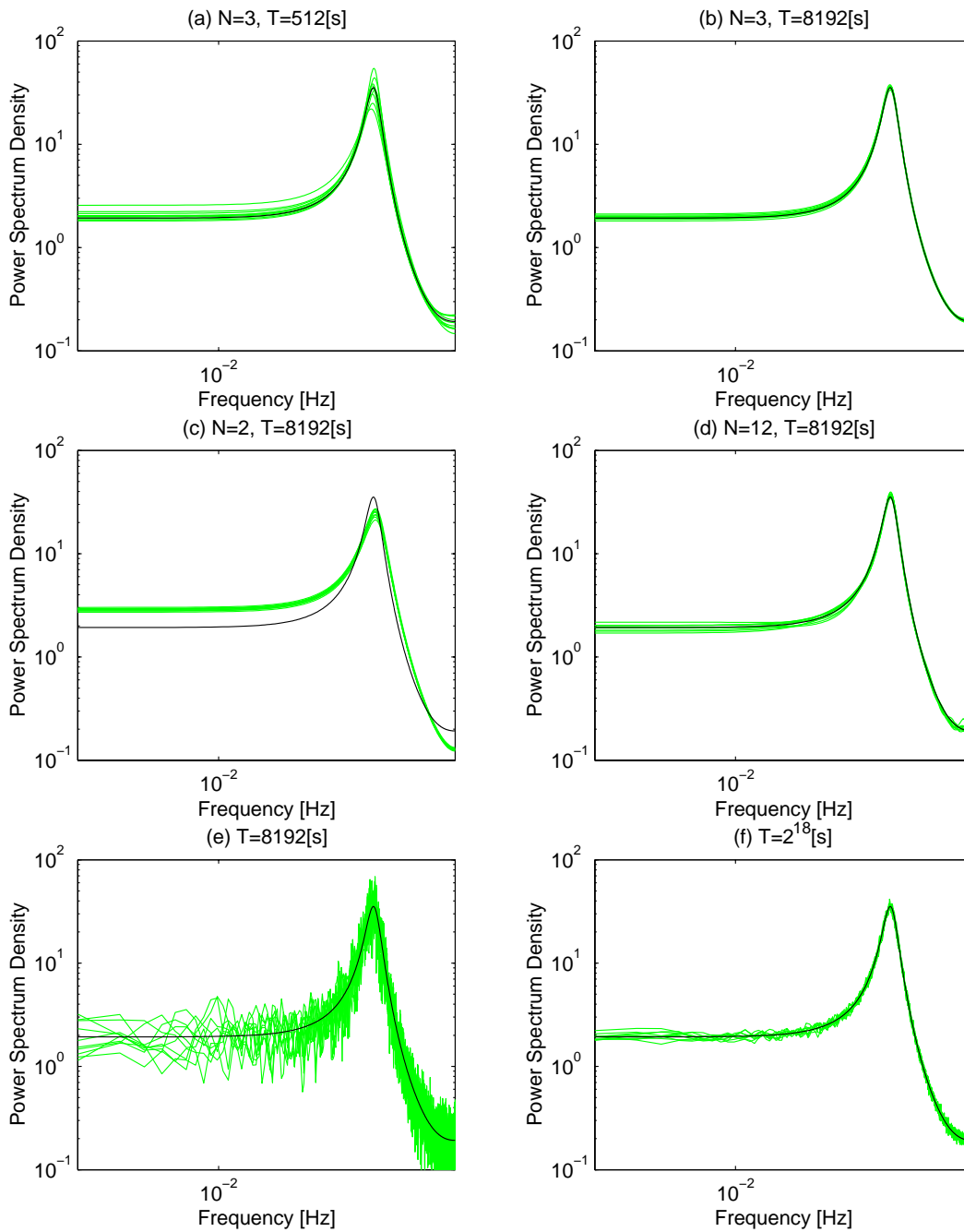


図 5.4: 演習 8 の解答 : AR(N) モデルを用いて推定されたパワースペクトル密度 ((a) ~ (d)) とペリオドグラム法により推定されたパワースペクトル密度 ((e),(f)) . 推定値 (緑色の実線) と真値 (黒色の実線) , 真のシステム : AR(3) , $a_1 = -0.98, a_2 = 0.5, a_3 = 0.2$

ピークは，低い周波数側から順に第一，第二ホルマントと呼ばれ，その周波数が母音の特徴を表している．また，右列のスペクトルのピークは，声帯で発せられた音声の基本周波数であり，ピッチと呼ばれている．AR モデルを用いたパワースペクトル密度の推定は，スペクトルのなだらかな特徴を捉えるのに向いており，一方ペリオドグラム法は，ピッチなどスペクトルのピークを捕らえるのに向いていることがわかる．

