確率密度関数の推定としての
正規混合分布の解析とその周辺に関する研究

Research on the Gaussian Mixture distribution analysis as estimation of Probability Density Function and it's the periphery

平成２７年　３月

前橋工科大学 大学院工学研究科博士後期課程

環境・生命工学専攻

塚　越　　清

Abstract

In statistics, Mixture distribution model is a stochastic model for a measured data set to express existence of the subpopulation in a population, without requiring that the subpopulation to whom each observational data belongs should be identified.

Formally, Mixture distribution model is equivalent to expressing the probability distributions of observational data in a population.

However, it is although it is related to the problem relevant to Mixture distribution pulling out a population's characteristic out of subpopulation.

Mixture distribution model is used without subpopulation's identity information in order to make the statistical inference about the characteristic of the subpopulation who was able to give only the observational data about a population simultaneously.

Some methods of fitting Mixture distribution model to observational data contain the step considered that subpopulation's assumed identity originates in each observational data (or gravity to such subpopulation).

This paper considered these matters from the similarity of the linear combination of an element function with the estimation problem of a Probability Density Function which used the Kernel function, and the estimation problem of the Probability Density Function using a Spline function.

How to take Translate in arrangement of knots of the estimation problem of the Probability Density Function using the method of Band width picking in the estimation problem of the Probability Density Function using a Kernel function and a Spline function and Wavelets analysis and Scale has a related thing.

At the end of this doctoral thesis, Application to an analysis of the problem of resistant bacteria and the scatter situation of the pollen and a problem of quality control is described.

目　　次

記号の説明

$\hat{f}_h(x)=\frac{1}{n}\sum_{i=1}^{n}K_h(x-x_i)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right)$　Kernel 関数法による推定密度関数

$K(\cdot)$はKernel関数　　$K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right)$

h　　Band 幅　または　階級幅

$k=\log_2 n+1$　Sturges の規則による階級数

n　　データ数

IQR　　四分位範囲

混合分布の尤度関数　　$L(\theta)=\prod_{i=1}^{n}\left(\sum_{j=1}^{3}\omega_j\frac{1}{\sqrt{2\pi}\sigma_j}e^{-\frac{(x_i-\mu_j)^2}{\sigma_j^2}}\right)$

ω_j　　混合比率

混合分布の l^2 ノルム　　$l^2=\sum_{i=1}^{k}\left\{P(x_i)-\sum_{j=1}^{3}\omega_j\frac{1}{\sqrt{2\pi}\sigma_j}e^{-\frac{(x_i-\mu_j)^2}{\sigma_j^2}}\right\}^2$

m 階の差分商　　$M_{j,m}(x)=M_{j,m}(x;u_j,u_{j+1},\cdots,u_{j+m})\quad(j=0,1,\cdots,k-m)$

を m 階の B spline という。

B spline $M_{j,m}(x)$は，$u_j\le x\le u_{j+m}$ の x に対して正で，それ以外は，0 である。

標準化 B spline　　$N_{j,m}(x)=\frac{u_{j+m}-u_j}{m}M_{j,m}(x)$

nodes　　$\{\xi_{j,m}\}_{j=0}^{l}\qquad \xi_{j,m}=\frac{1}{m-1}(u_{j+1}+....+u_{j+m})$

V.D spline 関数　　$S(x)=S(x;f)=\sum_{j=0}^{l}f(\xi_{j,m})N_{j,m}(x)\quad(a\le x\le b)$

$V(f)$　　任意の $f\in[a,b]$ の開区間 (a,b) における符号変化の数

knots　　折れ線関数 $f(x)$ の knots　$\{t_i\}_{i=0}^{n}$,

選択された knots　$\{x_j\}_{j=0}^{e}$

多重度を持たせた knots $\{u_j\}_{j=0}^{k}$

ウェーブレット変換　$CWT(a,b)=\int_{-\infty}^{\infty} f(x)\frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right)dx$

ウェーブレット関数　$\psi\left(\frac{x-b}{a}\right)$

a　　**スケールパラメータ:伸縮　拡大　ダイレーション**

b　　トランスレート：平行移動，シフト　時間軸上での移動

Wavelets Power Spector　$E(a,b)=\left|CWT(a,b)\right|^2$

1. 序章

1.1 研究の背景と位置づけ

社会の情報化はコンピュータやコンピュータ・ネットワークの普及によって急速に進みつつある。それにともなって世の中には膨大な情報やデータが氾濫し，ともすれば人間をその渦の中に巻き込んでしまいがちである。それを避けるためには，ユーザにとって必要なデータや情報を見極め，その背後にある構造を適切に抽出する必要がある。

R.A.Fisher[1]は統計学の問題を次の３っに分類している。

Ⅰ 集団の研究

Ⅱ 変動の研究

Ⅲ データの簡約方法に関する研究

(有用な情報を比較的少数の数値で表す。)

また，データの簡約の際に起こる問題は

ⅰ 母集団の定式化

分布の数学的な形を選ぶ

ⅱ 推定

未知 Parameter の推定に適した統計量を標本から計算する方法を選択する。

ⅲ 標本分布

Parameter の推定値の分布や母集団の定式化が妥当かどうかの検定に用いる他の統計量の分布に関して正確な情報を数学的に導く。

としている。そのための知的な情報処理手法が，データからの情報の取得のための学習の問題として，人工知能，パターン認識，統計学などを中心とした学際領域で盛んに研究されるようになり，さらにはデータマイニングと呼ばれる一領域も形成されている。

また，統計的データ解析 ・データマイニング(統計科学)すなわちデータから有用な情報

を取り出すための数学的方法論の研究も学問分野の一つを形成されている。

多量のデータを高速なコンピュータで処理して意味のある結論を導くには,複雑な現象を確率モデルで表現する方法が有用であることが多くの分野で示されている。

この情報処理の「手法」を探求するのが統計的データ解析 ・データマイニング(統計科学)である。

近年ではゲノム科学とコンピュータ科学の融合ともいえるバイオインフォマティクスでも重要な役割を果たしている。

コンピュータに高度に依存した統計的方法の理論と実践（モデル選択，ブートストラップ法など），確率モデルによる推測，情報処理の方法論（情報量規準，情報幾何理論，確率シミュレーションAlgorithmなど）バイオインフォマティクスなどデータからいかにして有用な情報を取り出すか，というのが興味深い事柄である。

統計的データ解析 ・データマイニング(統計科学)は，数学，コンピュータサイエンス，データという三つの要素の交わるところであり，非常に魅力的な分野である。

しかしながら長い歴史の結果，「データ」や「情報」を特定の文脈に限定する弊害が目立つようになってきているというのが現状である。

これを踏まえた上で，確立した方法論から得られる有用なアイデアを継承し，かつ，これまでの枠組みにとらわれない方法論を探求していく。

近年のコンピュータによる計算環境の進歩はデータ解析の質的な変化をもたらしている。それまで時間をかけていた計算を迅速にするのだけではなく,今まで出来なかった計算が行われるようになってきた。

コンピュータのハードウエアの進歩と計算Algorithmの発展により，もはや解析的に解けるクラスに問題を限定する必要はなく，多様なモデリングが科学・工学の様々な分野で現在行われている。

この状況で必要になるデータ解析の方法論の必要性が高まっている。数理的な考え方や手法が重要になるが，数理のための数理に陥らないために常に現実世界への応用を意識し，そこから新しい問題を定式化することが必要である。

数理統計学の推論は次のような仮定と手順を踏んで行われる。

ⅰ 観測値は一定の確率分布に従ってランダムに変動する。

ⅱ 観測値の従う確率分布は一定の分布型に従うが,その中に幾つかの未知母数を含む。

iii 得られた観測値から,未知母数の推測が行われる。

このとき，数理統計学の目的は，その型,および母数の観点から母集団分布を明らかにすることである。

データ解析とは得られた データ の性質を十分に把握することにより,調査観測対象についての情報量を最大にして，調査観測対象の特性をより明確にして，重要な構造，因果関係を見附だして行くことである。(集団の規則性の探求)

このように，データ解析と１口にいっても，そこで用いられる手法は多岐にわたり，また対象とする領域も工学，農学，生物学，医学，経済学，心理学等，さまざまである。

しかし，データ解析手法の目的は

１） データ の抽出，要約

何を知りたいか，どのような結果が欲しいかの目標を設定し，その目的のためデータを収集し，必要な形に要約する。

２） データ の表現，記述

データ を解析目標にあった形で統計的に表現する。

３） データ の解析，解釈

さまざまな統計手法を用いて解析し，統計的にだけでなく・対象領域も考慮して解釈する。

ということには基本的には変化がない。

統計学（データ解析）の呼称はその立場によって，記述統計学・推測統計学・数理統計学などと幾つも存在する。また，J.W.Tukey[2]によれば データ解析は検証的データ解析（Confirmatory Data Analysis）と，探索的データ解析（Exploratory Data Analysis）に分類され，検証的データ解析は データ からあらゆる種類の統計量を計算し，その統計量の信頼性等に重点を置き，仮設の採択・棄却にしか興味を持たない。

これに対して，探索的データ解析は与えられた データ に対してその誤差を十分に考慮した上で多くの理論的モデルを設定して,そのモデルの中でどれが最適で有るかを見つけ出すため,同一データに対してさまざまな手法を適用して理論の検討を行っていく方法をとっている。

このため，探索的データ解析では，統計的な数値解析手法だけではなく，データの視覚表現手法も重要な手法である。

検証的データ解析と呼ばれる従来の統計手法（統計的検定・推定）は データが正規母集団

からのランダム抽出であることを前提として理論展開がなされている。

しかし，現実には前提である正規母集団からのランダム抽出されたデータばかりとは限らない。

この,探索的データ解析は,先入観や偏見をもたずに データの示唆するものを抽出する。そしてその示唆の中から理論的正確性を追求していく。このことは，品質管理の基本である“データ でものをいう”,“事実に基づいて管理する”という考え方に通じ，品質管理の道具の一つとして導入されてきている。その手法は，一部のデータに多少の変化があっても影響を受けることが少ない抵抗統計量 Median を用いている。

統計的データ解析を行う場合，できる限りの情報を収集し，さまざまな解析手法を用いて現象・調査観測対象の背後にある重要な構造，因果関係を見つけだしモデルを作成するのであるがその際に次のようなことに注意して欲しい。

① 調査観測対象に関して知られていること，漠然とであっても，解っていることを詳しく調べて，モデル の中にできる限り取り入れること。正しい データ は多ければ多いほど統計的に精度は上がる。

② データには誤りがある。

人間が介入すればするほど データ の誤りは多くなる。(観測ミス，転記ミス，入力ミス等の人為的なもの)

③ データ分析においては，思いつく限りのモデルを想定・計算し，その中から試行錯誤の過程により，最も良いものを選択すれば良い。(従って，コンピュータを用いた対話処理は有効な方法である。)

④ 失敗例（モデル が適合しない，用いた手法が適切な答を与えない。）も有益な情報を与える。(失敗例による反省)

⑤ 得られた結論は全て相対的な正しさを持つにすぎない。

（結果は相対的，確率的な正しさを示し，理論的，普遍的な正しさを示しているのではなく モデル は，現実の一つの近似にすぎない。）

したがって，対象分野の固有技術によって論理的な裏付けをするべきである。

一般に，統計データは次のような構造であると考えられている。

統計データ ＝ 構造（規則性・法則性） ＋ 誤差（変動・偏り・歪み）

いずれにせよ，統計的データ解析はこのような データ にもとづき対象となる母集団にたいする意味ある情報を引き出し，母集団の構造，因果関係，法則性を見つけだすことである

からデータの性格に関して正しい認識を持ち，誤差の性格に関する確認をして欲しい。

その結果,適切な解析方法をとることが可能になりモデルの正しい構造を表現できるようになる。不確実なモデルの表現は確率分布により表現し，確率分布は確率密度関数を用いて表すのが常套手段である。

このように，確率密度関数は統計的データ解析の基本的な概念である。確率変数の分布は確率密度関数,または,確率関数によって表現される。従って，統計的推測理論において確率密度関数を推定することは基礎的問題である。

未知の確率分布に従う確率変数の実現値の集合を考える。この問題は大きく分けて2つの観点から考えられ得る。

未知の確率分布に関する推論を行う方法には,未知の分布の分布族を仮定する方法(これは確率密度関数の関数型は既知であるが未知な母数を含む場合)と,分布に関する仮定を置かない方法(これは確率密度関数の関数型が未知の場合)とがある。

前者が Parametric 法,後者が Nonparametric 法である。

最尤法を代表的な例とする，Parametric 法は,未知の分布が特定の分布（たとえば正規分布）に従うと仮定して,データから期待値や分散等の分布形を決定する Parameter を推定し推論を行う。

これに対して,古くからの方法として Histogram を１つの例とする，Nonparametric 法は,観測値の分布を規定するような仮定は置かない。

Parametric 法は，新しいデータに対する確率密度の計算が比較的簡単であるが，真の分布と仮定したモデルが異なる場合には必ずしも良い推定結果が得られるとは限らない。

一方，Nonparametric 法は，真の確率密度分布がどんな関数系であっても推定できるが,新しいデータに対して確率密度を評価するための 計算量が学習用のデータ数が増えると増大していく。

Semi-parametric な手法は，Parametric モデルに基づく方法と Nonparametric な 方法の中間的な手法であり,これらの手法の良い点を取り入れ,欠点を改善するような手法である。Semi-parametric 法の代表例として， 混合分布モデル(mixture model)に基づく方法がある。

この論文で取り上げる問題はNonparametric 法と Semi-parametric 法であり,密度推定は,観測値の背後に確率密度関数が存在する（もしくは確率密度関数について分かっている部分もあるけど分からない部分もある） ことだけを仮定し,その確率密度関数を推定する。確率密度関数の推定は,与えられた観測値がどのような性質を持つものであるのかを視覚的に捕

らえ,データ解析に役立てる方法として非常に有益である。

Nonparametric 密度推定には大きく分けて 2 つの利用法が考えられる。

1 つ目は,Parametric な方法が妥当なものであるのか検証する手段としての利用。

2 つ目はデータから確率密度関数そのものを推定する手段としての利用である。

1980 年代後半から盛んになった Semi-parametric 法は,Nonparametric 密度推定の上に成り立つものである。

このように，確率密度関数の推定問題は統計的推測において興味ある問題であり，多くの研究者によってこの問題の研究が続けられてきた。また,応用面においても重要である。

Nonparametric な観点での良く知られている推定方法の手法 1 つは Histogram であるが，階級幅の設定の選択が難しいという面をもっている。Freedman&Diaconis (1981) [3], Scott(1996)[4], Sturges(1926)[5]。その他の推定方法として，Kernel 関数法，最近傍法，直交系列法等がある。

Nonparametric な観点から Kernel 関数を用いた方法を用いた確率密度関数の推定問題は，Rosenblatt(1956)[6]によって考察されて以来，Parzen (1962)[7]等によって研究されてきている。それ以降，確率密度関数の推定は多くの人々により，様々な方法によって行なわれいる。Semi-parametric 法としての混合分布モデル(特に有限混合分布) を扱う。混合分布は統計学では Pearson[8]，Newcomb[9]以来古い歴史をもち，統計学におけるさまざまな知見の積み重ねがある。

しかしながら標本として抽出したデータを見ると,教科書に出てくるような整った形で表現できる分布の母集団はほとんど見当たらないのが現状である。本研究では，確率密度関数の推定問題を，Nonparametric 法と Semi-parametric 法の両方について考察してゆきたい。

1.2 論文の構成

本論文は大きく分けて，準備の部分である第 2，3，4 章と，オリジナルの結果をまとめた第 5，6，7 章とからなる。

まず，第 2 章で確率密度関数推定問題全体についての概説を行い，Nonparametric ・Semi-parametric 法についての説明を行う。

第 3 章では Nonparametric 法についての説明と準備をし， 第 4 章では Semi-parametric 法としての有限正規混合分布についての概説を行う。第 2，3，4 章では，従来なされてきた研究を概観するとともに，後の章で必要な基本事項をまとめる。

確率密度関数の推定問題は，統計学では，その起源からある，問題で古くから Welden などにより，さまざまな研究がなされてきた。また，有限正規混合分布の問題も 19 世紀の終わりから多くの人たちによりさまざまなアプローチで研究がなされてきた。本論文では，確率密度関数の推定問題では Spline 関数による表現する方法を試み，第 6 章，第 7 章での，有限正規混合分布の問題のための入力信号として用いる。

第 5 章では，確率密度関数の推定問題のための手法として，確率密度関数を Spline 関数にって表現する方法を提案し，その Algorithm の特徴と有効性を述べていく。そのなかで，Histogram における階級数の決定，Kernel 関数法による Band 幅の決定は特に注意深い問題である。Histogram の問題点，Kernel 関数法の問題点を明記し，Spline 関数による表現法の利点を述べる。第 6 章，第 7 章で提案する Semi-parametric 法への入力信号として重要な章である。

第 6 章では，Nonparametric 法により作られた入力信号を用いて，非線形最適化の手法による有限正規混合分布の問題のための定式化を提案し，その応用例を示す。

第 7 章では，Pearson，Newcomb 以来の混合分布問題の歴史と各手法を明示し，信号解析の手法の Wavelets を用いた有限正規混合分布の問題の解析法を提案し，その特徴を示す。また，各手法の利点並びに欠点を明らかにすることにより，適用の場により利点・欠点を理解しつつ状況に応じて補完的にうまく使い分けるのが賢いあり方である。その使い分けについて，考察する。

この章では Wavelet を用いて要素分解を行うことを中心にまとめ，本論文で扱う問題点を

明確化する。

第 8 章では，結論として，確率密度関数の推定問題において，特定の関数形の重ね合せでデータ集合の分布を近似する混合分布問題としての推定法の特徴を記す。

2. 確率密度関数の推定

2.1 分類

関数型が既知で，未知なParameterを持つ分布から得られる標本に基づいて，平均・分散が未知の正規分布のように,Parameterを推定し,確率密度関数を推定する方法をParametric法という。これに対して,関数型が未知な場合の推定方法(当然，Parameterも未知)をNonparametric法という。

このNonparametric法は分布型が想定できない場合に有力な手法となる。ところで，Rosenblatt がKernel法を用いてNonparametric確率密度関数の推定問題を考えて以来，Parzen を始め,非常に多くの研究者によってこの問題の研究が続けられてきた。

また，特定の関数形の重ね合せ（混合分布）で標本データの分布を近似する，Semi-parametricな方法もある。

本論文では，Parametric法は扱わず，Nonparametric法とSemi-parametric法についてそれぞれ特徴ある手法を提案する。

2.2 特徴

2.2.1 Nonparametric 法の特徴

確率密度関数の推定問題は統計的推測において興味ある問題であり,またパターン認識などの様々な応用面においても重要である。Nonparametric法では，観測値の分布を規定するような仮定は置かない。Nonparametricな確率密度関数の推定は，観測データの背後に確率密度関数が存在することだけを仮定し，その確率密度関数を推定する。このことは，何故かというと，これらは関数の形状ではなく，モデルの複雑さを主に調整する事が目的であると

思う。

確率密度関数推定は,与えられた観測データがどのような性質を持つものであるのかを視覚的に捕らえ,データ解析に役立てる方法として非常に有益である。

この方法には大きく分けて2つの利用法が考えられる。

1つ目はParametricな方法が妥当であるのか検証する手段としての利用。

2つ目はデータから確率密度関数そのものを推定する手段としての利用である。

Nonparametric推定方法はHistogram法・Kernel関数を用いた, Kernel法・直交級数法などが提案され,研究がなされている([10],[11],[12]など)。

母集団から独立に抽出された観測値に基づいて,母集団分布に関する推論を行う際,Histogramを利用することは古典的であり, 一般的である。しかし,Histogramは一致性を持たず, Baisのある推定量である。Nonparametric-Kernel推定法はHistogramより,一般性のある確率密度関数の推定法である。

Histogram の有効性は階級幅の選択に依存する。同様に, Kernel 推定量の有効性も Histogram の階級幅に相当する'Band 幅'に依存する。Band 幅の決定方法としてさまざまな方法が考案されてきた。Histogram の階級幅の決定方法としては Sturges の公式, Scott の選択, Square-root choice, Freedman-Diaconis の選択 などがある。

母集団分布を表現する Nonparametric な従来の方法。

ⅰ Histogram

(階段関数による表現であるから０次の Spline 関数による表現方法に相当する。)

ⅱ 度数多角形

(折れ線による表現であるから１次の Spline 関数による表現方法に相当する。)

最も一般的なNonparametricな確率密度推定法であるHistogramの手順と長所・問題点は次のようになる。

(1)手順

1. 定義域を適当な階級に区切る
2. 各階級に含まれているデータの個数を数える
3. それを棒グラフに描く

(2)長所

1. Histogram がそのまま確率密度関数の代用となる

2. Histogram に登録すると元データは捨ててよい

(3)問題点

1. 始点（階級設定の際の左端）の選び方によって印象が変わる
2. 階級幅の選び方によって印象が変わる
3. $f'(x)$ などの微分や，その他の目的で使う際に有効でないことがある。
4. 多変量だと難しい。
5. Histogram は各階級の境界で不連続であり滑らかでない
6. データ数にくらべ階級数が多いとほとんどの各階級は空になる

Histogramは,データの度数分布のグラフ表示である。それは,データの全体的な特徴をつかむために確率モデルの分析に先立つ予備的なデータ解析の道具としてしばしば用いられる。例えば,Histogramの形によって,正規モデルを使うことが適切であるかどうか判断することができる。

Histogramを作るためには，データをグルーピングするための階級が事前に設定されていなければならない。階級をいかに設定するかによって，Histogramの形状は著しく異なってくる。

もしも階級の個数が多すぎれば,細部ばかりを強調することになり，その結果Histogramはデータの適切な縮約から程遠いものとなってしまう。逆に階級が少なすぎるHistogramは,平坦な特徴のない度数分布を与える。もしも各階級の幅が異なることを許せば，階級設定の問題はざらに複雑となる。

Kernel法による確率密度関数推定はHistogramの問題点を緩和する推定法である。ある点 x_0 に対する密度 $f(x_0)$ を推定することを考えてみよう。

$\{x_i\}_{i=1}^n$ を独立かつ同一な分布に従う確率変数の標本としたとき，その確率密度関数のKernel密度推定は次のようになる。

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \qquad (2.1)$$

$K(\cdot)$はKernel　　$K_h(x) = 1/h\, K(x/h)$

h はBand幅（平滑化Parameter）である。K としては，標準正規分布関数（平均が 0 で分散が1 ）を採用することが多い。

最近傍確率密度関数推定法は次のような考え方に基づいた方法である。

Kernel密度推定の欠点である「Band幅 h を小さめにすると密度の低い領域でノイズの多い推定をしてしまう」を回避するために「密度に応じてBand幅 h を設定する」.

$f(x_0)$ が大きい⇒ x_0 の近くに多くの観測値

$f(x_0)$ が小さい⇒ x_0 の近くはあまり観測値がない

したがって， $f(x_0)$ の値を推定するためには， x_0 の近くにある観測データの個数が有効な情報を持っている。

もちろん，Histogramもこの考え方に由来するものである。

Kernel密度推定などは,各データ点を中心としたKernel関数の重ね合わせのモデルを使い表現している。

最近傍確率密度関数推定法

クラスのラベルが付加された訓練事例が与えられているクラス分類の場合： 分類したい事例から近い方から順に k 個の事例を見つける。これら，k 個の事例のうち，最も多数をしめるクラスに分類する.