5.5 最適次数決定基準

AR モデル $AR(N)$ を用いた確率系列 $x_n, n = 1, 2, \dots, T$ の線形予測における予測誤差 $u_n, n = 1, 2, \dots, T$ は，モデル次数 N を増やせば単調に減少してゆく．しかし，この予測誤差は，AR 係数を推定する際に用いた確率系列に対する予測誤差であり，未知の確率系列，例えば， $x_n, n = T + 1, T + 2, \dots$ に対しても同様に予測誤差を小さくする保証はない．未知の確率系列に対して，予測誤差を最小にするためには，その確率系列を生成したシステムを表現するために必要かつ十分な最適なモデル次数 N^* を選択する必要がある．そのための手法として，ある種の情報量を最小化するように次数を決定する手法が提案されている．これまで提案された代表的な 3 種類の情報量の定義を紹介しよう．データ数を T ，AR モデルの次数を N ，その際の予測誤差の分散を $\sigma_u^2(N)$ とした時，情報量は，赤池情報量基準 (Akaike's information criterion; AIC) では，

$$AIC(N) = \log(\sigma_u^2(N)) + \frac{2(N+1)}{T} \quad (5.40)$$

最終予測誤差 (final prediction error criterion; FPE) では，

$$FPE(N) = \sigma_u^2(N) \left(\frac{T+N+1}{T-N-1} \right) \quad (5.41)$$

MDL (Minimum description length) では，

$$MDL(N) = \frac{T}{2} \log_2 \sigma_u^2(N) + \frac{N+1}{2} \log_2 T \quad (5.42)$$

として定義される．そして，それぞれの情報量を最小にする N を最適な次数 N^* として決定するものである．こうした手法は，必ずしも正しい次数を選択するとは限らないが¹，1 つの次数決定法として利用されている．

[演習 10] $0.9e^{\pm 0.6i}, 0.98e^{\pm 1.6i}, 0.95e^{\pm 2.5i}$ に，合計 6 個の極を持つ全極型システムを考えよう．このシステムに，平均 0，分散 1 の正規分布に従う白色雑音を入力した際の出力 $x_n, n = 1, \dots, 1000$ を求めよう． x_n を $N = 1, 2, \dots, 20$ 次の AR モデルにより同定した際の AIC, FPE, MDL の各情報量を算出せよ．その結果，各情報量を最小にする AR 次数として， $N^* = 6$ が選択されることを確かめよう．

[解答] 図 5.6(a),(b),(c) に，各 AR 次数 N に対する AIC, FPE, MDL による情報量を求めた結果の一例を示す．図からでは分かりづらいが，いずれの手法でも $N = 6$ で情報量が最小になっていることがわかる¹．