Kernel 密度推定法と最近傍確率密度関数推定法の長所と問題点

(1)長所

1 確率密度関数の関数形を仮定しなくてもよい

2 標本データの分布を忠実に反映する

(2)問題点

1 標本データを全て記憶しておく必要がある

2 得られる分布が，Parameter のサイズに敏感である

標本データからの区分で確率密度関数または累積分布関数の推定によって形成される確率分布（区分的分布）などがある。区分的線形分布(１次の Spline 関数)では標本データの個々のポイントで各累積分布関数値を計算して全体の累積分布関数を推定し,これらの値を線形に結合して連続的な曲線を形成する。この，区分的線形分布(１次の Spline 関数)を滑らかな曲線で表現する方法(Spline 関数)もある。本論文で提案する方法はこの Spline 関数による手法を用いる。

2.2.2 Semi-parametric な方法

特定の関数形の重ね合せ（混合分布）で標本データの分布を近似する。混合正規分布などの明示的な関数で表されたモデルを用いていない。Parametric性とNonparametric性の両方を兼ね備えていて，様々な解析手法が考えられている。

混合分布では要素分布の数を変えることによって，Parametric な性質と Nonparametric な性質を合わせ持っている。すなわち，要素分布の数を少なくすると，複雑な対象を少数の Parameter で記述するモデルになり Parametric モデルとして働く。一方，要素分布の数をサンプル数と同程度かそれ以上に増やして行くと個々のサンプルにフィットした Nonparametric な性質が現われてくる。

これは，データのもつ構造に対して大まかな視点と微細な視点とを自在に制御できるという，混合分布の柔軟性を表す性質であるといえる。

(1) 最尤法

正規分布の混合モデルで，最尤推定により分布 Parameter や混合比を決定する.

(2) 非線形最適化手法を用いる方法

非線形最適化手法により，Parameter を決定する。

(3) E-M Algorithm [22]

学習データを用いて，expectation step → maximization step を反復して，混合モデルの反復解法を与える。

・関数形の扱い易さとそれらの重ね合せによる柔軟性の両面を目指す。

(4) Wavelet 解析を用いる方法

などがある。

本論文では,非線形最適化手法を用いる方法と Wavelet 解析を用いる方法について提案し,その説明を行う。E-M Algorithm，非線形最適化手法を用いる方法は初期値が必要なため，予め要素分布の数を与える必要がある。しかし，提案する Wavelet 解析を用いる方法については初期値を与える必要がない。

3. Nonparametric 法による確率密度関数の推定法

Histogram は，最も簡単な Nonparametric な手法のひとつである。しかし，Histogram によって推定された確率密度関数は，滑らかではない。また，拡張が難しい等の問題がある。ここでは，もう少し凝った手法として，Kernel 関数に基づく方法(kernel-based methods)について紹介する。

3.1 Histogram について

Histogram は,密度関数の区分定数近似（piecewise constant aproximation）である。一般的にデータはノイズによって汚染されるため，あまりにも細かい（データへの当てはまりがより優れた）推定量は必ずしも「より優れている」というわけではない。Histogram についての階級幅の選択は，平滑化母数の選択となる。狭い階級幅はデータを未平滑化（undersmooth）する可能性があり，細かくなりすぎる。一方,階級幅が広くなると過平滑化（oversmooth）する可能性があり，それは重要な特徴を覆い隠してしまう。一般にいくつかの規則が階級幅の最適選択に利用される。これらの規則を以下に説明する。平滑化母数や,階級の中心の選択は,研究でいつも興味が持たれる難しい問題である。

最も初歩的な Nonparametric 密度推定量は Histogram である。Histogram は,データの度数分布のグラフ表示である。それは，データの全体的な特徴をつかむために，モデル解析に先立つ予備的なデータ解析の道具としてしばしば用いられる。

1. 定義域を適当な間隔に区切る $\left[x_i \le x < x_{i+1}\right)$
2. 階級 $\left[x_i \le x < x_{i+1}\right)$ に含まれているデータの個数 ν_i を数える
3. $\hat{f}(x)=\frac{\nu_i}{nh}$ n:データ数　h:階級幅　を棒グラフに描く

Histogram の問題点

1. 階級の境界の設定によって，印象が全く異なってくる。
2. 階級幅の選び方によって印象が変わる
3. $f'(x)$ などの微分や，その他の目的で使う際に有効でないことがある。

Histogram の最も重要な Parameter は階級幅(bin 幅)である。Histogram が真の分布に関して過剰に詳細な「非平滑化」，もしくは，詳細すぎる「過剰平滑化」になり，表示することの間の確率密度関数の推定における構造と誤差のトレードオフを制御する。

このように，Histogram を作るためには，データをグルーピングするための階級が事前に設定されていなければならない。階級をいかに設定するかによって,Histogram の形状は著しく異なってくる。もしも階級の個数が多すぎれば,細部(誤差)ばかりを強調することになり，その結果 Histogram はデータの適切な縮約から程遠いものとなってしまう。逆に階級が少なすぎる Histogram は，‘過剰平滑された平坦’の特徴のない度数分布を与え，かもするとデータの構造についての重要な表現までも損なう恐れがある。

3.1.1 Sturges の規則

k 個の 階級 があって，i 番目 $(0 \leqq i \leqq k-1)$ の 階級 には ${}_{k-1}C_i$ 個のデータがあるとする。このとき全てのサンプル数 n は $n=\sum_{i=0}^{k-1} {}_{k-1}C_i = \sum_{i=0}^{k-1} {}_{k-1}C_i \times 1^i \times 1^{(k-1)+i} = (1+1)^{k-1} = 2^{k-1}$

となる。 $\sum_{i=0}^{k-1} \frac{(k-1)!}{i! \times (k-1-i)!}$

あとは， $n = 2^{k-1}$ において底が 2 の対数をとると $\log_2 n = (k-1)\log_2 2 = k-1$

$$k = \log_2 n + 1 \qquad (3.1)$$

が得られる。

法則は n が 200 未満とときによく機能するが,大きな n のときに不正確であることが判明している。

Sturges の規則は左右対称を前提としているが，歪んだ場合には，3 次の平均周りの moment をもちいて，Doane[13]が次のような方法を提案している。

$$m_3=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^3 \qquad s=\sqrt{V} \qquad \sqrt{\beta_1}=m_3/s^3 \qquad \sigma\left(\beta_1\right)=\sqrt{\frac{6(n-2)}{(n+1)(n+3)}}$$

$$k=1+\log_2(n)+K_e \qquad K_e=\log_2\left(1+\frac{\left|\sqrt{\beta_1}\right|}{\sigma\left(\sqrt{\beta_1}\right)}\right) \tag{3.2}$$

3.1.2 Scott の選択

Scott の選択は正規分布に従うデータには確率密度関数の推定の平均二乗誤差を最小化するという意味では最適である。

$$MSE\left(\hat{f}(x)\right)=E\left(\hat{f}(x)-f(x)\right)^2=Var\left(\hat{f}(x)\right)+\left(E\left[\hat{f}(x)\right]-f(x)\right)^2$$
$$IMSE=\int MSE\left\{\hat{f}(x;h)\right\}dx=\int Bias\left[\hat{f}(x;h)\right]^2dx+\int Var\left[\hat{f}(x;h)\right]dx$$

漸近的な $IMSE\left(AIMSE\right)$ は h に依存することから $AIMSE\left(h\right)$ と表記すると

$$AIMSE(h)=\frac{1}{nh}+\frac{1}{12}h^2\int_{-\infty}^{\infty}f'(x)^2\,dx \tag{3.3}$$

AIMSE を最少にする h は

$$h=\left[\frac{6}{\int_{-\infty}^{\infty}f'(d)^2\,dx}\right]^{\frac{1}{3}}n^{-\frac{1}{3}} \tag{3.4}$$

もし $f(x)$ が $N\left(\mu,\sigma^2\right)$ に従うならば $\int_{-\infty}^{\infty}f'(d)^2\,dx=1\Big/\left(4\sqrt{\pi}\sigma^3\right)$ となり $h=\dfrac{3.49\sigma}{n^{1/3}}$

σ はサンプルの標準偏差
となる。

3.1.3 Freedman-Diaconis の選択

Scott の選択で 3.5σ を $2IQR(x)$ に変えればよい。

$$h=2\frac{IQR(x)}{n^{1/3}} \qquad IQR:\text{四分位範囲} \tag{3.5}$$

この方法は，分散と四分位範囲の性格の違いが適用されるため，データの外れ値に対して

敏感ではなくなる。

下に,これらの方法で同じデータに対して,階級数を決めた Histgram を 2 種類示す。1 つの例ではサンプル数は 500 で正規乱数を発生させたデータである。もう 1 つ例は,サンプル数は 1047(データについての詳細は後述する)である。また,データの単位は省略する。

この 2 つの図からも解るように,Sturges の規則を用いた方法では滑らかに表現されているように思われる。

また,データの性質によって,それぞれの方法がその特徴がある。分散が小さいときには Scott の選択も Sturges の規則と同程度の階級数になっている。データの性質を見極めて手法の選択が必要である。

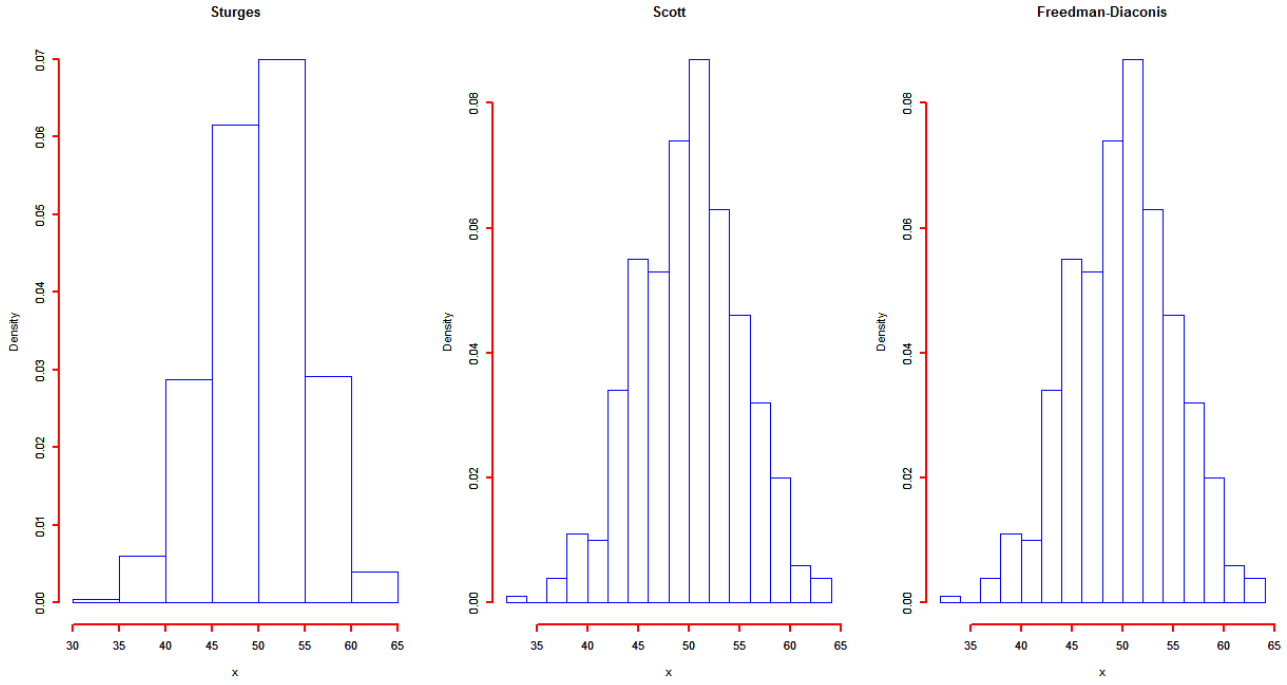


図 3.1　サンプル数 500　Sturges の規則,Scott の選択,Freedman-Diaconis の選択

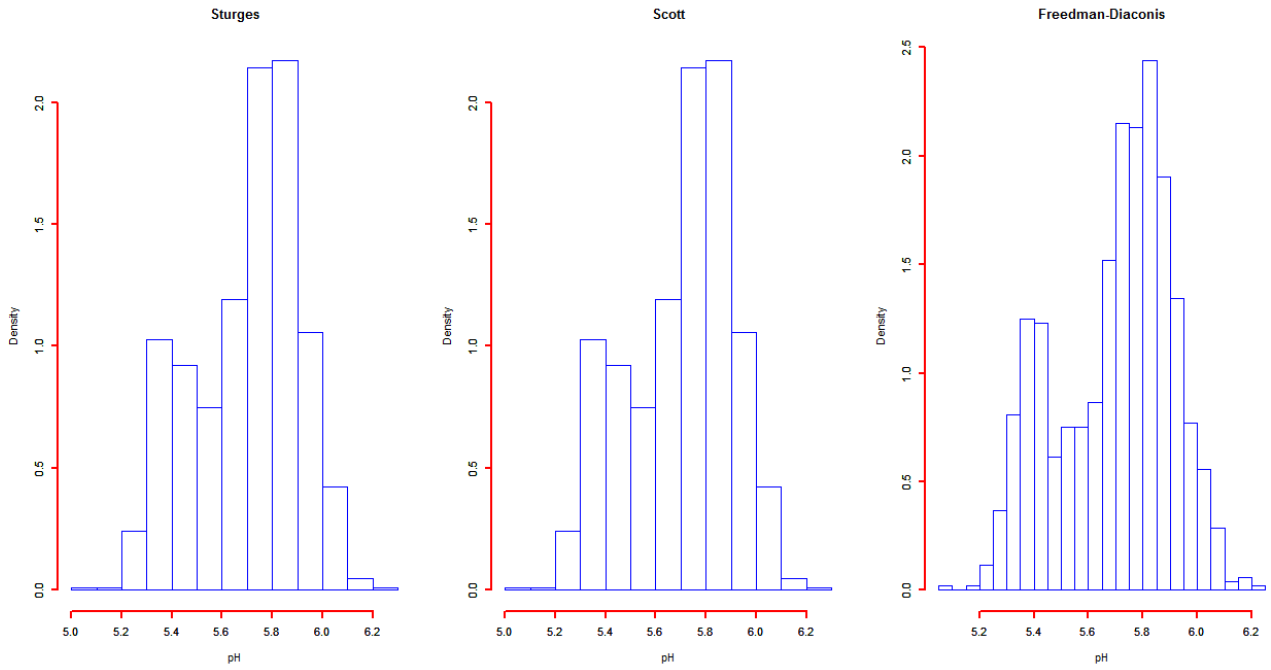


図 3.2　サンプル数 1041　Sturges の規則, Scott の選択, Freedman-Diaconis の選択

結果と，解析目的を吟味し手法の選択を行う，探索的な解析を試みることが必要である。

図 3.2 に用いたデータは第 5 章の提案する確率密度関数の推定法，第 6 章，提案する正規混合分布の解析方法 1(非線形最適化手法を用いる方法)，第 7 章，提案する正規混合分布の解析方法 2(Wavelet 解析による正規混合分布の解析方法)においても用いる。

これらの他に，根拠もなく，品質管理の教科書によく記載されている方法としてデータ数の平方根に近い整数を階級数に用いる方法などもある。

3.2 Kernel 確率密度関数推定について

階級の境界に依存せず，母集団の分布を推定できないかにこたえる方法として Kernel確率密度関数推定法がある。Kernel確率密度関数推定法はHistogramと異なり階級の境界を定める必要がない。

しかし，Histogramの階級幅と同様に，ひとつひとつの観測値の周りにいくつのブロックを積むか（Band幅）は決めなければいけない。

Kernel確率密度関数推定の結果は，Band幅の選び方に依存して大きく異なる。

Band幅の選び方に絶対的な方法はないが，ひとつの目安として

$$h = \frac{\alpha\hat{\sigma}}{n^{\frac{1}{5}}} \tag{3.4}$$

が用いられる。ここで，$\hat{\sigma}$ は標準偏差 $\hat{s}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$ と四分位範囲のいずれか小さい方を用いる。

（$\alpha=1.06$ をScottのルール，$\alpha=0.9$ をSilvermanのルール）

実際には，いろいろなBand幅を試してみて良さそうなものを選べばよい。

各観測値の周りに平たくブロックを積むのではなく，各観測値を中心とした分布を想定し，それを積み上げれば，より滑らかな形状の分布が得られる。

Kernel関数およびBand幅を決定すれば，Kernel密度関数は以下のように推定できる。

$$\hat{f}_K(x)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right) \tag{3.5}$$

ここで，h は Band幅，$K(\bullet)$ は Kernel関数である。

$\hat{f}_K(x)$が一致性を持つためにKernel関数には次のような仮定が置かれる。

仮定 $K(\bullet)$は次のような性質を持つものとする。

$$\text{（Ⅰ）}\ \int K(x)dx=1$$

$$\text{（Ⅱ）}\quad K(x)=K(-x) \tag{3.6}$$

$$\text{（Ⅲ）}\quad \int x^2K(x)dx=\kappa_2>0$$

（積分範囲は積分する変数の定義域全体とする。）

Kernel確率密度関数推定のメリット

・Histogramに比べて，分布の多峰性などの特徴がわかりやすい。

分布に峰（peak）が複数ある場合，データが分布の異なる複数の母集団から抽出されている可能性が疑われる。

・ 重ね合わせることにより，複数の分布の視覚的な比較を容易に行える。

分布が同一であるかは，Kolmogorov-Smirnov　検定により判断する。

表3.1　Kernel関数の種類

Kernel関数	関数の型
Gaussian	$K(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$
Rectangular	$K(x)=\begin{cases}\frac{1}{2} & (\lvert x\rvert\le 1)\\ 0 & (otherwise)\end{cases}$
Triangular	$K(x)=\begin{cases}\lvert x\rvert & (\lvert x\rvert\le 1)\\ 0 & (otherwise)\end{cases}$
Epanechnikov	$K(x)=\begin{cases}\frac{3}{4}(1-x^2) & (\lvert x\rvert\le 1)\\ 0 & (otherwise)\end{cases}$
Biweight	$K(x)=\begin{cases}\frac{15}{16}(1-x^2)^2 & (\lvert x\rvert\le 1)\\ 0 & (otherwise)\end{cases}$

各Kernel関数の形状を下図に示す。

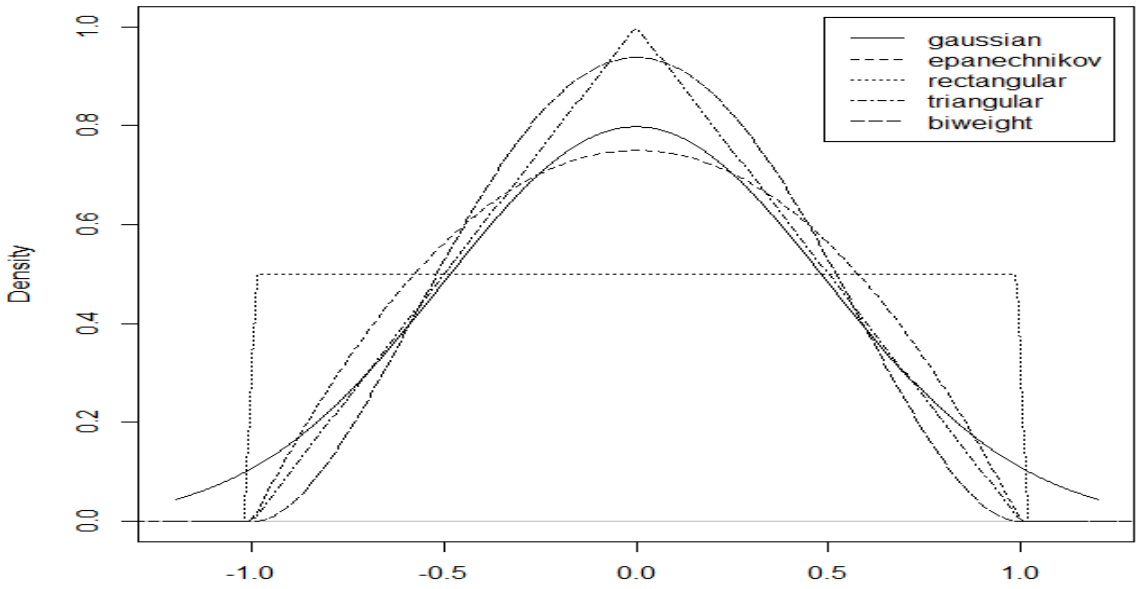


図3.3　Kernel関数の種類と形

ここではHistogram のところで示したデータを用いてGaussian Kernel関数での確率密度関数推定を示す。Kernel関数による確率密度関数の推定は，データpointを中心にこれらの

関数を重ね合わせていく方法である。Band幅が広くなればその分，確率密度関数が滑らかになるのは自明である。

Kernel 関数 $K(\bullet)$ と Band 幅 h が解析する時によって選択されるが，これらの Kernel 関数 $K(\bullet)$ 中からどれを選択しても推定の良さにはあまり影響せず，Band 幅の選択が重要な問題であることが次の例から解る。

図3.4 には データ数 500に対してBand幅 1.6040 と その$\frac{1}{3}$ のBand幅と3倍のBand幅を用いた時の確率密度関数の推定を示す。赤い線のBand幅 1.6040x$\frac{1}{3}$では変動がみられるが，3倍では滑らかに成りすぎているように思われる。