¹統計的な変動により稀に $N > 6$ の次数で情報量が最小となることもある

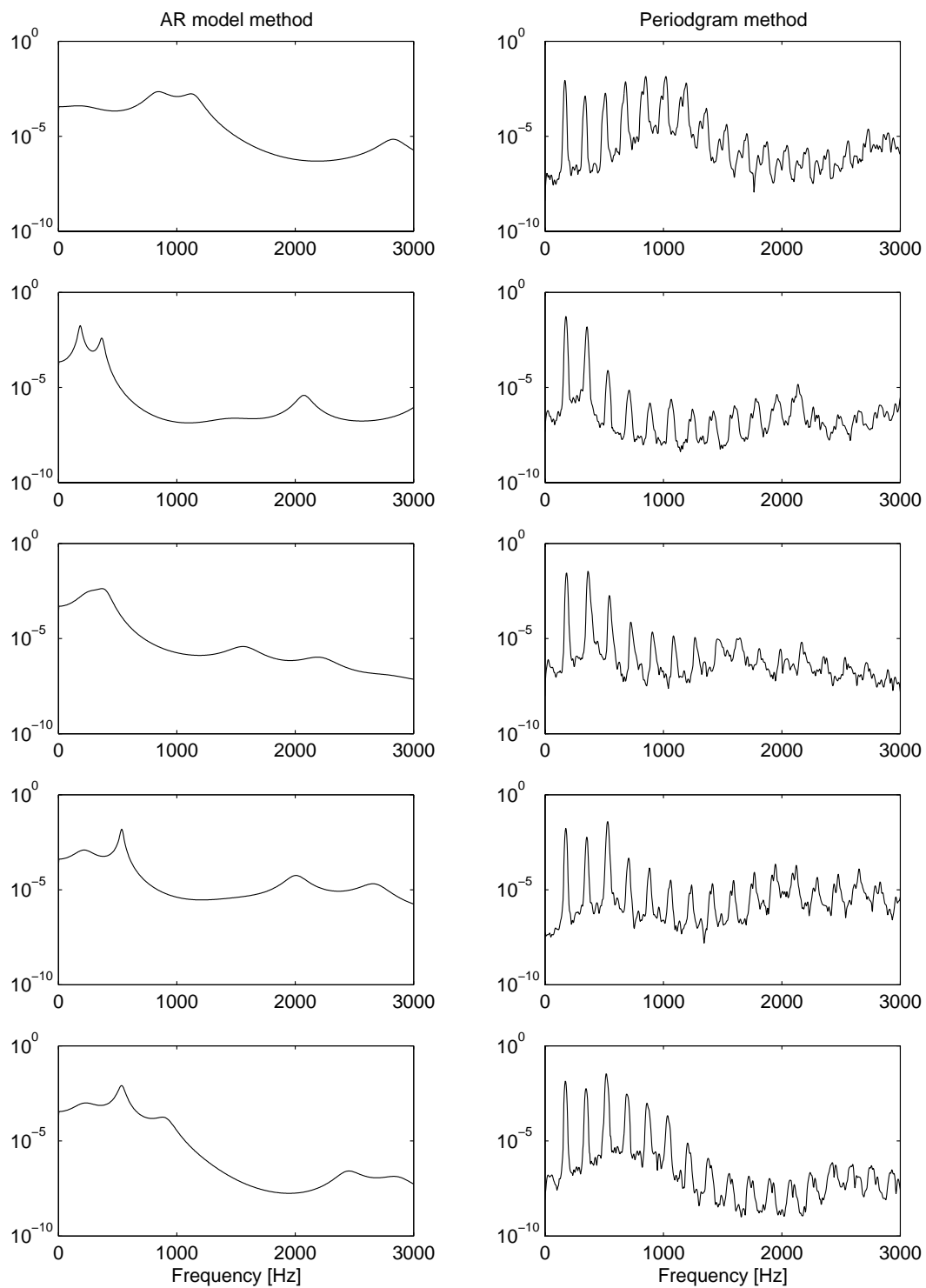


図 5.5: 演習 9 の解答：母音のパワースペクトル密度（左列：AR(14) モデル，右列：ペリオドグラム法を用いた結果，上から順に「あ」、「い」、「う」、「え」、「お」）

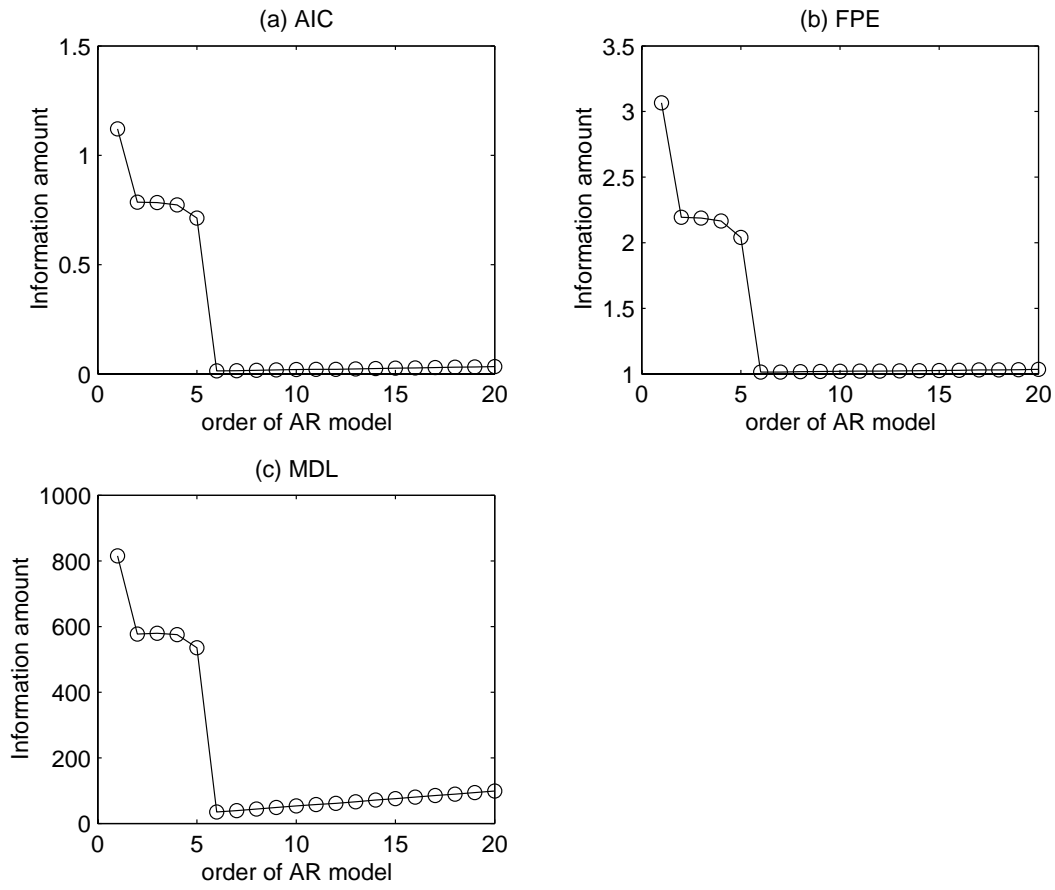


図 5.6: 演習 10 の解答 : AIC , FPE , MDL の各情報量の比較

第6章 ARモデルの係数推定法

6.1 Levinson–Durbin アルゴリズム

ある次数の AR モデルで AR 係数を推定した際、モデル次数が足りなかったことに気づき、次数をさらに増やしたい場合が往々にしてある。前述したように、AR 係数は、互いに独立変数ではないため、次数を増やすことにより、今までに推定した AR 係数もまた影響を受ける。したがって、正規方程式を用いた AR 係数推定では、再び AR 係数の推定を最初から行わなければならない。そこで、次数を増やした時、今までに推定した係数を用いて、逐次的に更新できると便利である。こうした発想から、以下説明する Levinson–Durbin アルゴリズムが提案された。

モデルの次数を明確にするため、 m 次 AR モデルの AR 係数の推定値を $a_1^{(m)}, a_2^{(m)}, \dots, a_m^{(m)}$ と表し、見やすさを考慮して、自己相関関数 $\hat{R}_x(0), \hat{R}_x(1), \dots$ を $R_x(0), R_x(1), \dots$ と表記する。AR 係数は、一般に、

$$a_j^{(m)} \neq a_j^{(n)} \quad \text{for } m \neq n \quad (6.1)$$

となることに注意すべきである。さて、 m 次 AR モデルの Yule–Walker 方程式は、

$$\begin{pmatrix} R_x(0) & \cdots & R_x(m-1) \\ \vdots & \ddots & \vdots \\ R_x(m-1) & \cdots & R_x(0) \end{pmatrix} \begin{pmatrix} a_1^{(m)} \\ \vdots \\ a_m^{(m)} \end{pmatrix} = - \begin{pmatrix} R_x(1) \\ \vdots \\ R_x(m) \end{pmatrix} \quad (6.2)$$

であるが、自己相関行列が対称 Toeplitz 行列であることを利用すれば、

$$\begin{pmatrix} R_x(0) & \cdots & R_x(m-1) \\ \vdots & \ddots & \vdots \\ R_x(m-1) & \cdots & R_x(0) \end{pmatrix} \begin{pmatrix} a_m^{(m)} \\ \vdots \\ a_1^{(m)} \end{pmatrix} = - \begin{pmatrix} R_x(m) \\ \vdots \\ R_x(1) \end{pmatrix} \quad (6.3)$$

と順番を逆にすることができる。

ここで、AR モデルの次数を増やす場合を考える。 $m+1$ 次 AR モデルの Yule–Walker 方程式は、

$$\left(\begin{array}{cccc|c} R_x(0) & R_x(1) & \cdots & R_x(m-1) & R_x(m) \\ R_x(1) & R_x(0) & \cdots & R_x(m-2) & R_x(m-1) \\ \vdots & & \ddots & & \vdots \\ R_x(m-1) & R_x(m-2) & \cdots & R_x(0) & R_x(1) \\ \hline R_x(m) & R_x(m-1) & \cdots & R_x(1) & R_x(0) \end{array} \right) \begin{pmatrix} a_1^{(m+1)} \\ a_2^{(m+1)} \\ \vdots \\ a_m^{(m+1)} \\ a_{(m+1)}^{(m+1)} \end{pmatrix} = - \begin{pmatrix} R_x(1) \\ R_x(2) \\ \vdots \\ R_x(m) \\ R_x(m+1) \end{pmatrix} \quad (6.4)$$