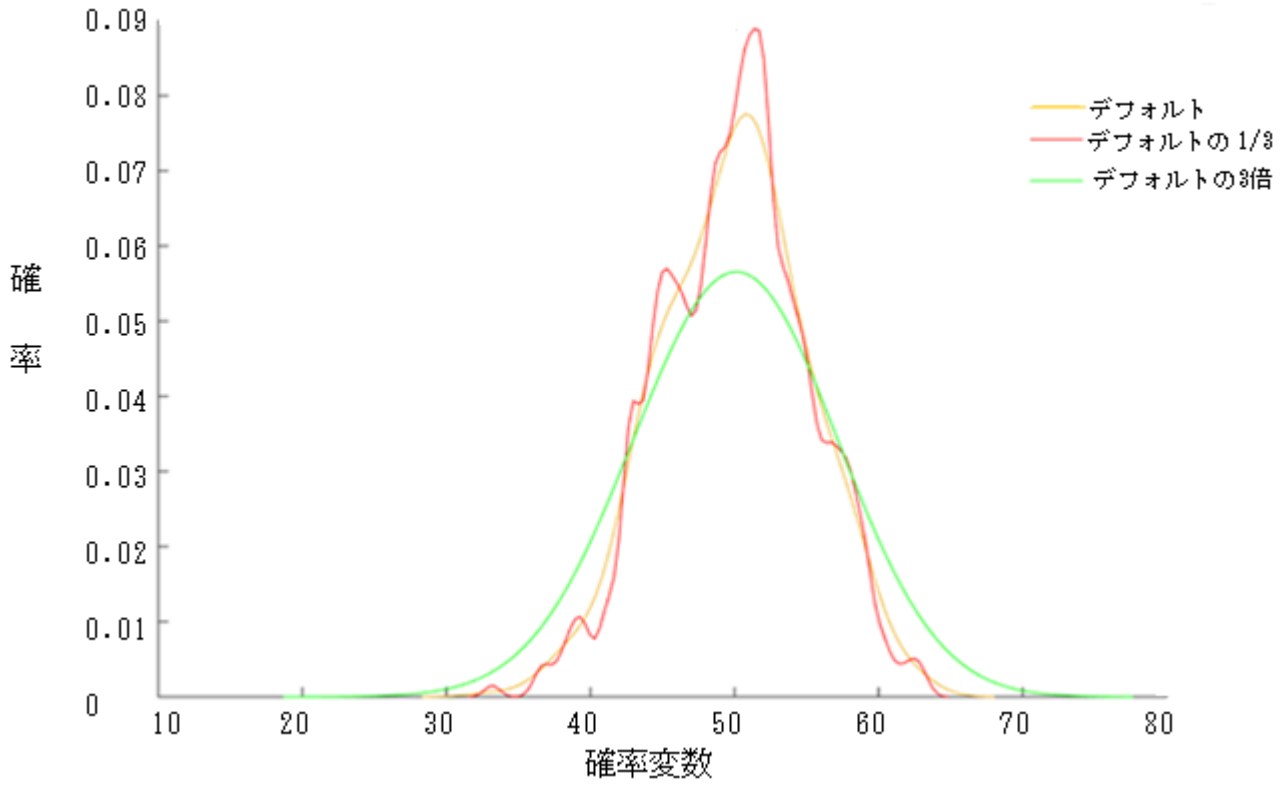


図3.4 Band幅を1.6040, 1.6040/3, 1.3x1.6040にしたときの
Gaussian Kernel関数での推定

図3.5 では，Silverman間欠泉のデータ[14]を用いた確率密度関数の推定におけるKernel関数の配置状況を下図で示す。Gaussian Kernel関数を用いた確率密度関数の推定(黒い線)をおこなった。Band幅は 10, 15, 17, 20 とした。

この図での，赤い線はKernel関数であり，その値は50倍にして表示してある。これらの赤

い線が，積み重なって確率密度関数の推定がなされる。

Band幅が小さいとKernel関数が尖り推定された確率密度関数も小さな変動を敏感にとらえていることがわかる。

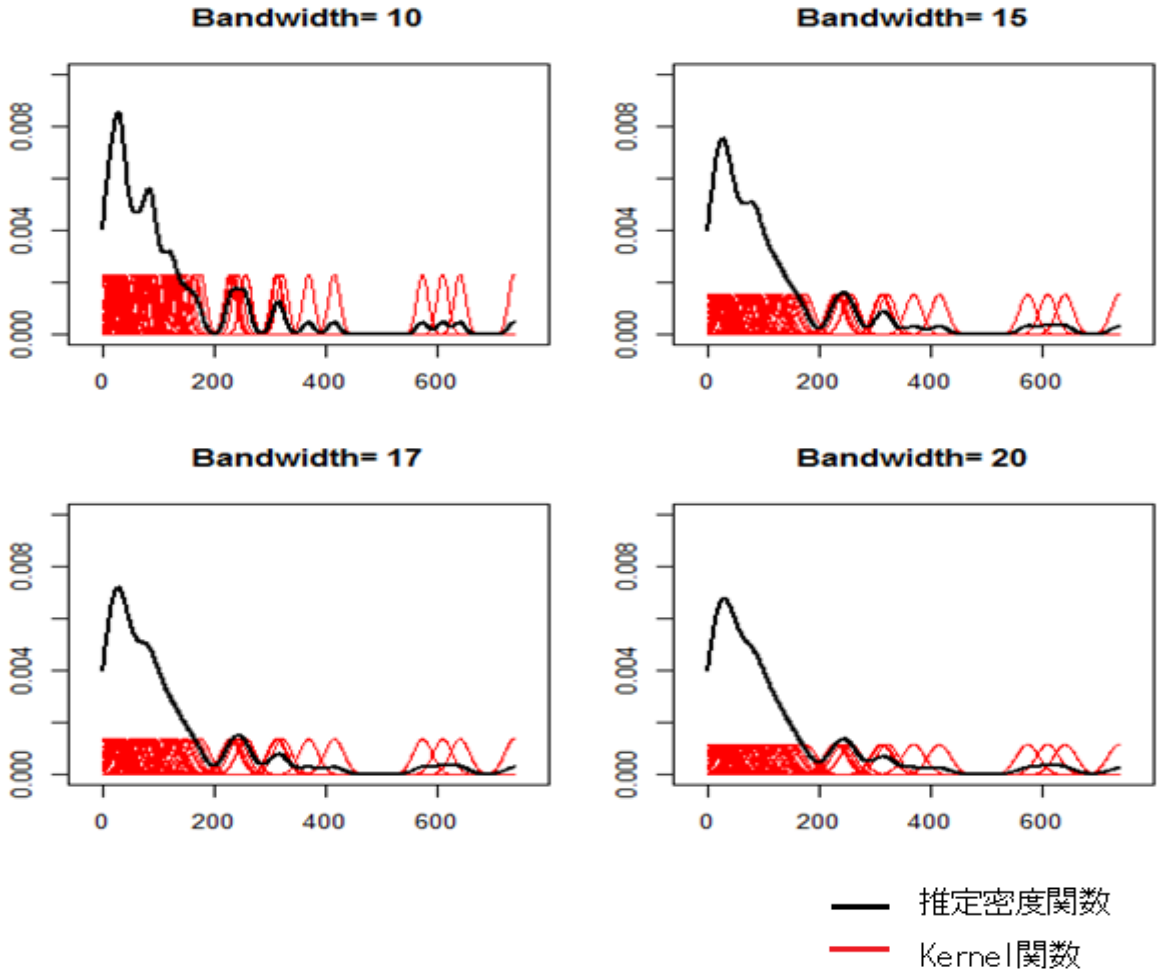


図3.5　Silverman間欠泉のデータに対するGaussian Kernel関数による確率密度関数の推定

図3.5，図3.6　ではデータの変動幅が大きく異なるため，Band幅も図3.5では大きく，図3.6ではBand幅は0.1875，　0.3126，　0.4375，　0.625と当然小さくなっている。

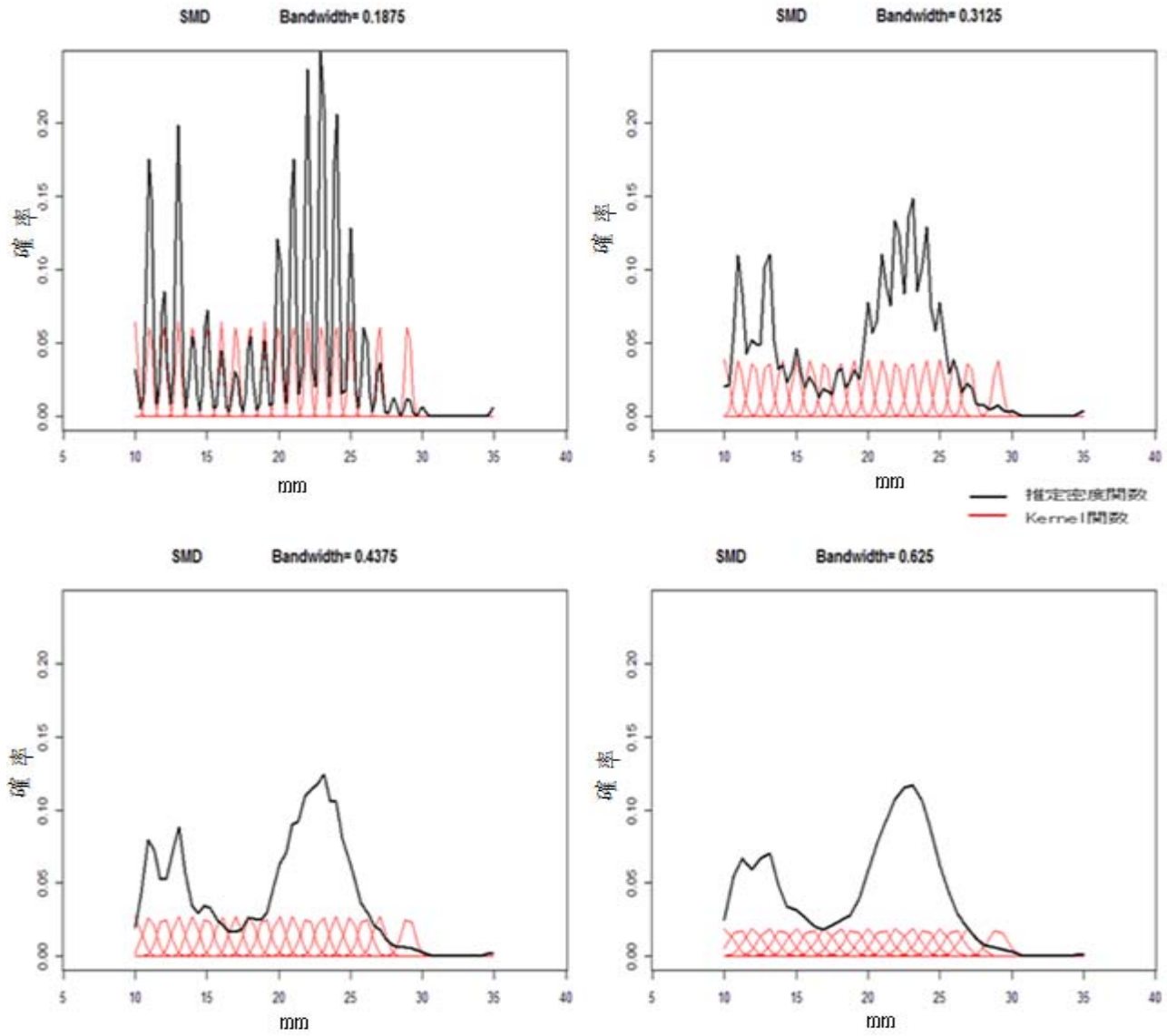


図3.6　SMに対する耐性菌のデータに対するGaussian Kernel関数
による確率密度関数の推定

また，下の3つの図はKernel関数の違いによる確率密度関数の表現を示す。この図での，Kernel関数の値は50倍にして表示してある。Band幅が小さいとKernel関数が尖り推定された確率密度関数も小さな変動を敏感にとらえている。また，下の3つの図，図3.6 図3.7 図3.8はそれぞれKernel関数にEpanechnikov Kernel関数，Biweight Kernel関数，Rectangular Kernel関数を用いてその違いによる確率密度関数の表現を示す。Band幅はGaussian Kernel関数の時と同様に　10，15，17，20 とした。

その結果，Epanechnikov Kernel関数，Biweight Kernel関数，Rectangular Kernel関数を用いいた推定はGaussian Kernel関数の時と滑らかさで大差はない。

しかし，図3.3 から見られる様にRectangular Kernel関数は角張った形状をもつため，Rectangular Kernel関数を用いた推定においては刺々しい形になっている。