と表される。上式は、AR モデルの次数が増えたときの Yule–Walker 方程式の変化が分かるように、自己相関行列において新たに増えた箇所を明記した。さらに、上式を 2 つの式に分離すると

$$\begin{pmatrix} R_x(0) & \cdots & R_x(m-1) \\ \vdots & \ddots & \vdots \\ R_x(m-1) & \cdots & R_x(0) \end{pmatrix} \begin{pmatrix} a_1^{(m+1)} \\ \vdots \\ a_m^{(m+1)} \end{pmatrix} + k_{m+1} \begin{pmatrix} R_x(m) \\ \vdots \\ R_x(1) \end{pmatrix} = - \begin{pmatrix} R_x(1) \\ \vdots \\ R_x(m) \end{pmatrix} \quad (6.5)$$

$$\begin{pmatrix} R_x(m) & \cdots & R_x(1) \end{pmatrix} \begin{pmatrix} a_1^{(m+1)} \\ \vdots \\ a_m^{(m+1)} \end{pmatrix} + k_{m+1} R_x(0) = -R_x(m+1) \quad (6.6)$$

となる．ただし

$$k_{m+1} \equiv a_{m+1}^{(m+1)} \quad (6.7)$$

とし， k_m は，反射係数 (reflection coefficient)，または偏相関係数 (partial correlation coefficient : **PAR-COR**) と呼ばれる重要な係数である．さて，上式のベクトル，行列を以下の記号で表すことにする．

$$\mathbf{R} = \begin{pmatrix} R_x(0) & \cdots & R_x(m-1) \\ \vdots & \ddots & \vdots \\ R_x(m-1) & \cdots & R_x(0) \end{pmatrix}, \quad \mathbf{a}^{(m)} = \begin{pmatrix} a_1^{(m)} \\ \vdots \\ a_m^{(m)} \end{pmatrix}, \quad \mathbf{a}'^{(m)} = \begin{pmatrix} a_m^{(m)} \\ \vdots \\ a_1^{(m)} \end{pmatrix},$$

$$\mathbf{r}^{(m)} = \begin{pmatrix} R_x(1) \\ \vdots \\ R_x(m) \end{pmatrix}, \quad \mathbf{r}'^{(m)} = \begin{pmatrix} R_x(m) \\ \vdots \\ R_x(1) \end{pmatrix}, \quad \mathbf{a}^{(m+1)} = \begin{pmatrix} a_1^{(m+1)} \\ \vdots \\ a_m^{(m+1)} \end{pmatrix}$$

こうした記号を用いれば，式 (6.2)，式 (6.3)，式 (6.5)，式 (6.6) は，それぞれ次のように表される．

$$\mathbf{R}\mathbf{a}^{(m)} = -\mathbf{r}^{(m)} \quad (6.8)$$

$$\mathbf{R}\mathbf{a}'^{(m)} = -\mathbf{r}'^{(m)} \quad (6.9)$$

$$\mathbf{R}\mathbf{a}^{(m+1)} + k_{m+1}\mathbf{r}'^{(m)} = -\mathbf{r}^{(m)} \quad (6.10)$$

$$\mathbf{r}'^{(m)T}\mathbf{a}^{(m+1)} + k_{m+1}R_x(0) = -R_x(m+1) \quad (6.11)$$

さて，まず式 (6.8)，式 (6.9) を式 (6.10) に代入し， \mathbf{R} ， $\mathbf{r}^{(m)}$ ， $\mathbf{r}'^{(m)}$ を消去すれば，

$$\mathbf{a}^{(m+1)} = \mathbf{a}^{(m)} + k_{m+1}\mathbf{a}'^{(m)} \quad (6.12)$$

となる．次に，上式の k_{m+1} を求めるために，上式を式 (6.11) に代入し， $\mathbf{a}^{(m+1)}$ を消去すると

$$\mathbf{r}'^{(m)T}(\mathbf{a}^{(m)} + k_{m+1}\mathbf{a}'^{(m)}) + k_{m+1}R_x(0) = -R_x(m+1) \quad (6.13)$$

であり， k_{m+1} について解けば

$$k_{m+1} = -\frac{R_x(m+1) + \mathbf{r}'^{(m)T}\mathbf{a}^{(m)}}{\mathbf{r}'^{(m)T}\mathbf{a}'^{(m)} + R_x(0)} \quad (6.14)$$

となる．一方， m 次 AR モデルの予測誤差の分散 $\hat{\sigma}_m^2$ は，

$$\begin{aligned} \hat{\sigma}_m^2 &= R_x(0) + \sum_{k=1}^m R_x(k)\hat{a}_k^{(m)} \\ &= R_x(0) + \mathbf{r}^{(m)T}\mathbf{a}^{(m)} \\ &= R_x(0) + \mathbf{r}'^{(m)T}\mathbf{a}'^{(m)} \end{aligned} \quad (6.15)$$

となる．したがって，上式と式 (6.14) より

$$k_{m+1} = -\frac{R_x(m+1) + \mathbf{r}'^{(m)T}\mathbf{a}^{(m)}}{\hat{\sigma}_m^2} \quad (6.16)$$

となる．次に，次数を1つ増した $m+1$ 次 AR モデルの予測誤差の分散 $\hat{\sigma}_{m+1}^2$ は，式 (6.7) より

$$\begin{aligned}\hat{\sigma}_{m+1}^2 &= R_x(0) + \sum_{k=1}^{m+1} R_x(k) \hat{a}_k^{(m+1)} \\ &= R_x(0) + \sum_{k=1}^m R_x(k) \hat{a}_k^{(m+1)} + R_x(m+1) a_{m+1}^{(m+1)} \\ &= R_x(0) + \mathbf{r}^{(m)T} \mathbf{a}^{(m+1)} + R_x(m+1) k_{m+1}\end{aligned}\quad (6.17)$$

となる．そこで，上式に式 (6.12) を代入して， $\mathbf{a}^{(m+1)}$ を消去すれば

$$\begin{aligned}\hat{\sigma}_{m+1}^2 &= R_x(0) + \mathbf{r}^{(m)T} \left(\mathbf{a}^{(m)} + k_{m+1} \mathbf{a}'^{(m)} \right) + R_x(m+1) k_{m+1} \\ &= R_x(0) + \mathbf{r}^{(m)T} \mathbf{a}^{(m)} + k_{m+1} \mathbf{r}^{(m)T} \mathbf{a}'^{(m)} + R_x(m+1) k_{m+1} \\ &= \left(R_x(0) + \mathbf{r}^{(m)T} \mathbf{a}^{(m)} \right) + k_{m+1} \left(\mathbf{r}^{(m)T} \mathbf{a}'^{(m)} + R_x(m+1) \right)\end{aligned}\quad (6.18)$$

となる．上式の最後の式の括弧にそれぞれ，式 (6.15)，式 (6.16) を代入して

$$\begin{aligned}\hat{\sigma}_{m+1}^2 &= \hat{\sigma}_m^2 + k_{m+1} (-k_{m+1} \hat{\sigma}_m^2) \\ &= (1 - k_{m+1}^2) \hat{\sigma}_m^2\end{aligned}\quad (6.19)$$

となる．式 (6.16)，式 (6.12)，式 (6.19) を要素を明らかにして表すと次のようになる．

$$k_{m+1} = -\frac{1}{\hat{\sigma}_m^2} \left\{ R_x(m+1) + \begin{pmatrix} R_x(m) & \cdots & R_x(1) \end{pmatrix} \begin{pmatrix} a_1^{(m)} \\ \vdots \\ a_m^{(m)} \end{pmatrix} \right\}\quad (6.20)$$

$$\begin{pmatrix} a_1^{(m+1)} \\ \vdots \\ a_m^{(m+1)} \end{pmatrix} = \begin{pmatrix} a_1^{(m)} \\ \vdots \\ a_m^{(m)} \end{pmatrix} + k_{m+1} \begin{pmatrix} a_m^{(m)} \\ \vdots \\ a_1^{(m)} \end{pmatrix}\quad (6.21)$$

$$\hat{\sigma}_{m+1}^2 = (1 - k_{m+1}^2) \hat{\sigma}_m^2\quad (6.22)$$

これらの式が，Levinson-Durbin の漸化式である．ただし，

$$a_0^{(i)} = 1 \quad \forall i\quad (6.23)$$

$$a_{m+1}^{(m+1)} = k_{m+1}\quad (6.24)$$

であり，初期値としては

$$a_1^{(1)} = k_1 = -\frac{R_x(1)}{R_x(0)}\quad (6.25)$$

$$\hat{\sigma}_1^2 = (1 - k_1^2) R_x(0)\quad (6.26)$$

を与えておくものとする．また，得られた m 次 AR モデルが安定であるための必要十分条件は，

$$\hat{\sigma}_i^2 > 0 \quad \text{for } 1 \leq i \leq m\quad (6.27)$$

となることである．

Yule-Walker 方程式を解く際の計算量は，次数の3乗のオーダーであるが，Levinson-Durbin のアルゴリズムを用いれば，2乗のオーダーまで減少し，かなりの高速化が達成されている．しかも，計算の過程で，より低次の AR モデルの AR 係数，その際の予測誤差の分散が求められているといった利点もある．さらに，AR モデルは，フィードバックループを持つため，その安定性の判別が重要な問題となるが，Levinson-Durbin のアルゴリズムでは，予測誤差の非負性として，計算過程において随時，安定判別が行われている．