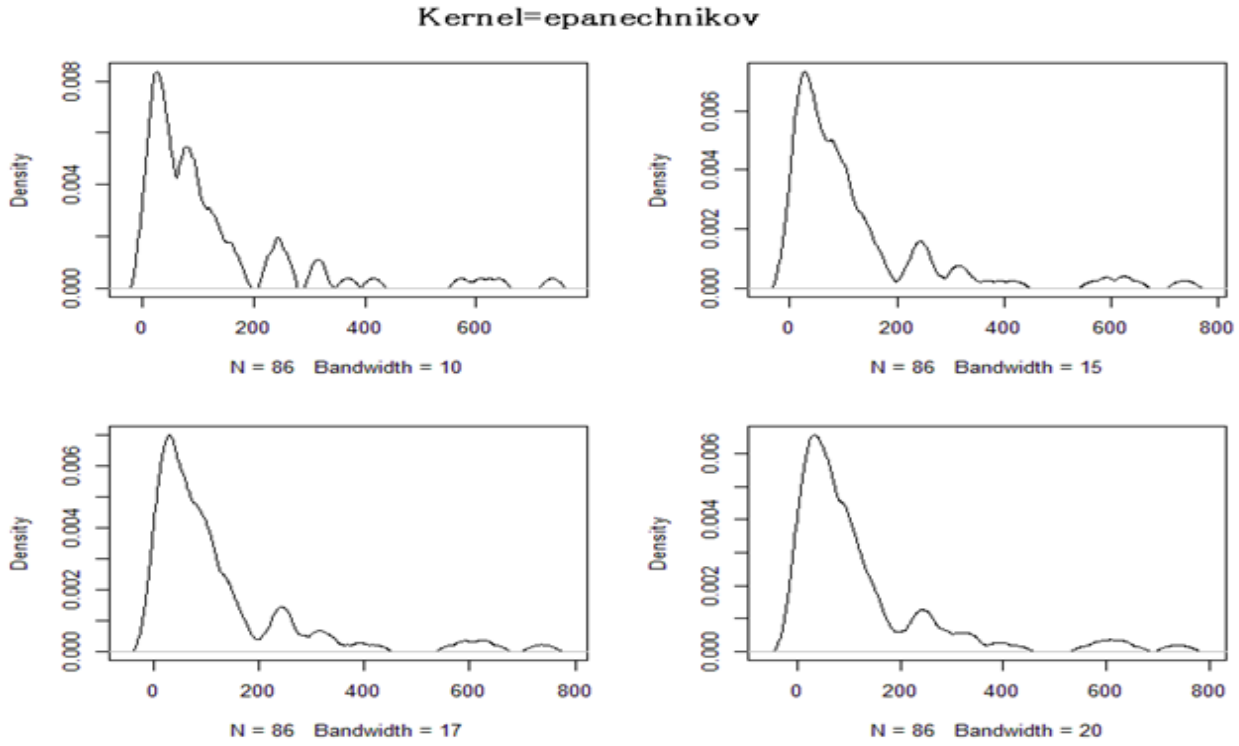


図3.7 Epanechnikov Kernel関数

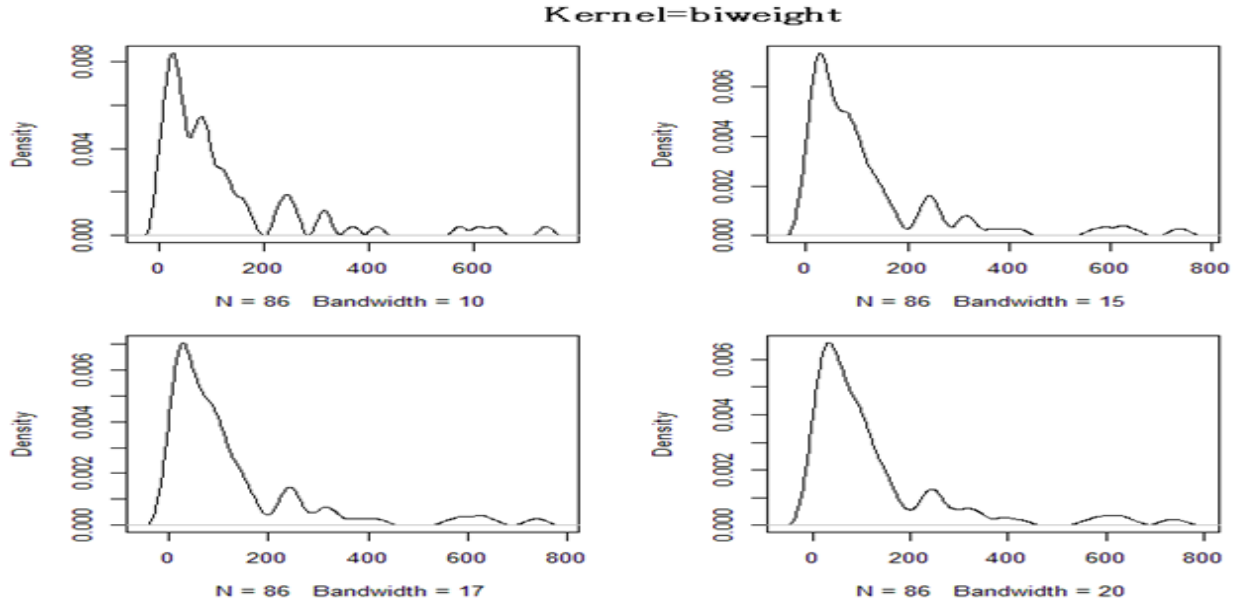


図3.8 Biweight Kernel関数

これらの図から，確率密度関数の推定においては，Kernel関数の違いよりBand幅の違いが表情の大きな変化をもたらしているのが理解できる。

どの，Kernel 関数を用いるか，Band 幅をいくつにするか，確率密度関数の推定においては決めなくてはいけないことがいくつもある。そのためには，会話的に，結果と，解析目的を吟味し手法の選択を行う，探索的な解析を行うことが必要である。

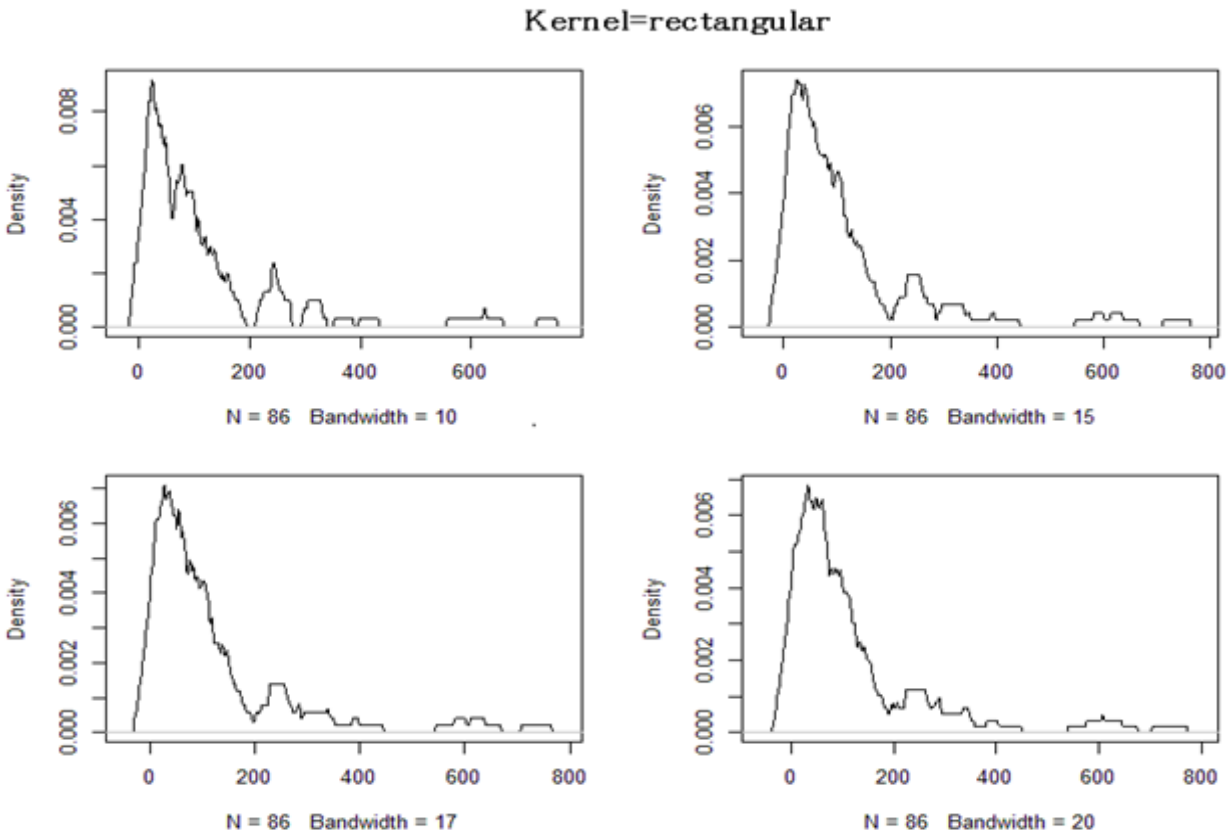


図3.9 Rectangular Kernel関数

Rectangular Kernel関数は他のKernel関数に比べて滑らかさに欠けるが，他のKernel関数はその変動に大きな差はない。

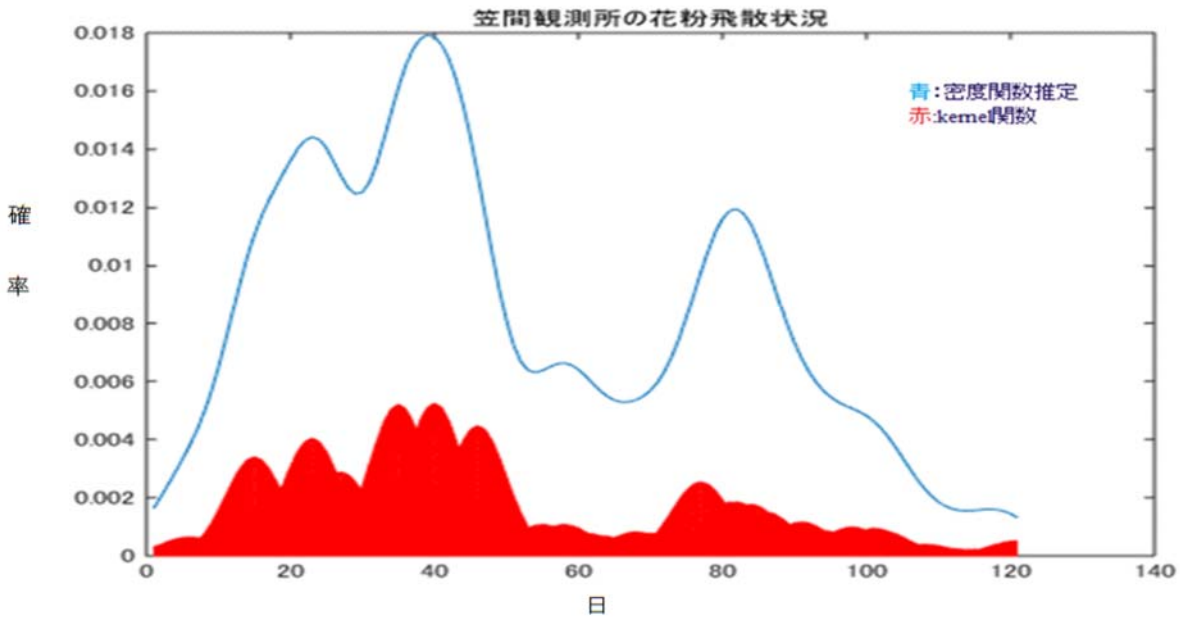


図 3.10　笠間観測所　花粉飛散データ　Bandwidth=4

(赤色:Kernel 関数を積み重ねたもの　2294 個のデータ)

図 3.5・3.6 に見られる kernel 関数を積み重ねたものが図 3.10 の赤い色の線である。Kernel 関数法ではすべてのデータを用いなければ推定密度関数の再計算はできない。