6.2 ラティスフィルタによる AR モデルの構成

Yule–Walker 方程式を，AR モデルの次数の方向に逐次的に解く Levinson–Durbin のアルゴリズムの漸近式は，式 (6.20),(6.21),(6.22) で与えられた．式 (6.21) より，Levinson–Durbin アルゴリズムでは，反射係数 k_0, k_1, \dots, k_m が決まれば，一意に m 次 AR モデルの係数が決定できることに気づく．このことを利用して，以下に述べる格子形 (ラティス) フィルタを構成する．

式 (6.21), 式 (6.24) は， $m+1 \rightarrow m$ に置き換えれば，次のように表わされる．

$$\begin{pmatrix} 1 & a_1^{(m)} & \cdots & a_m^{(m)} \\ a_m^{(m)} & \cdots & a_1^{(m)} & 1 \end{pmatrix} = \begin{pmatrix} 1 & k_m \\ k_m & 1 \end{pmatrix} \begin{pmatrix} 1 & a_1^{(m-1)} & \cdots & a_{m-1}^{(m-1)} & 0 \\ 0 & a_{m-1}^{(m-1)} & \cdots & a_1^{(m-1)} & 1 \end{pmatrix} \quad (6.28)$$

さて，

$$\begin{pmatrix} A_m(z^{-1}) \\ A'_m(z^{-1}) \end{pmatrix} \equiv \begin{pmatrix} 1 & a_1^{(m)} & \cdots & a_m^{(m)} \\ a_m^{(m)} & \cdots & a_1^{(m)} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ z^{-1} \\ \vdots \\ z^{-m} \end{pmatrix} \quad (6.29)$$

とおけば，

$$\begin{aligned} \begin{pmatrix} A_m(z^{-1}) \\ A'_m(z^{-1}) \end{pmatrix} &= \begin{pmatrix} 1 & k_m \\ k_m & 1 \end{pmatrix} \begin{pmatrix} A_{m-1}(z^{-1}) \\ z^{-1}A'_{m-1}(z^{-1}) \end{pmatrix} \\ &= \begin{pmatrix} 1 & k_m \\ k_m & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix} \begin{pmatrix} A_{m-1}(z^{-1}) \\ A'_{m-1}(z^{-1}) \end{pmatrix} \end{aligned} \quad (6.30)$$

となり，さらに，

$$\mathbf{K}_m \equiv \begin{pmatrix} 1 & k_m \\ k_m & 1 \end{pmatrix}, \quad \mathbf{Z} \equiv \begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix}$$

とおき，再帰的に代入して，次式を得る．

$$\begin{pmatrix} A_m(z^{-1}) \\ A'_m(z^{-1}) \end{pmatrix} = \mathbf{K}_m \mathbf{Z} \mathbf{K}_{m-1} \mathbf{Z} \cdots \mathbf{K}_2 \mathbf{Z} \mathbf{K}_1 \mathbf{Z} \begin{pmatrix} A_0(z^{-1}) \\ A'_0(z^{-1}) \end{pmatrix} \quad (6.31)$$

ただし，

$$\begin{pmatrix} A_0(z^{-1}) \\ A'_0(z^{-1}) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (6.32)$$

である．

さて，ここで $A_m(z^{-1})$, $A'_m(z^{-1})$ の意味を考えてみる． $A_m(z^{-1})$ は

$$\begin{aligned} A_m(z^{-1})x_n &= x_n + a_1^{(m)}x_{n-1} + \cdots + a_m^{(m)}x_{n-m} \\ &= u_n \end{aligned} \quad (6.33)$$

となり，時系列 x_n を入力とし予測誤差 u_n を出力とする AR モデルの伝達関数に他ならない．一方， $A'_m(z^{-1})$ は，AR モデルの係数 $1, a_1, a_2, \dots, a_m$ をその順番を逆にした，すなわち AR 係数 $a_m, a_{m-1}, \dots, a_1, 1$ とした AR モデルの予測誤差と考えられる．あるいは，時系列 x_n を逆方向から (後ろ向きに) 予測したときの誤差と見ることもできる．そこで，これらを区別するため，通常の AR モデルの予測誤差 u_n を u_n^f と記し，係

数の順番を逆にした AR モデルの予測誤差を u_n^r と記すことにする． $u_n^{f(m)}$, $u_n^{r(m)}$ は，式 (6.30) より

$$\begin{aligned} \begin{pmatrix} u_n^{f(m)} \\ u_n^{r(m)} \end{pmatrix} &= \begin{pmatrix} 1 & k_n \\ k_n & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix} \begin{pmatrix} u_n^{f(m-1)} \\ u_n^{r(m-1)} \end{pmatrix} \\ &= \begin{pmatrix} u_n^{f(m-1)} + k_m u_{n-1}^{r(m-1)} \\ k_m u_n^{f(m-1)} + u_{n-1}^{r(m-1)} \end{pmatrix} \end{aligned} \quad (6.34)$$

と表される．上式から分かるように， K_m は 4 端子回路を構成した際，格子形になることから，こうしたフィルタをラティスフィルタという．図 6.1 にその様子を示す．

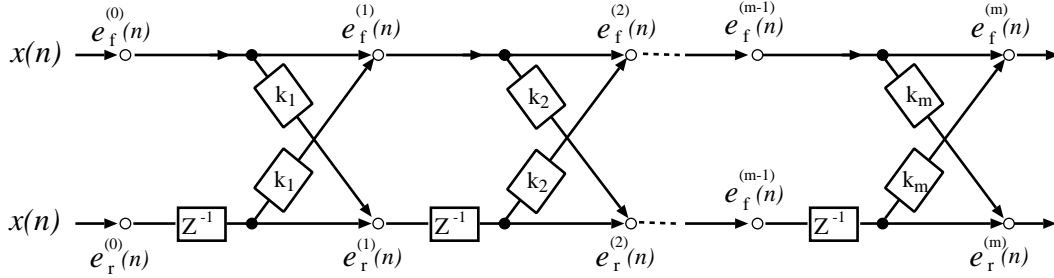


図 6.1: ラティスフィルタ

6.3 Burg 法

Yule-Walker 方程式は，(前向き) 予測誤差 u_n^f の 2 乗和を最小にすることで構成された．Yule-Walker 方程式を，その次数方向に逐次的に高速に計算するものとして Levinson-Durbin のアルゴリズムがある．Levinson-Durbin の漸化式から，AR 係数が反射係数 k_1, k_2, \dots, k_m のみから一意に決定されることを利用して，AR モデルをラティスフィルタにより構成した．その結果，後ろ向き予測誤差 u_n^r も同時に得られた．そこで次に，前向き予測誤差 u_n^f の 2 乗和と後ろ向き予測誤差 u_n^r の 2 乗和の和

$$E_m = \sum_{n=m}^{N-1} \left[\left\{ u_n^{f(m)} \right\}^2 + \left\{ u_n^{r(m)} \right\}^2 \right] \quad (6.35)$$

を最小にするように反射係数 k_1, k_2, \dots, k_m (等価的に AR 係数) を決定することを考える．式 (6.34) を上式に代入し，整理すると，

$$\begin{aligned} E_m &= \sum_{n=m}^{N-1} \left[(1 + k_m^2) \left\{ u_n^{f(m-1)} \right\}^2 + (1 + k_m^2) \left\{ u_{n-1}^{r(m-1)} \right\}^2 \right. \\ &\quad \left. + 4k_m u_n^{f(m-1)} u_{n-1}^{r(m-1)} \right] \end{aligned} \quad (6.36)$$

となり， k_m で偏微分して 0 とおき， k_m について解けば，次式が得られる．

$$k_m = - \frac{2 \sum_{n=m}^{N-1} u_n^{f(m-1)} u_{n-1}^{r(m-1)}}{\sum_{n=m}^{N-1} \left[\left\{ u_n^{f(m-1)} \right\}^2 + \left\{ u_{n-1}^{r(m-1)} \right\}^2 \right]} \quad (6.37)$$

Levinson–Durbin の漸化式 (6.20) の代わりに式 (6.37)，式 (6.22) の代わりに式 (6.34) を用いた AR 係数の推定法を **Burg 法** という。Burg 法は，Yule–Walker 法やその高速解法である Levinson–Durbin アルゴリズムとは違い，自己相関関数を推定せずに直接 AR 係数が推定できる。また，Burg 法は，Yule–Walker 法とは異なる評価値を採用しているため，一般に，得られる AR 係数も異なる。結果として得られるスペクトル推定値は，Yule–Walker 法よりも高い周波数分解能が得られるのが特徴である。しかしながら，周波数のピークが強調されすぎ，ピークが 2 つに分かれる現象 (line splitting) が起こることもある。さらに，Burg 法では，常に安定な AR モデルが得られる反面，非確率信号に対しては，バイアス誤差を含む等の問題点もある。