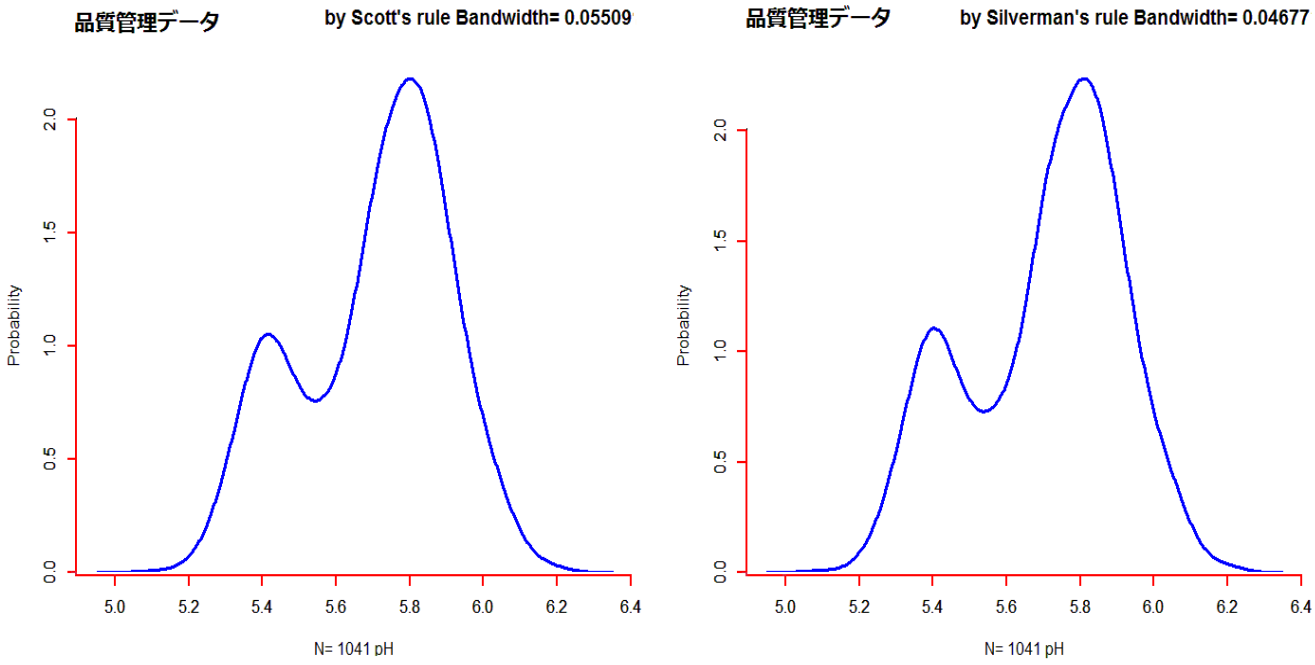


図3.11　1041個データに(3.4)のα=1.06(Scottのルール)
α=0.9(Silvermanのルール)を用いての比較

ScottのルールとSilvermanのルールではBand幅が0.00832の違いであるが表現した確率密度関数では5.5近傍の窪みの違いがみられるが他は粗変化はない。

4. Semi-parametric な推定方法 (混合モデルを用いる推定方法)

有限の混合分布モデルの使用に関する最初の主な分析は Newcomb(1889)[10]によるものや, Welden(1892 と 1893)[16]によって提供されるあるデータに, 2 つの正規分布の確率密度関数の混合分布の適合を Pearson(1894)[9]によって試みられた論文がある。

Pearson によって分析されたデータ集合は, ナポリ湾からサンプリングされた n = 1000 のカニの体長に対する額の比率上の測定から成った。

Welden は, これらのデータの Histogram 中の不調和がこの母集団が 2 つの新しい亜種の方へ発展させていた信号かもしれないと推測した。

Pearson は, 優れた適合を得るために彼が開発した Moment 法を使用し, カニの 2 つの種があったという証拠として 2 つの要素の存在を解釈した。

4.1 混合モデル

ここでは 2 つ, または 3 つの正規分布の混合分布をその成分要素に分離する問題として定式化を示す。

従って, 問題は密度関数 $f(x)$ を次のような形で推定することである。

$$f(x)=\omega_1\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{\sigma_1^2}}+\omega_2\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(x-\mu_2)^2}{\sigma_2^2}}+\omega_3\frac{1}{\sqrt{2\pi}\sigma_3}e^{-\frac{(x-\mu_3)^2}{\sigma_3^2}} \qquad (4.1)$$

において $\omega_1,\omega_2,\omega_3\ (\sum\omega_i=1)$, μ_1,μ_2,μ_3, $\sigma_1,\sigma_2,\sigma_3$ を推定する。

ただし

ω_i: 各要素分布の混合率　μ_i: 各要素分布の平均　σ_i: 各要素分布の標準偏差

統計的な Parameter の推定法としては, Moment 法, 最尤推定法, 最小二乗近似などが考えられる。

尤度関数を求めると

$$L(\theta)=\prod_{i=1}^{n}\left(\sum_{j=1}^{3}\omega_j\frac{1}{\sqrt{2\pi}\sigma_j}e^{-\frac{(x_i-\mu_j)^2}{\sigma_j^2}}\right) \tag{4.2}$$

となる。これを最大になるようにParameterを求める。

l^2 ノルムを求めると

$$l^2=\sum_{i=1}^{k}\left\{P(x_i)-\sum_{j=1}^{3}\omega_j\frac{1}{\sqrt{2\pi}\sigma_j}e^{-\frac{(x_i-\mu_j)^2}{\sigma_j^2}}\right\}^2 \tag{4.3}$$

となり,これを最小にするように Parameter を求める。

また,B.S. Everitt,D.J.Hand[17],C.G.Bhattacharya[18],G.D.Murray,D.M.Tttterington[19],E.A.C.Thomas[20],D.M.Titterington,A.F.M.Smith and U.E.Markov[21]などによる様々な方法がある。

4.2 E-M Algorithm

最尤推定法では確率の積となるデータ数 k が大きくなると,その積は限りなく 0 に近づいてしまうのでこのままの形では数値計算に不向きである，それを解消するために E-M Algorithm[22]を扱う。

4.2.1 E-M Algorithm とその特徴

「一度に計算できないなら，徐々に正解に近づけていこう」というのが，E-M Algorithm の基本的な考えである。観測できない隠れた Parameter(隠れ変数*)が存在する時に最尤推定を行うための汎用手法であり,混合分布以外にも隠れマルコフモデルやグラフィカルモデルの学習に応用さている。Newton 法(あるいは Fisher のスコアリング法)勾配法と同様，反復法によって局所最適解を求める Algorithm である。

・ 尤度が単調に増加することが保障されており，Algorithm の振る舞いが安定している。

混合分布では尤度が無限大になる無意味な解が存在するので，Algorithm の安定性は重要。

・ 速度に関しても収束の初期の段階では Newton 法と同程度の速さになることが知られている。

・ インプリテーションが簡単になることが多い。また，これと関係して 1 ステップに要する計算量が減らせる場合もある。Newton 法では尤度の Hessian を計算する必要があるが，混合分布などでは一般に複雑な形になり，多くの計算量を必要とする。

E-M Algorithm は，データに欠測値が存在した場合に，観測データと隠れ変数からなる完全データを考え，完全データの尤度関数の条件付き期待値を計算し，Parameter の最尤推定を行う方法である。E-M Algorithm には完全データの尤度関数の条件付き期待値を計算する E-step と最尤推定法を行う M-step がある。

E-M Algorithm の各々の繰り返しによって，尤度が単調に増加することが証明されている。従って，局所的には最適解に収束し，少なくとも初期解よりは良好な大域的収束性が経験的に知られている。

ただし，最初のうちは速い収束を示すが，収束の後期では遅くなるといわれており，E-step や M-step が必ずしも容易に実行できないという問題も存在する。

混合分布の場合，各データxが何番目かのクラスタから発生したかがわかると，Parameter 推定は各クラスタに属するデータだけ集めて行えばよい。

4.2.2 E-M Algorithm

E-M AlgorithmはParameterをある適当な初期値に設定し,Eステップ(Expectation step)とM ステップ(Maximization step)と呼ばれる二つの手続きを繰り返すことにより θ の値を逐次更新する方法であり，次のように定式化される。

1. Parameterの初期値を適当な点_$\theta = \theta^0$　にとる。

2. $p=0,1,2,\cdots$ に対して次の二つのステップを繰り返す。

(a) Eステップ: 完全データの対数尤度 $\log f(x|\theta)$ の データ y とParameter $\theta^{(p)}$ に関する条件つき平均を求める。

つまり

$$Q(\theta)=E\left[\log f(x|\theta)|y,\theta^{(p)}\right]=\int f\left(x|y,\theta^{(p)}\right)\log f(x|\theta)dx \qquad (4.4)$$

を計算する。 (Parameter固定の下で隠れ変数の分布について最尤推定)

(b) Mステップ: $Q(\theta)$を最大化する θ を $\theta^{(p+1)}$ とおく。

なお, 不完全データ y が与えられたときの完全データ x の条件つき分布はBayesの公式から

$$f(x|y,\theta)=\begin{cases} f(x|\theta)\Big/g(y|\theta) & ,x\in X(y) \\ 0 & x\notin X(y) \end{cases} \qquad (4.5)$$

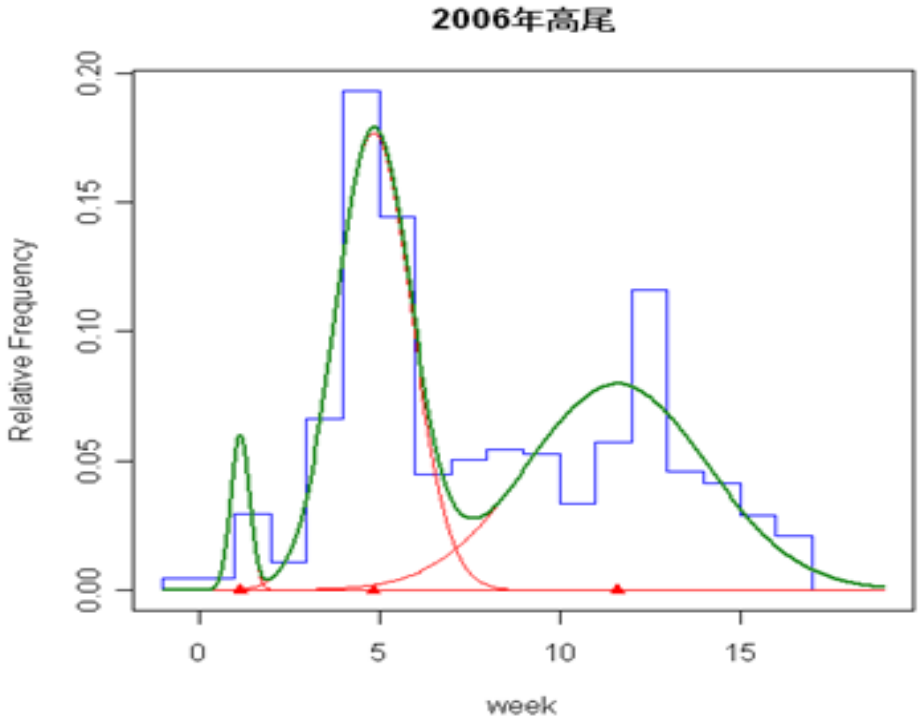
で与えられる。 (求めた隠れ変数の分布の下で Parameter について最尤推定)

E-step で行っていることは, θ を固定して, 尤度を最大にする隠れ変数を求めることに対応し, M-step は E-step で得られた隠れ変数を固定して, 尤度を最大にする θ を求めることに対応する。

*隠れ変数(潜在変数)・・・サンプリングによってその値が観測されることはないが, モデル中には存在する変数。

Algorithmの各々の繰り返しによって, 尤度が単調に増加することが証明されている。従って, 局所的には最適解に収束し, 少なくとも初期解よりはよい解が得られる。もちろん一般に大域的に収束する保証はないが,多くの応用例で良好な大域的収束性が経験的に知られている。

ただし, 最初のうちは速い収束を示すが, 収束の後期では遅くなると言われており, Eステップ やMステップが必ずしも容易に実行できないという問題も存在する。これらの記述から解るように, 当然, 要素数と各要素のParameter, 及び混合比率を初期値として与えなければならない。



4.2.3　E-M Algorithm の適用例

ここでは，花粉の飛散状況の分布データをもちいたE-M Algorithmによる計算結果だけを示す。環境省が発表している花粉飛散状況は2月に始まり，5月に観測の表示が終わり，関東地方の花粉はスギ花粉(前半)，ヒノキ花粉(後半)と中間に黄砂等が混ざり興味深い分布状況をしている。

ここでは，高尾の観測所における2006年の飛散状況を提示する。

表 4.5 2006 年高尾の解析結果表

高尾 2006	第一分布	第二分布	第三分布
混合率	0.038	0.463	0.499
平均	1.165	4.851	11.615
標準偏差	0.251	1.044	2.500

図4.5高尾観測所の花粉飛散データ

図4.5ではHistogram(青)と要素分布(赤)，合成された混合分布(緑)を表示している。

2月(2,3週)初めに小さな山があり，2月末から3月初頭(4,5週)に杉花粉のピークを迎え，4月末から5月初頭檜花粉(12,13週) のピークを迎えている。

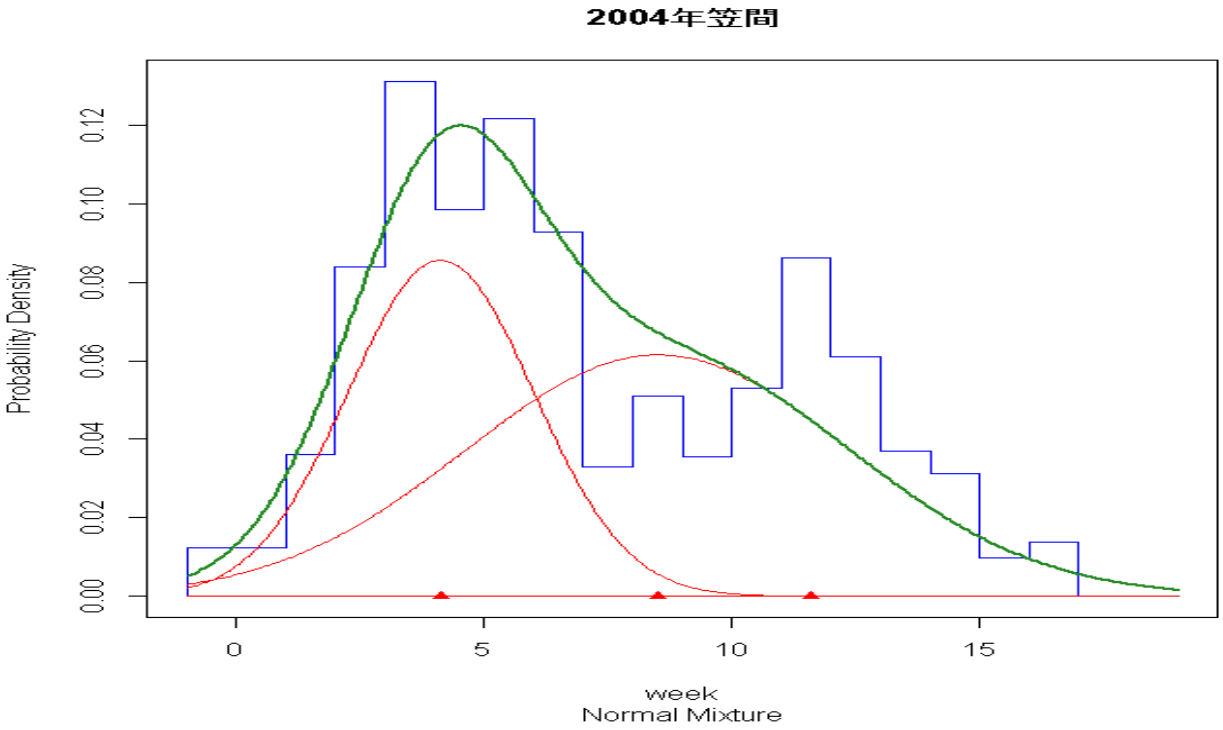


図4.6　2004笠間観測所の花粉飛散データ

表 4.6 2004 年笠間の解析結果表

高尾 2006	第一分布	第二分布	第三分布
混合率	0.402	0.597	0.000
平均	4.126	8.511	11.587
標準偏差	1.879	1.0442	5.276

第三分布は混合率が殆ど0である。

5. 提案する確率密度関数の推定法 (Variation Diminishing Spline 関数表現による確率密度関数の推定)

本章では，確率密度関数の Variation Diminishing Spline 関数表現と，その確率密度関数の特性関数を Spline 関数の knots と node で表現して，第 6 章・第 7 章のための入力信号として用いるための準備を行う。さらに，knots の選択法により，R.A.Fisher のいう統計学の問題における，有用な情報を比較的少数の数値で表すという “Ⅲ データの簡約方法に関する研究”への貢献がなされるような方法を考える。

5.1 区分的線形分布を滑らかな曲線で表現する方法

滑らかな，曲線を表現する方法として Spline 関数(T.N.E. Greville[23], J.H. Ahlberg, E.N.Nilson, J.L.Walsh[24], I.J. Schoenberg [25])は定評がある。Spline 関数の表現方法は，区分的多項式で表現する方法・Cardinal Spline による表現法と B-Spline による表現法の 3 っの表現方法がある。

下図に Cardinal Spline の variation を示す。この図から解るように Cardinal Spline 表現では負の部分が出てくるので，確率密度関数の推定においては使用を避けたい。そこで，負の部分が出てこない B-Spline の一次結合の形での表現を用いた Variation Diminishing Spline 関数表現方法を使う。

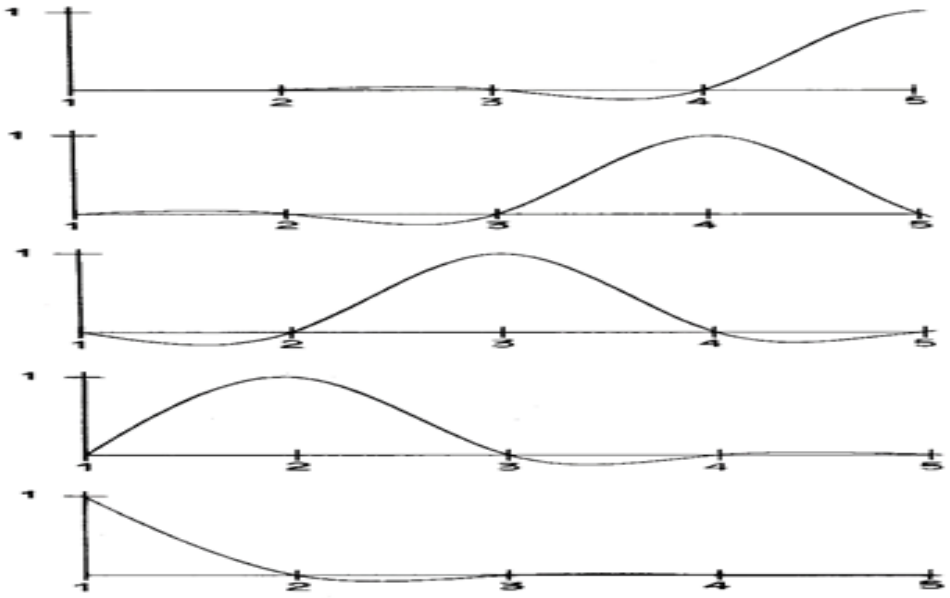


図 5.1 Cardinal Spline

ここでは,「形を維持する性質」を持つ Variation Diminishing Spline 関数（以下 V.D. Spline 関数と略す，I.J. Schoenberg [26,27]）によって，大標本から効果的に母集団の確率密度関数を導き出す。そして，それらは knot と node によって簡単に計算される。

V.D.Spline 関数は，適用された逆行列の理論を使用することによって，一次方程式の解を持つ折れ線関数に近似する。

5.2　折れ線関数による確率密度関数の近似

V.D. Spline 関数による効果的な母集団の確率密度関数を導き出すために，まず初めに折れ線関数による確率密度関数の近似推定を提案する。大標本が母集団から得られた場合，確率密度関数が１変数連続関数であると知られていると考える。それから近似関数に対する望ましい特性が簡単であり，変動に対して感度良く反応する。その折れ線関数(1 次の splne 関数)は，それらの要求に対して適当である。

確率密度関数 $\hat{f}(x)$ を得るために，大標本が母集団から得られるとして推定する。全ての標本を含む閉区間 $[a,b]$ は，n 個の等間隔で点 $\{t_i\}_{i=0}^{n}$ が $n+1$ 個得られる。

今，確率密度関数の近似関数は，次のように定義される。

$$\hat{f}_n(x)=\sum_{i=0}^{n} C_i N_{i,2}(x) \quad x\in[a,b] \tag{5.1}$$

ただし,

$$N_{i,2}(x)=\begin{cases}(x-t_{i-1})_+/(t_i-t_{i-1}) & if \quad x\le t_i\\ (t_{i-1}-x)_+/(t_{i+1}-t_i) & if \quad t_i\le x\end{cases} \tag{5.2}$$

$$t_i=a+ih,\quad i=-1,0,1,\cdots,n+1,\quad h=(b-a)/n$$
$$x_+=\max\{0,x\},\quad -\infty\le x\le\infty$$

$\hat{f}_n(x)$ では, n を大きくすることで, また $\{C_i=f_n(t_i)\}_{i=0}^{n}$ を得ることで, $\hat{f}_n(x)$ が $f(x)$ に近づいてくる。

n が十分大きくなり，(5.1)で $\{C_i\}_{i=0}^{n}$ が決まり，確率密度関数の特性を満たす。決定方程式は，次の特性で表される。

決定方程式を次のように行う。

まず, $\int_{-\infty}^{\infty} f(x)dx=1$ より,

$$\int_{-\infty}^{\infty}\hat{f}_n(x)dx=\int_{-\infty}^{\infty}\sum_{i=0}^{n} C_i N_{i,2}(x)dx=\sum_{i=0}^{n} C_i\int_{-\infty}^{\infty} N_{i,2}(x)dx=1 \tag{5.3}$$

ここで，$\int_{-\infty}^{\infty} N_{i,2}(x)dx$ は，底辺を $[t_{i-1},t_{i+1}]$，高さ 1 の三角形の面積である。

ところで，ここでの n は大きくなるので，区間 $\left[t_i,t_{i+1}\right]$ $(i=0,1,\cdots,n-1)$ を数個（ここではm個，mはたとえば４のような偶数）ずつひとまとめにして一つの区間とし，この部分区間における与えられた標本の相対度数を β_i/P としたとき，各部分区間上での $f_n(x)$ の積分が β_i/P であるとして，次の方程式を作る。

$$\int_{t_{im}}^{t_{(i+1)m}}\hat{f}_n(x)dx=\sum_{j=im}^{(i+1)m} C_j\int_{t_{im}}^{t_{(i+1)m}} N_{j,2}(x)dx=\beta_i/P \quad i=0,1,\cdots,\alpha-1 \tag{5.4}$$

更に，上のようにして定められた部分区間に関して隣り合う２つの部分区間の結合点付近の C_i について，その変動を考慮するため式(5.4)の積分区間を $m/2$ ずらして

$$\int_{t_0}^{t_{m/2}}\hat{f}_n(x)dx=\sum_{j=0}^{m/2} C_j\int_{t_0}^{t_{m/2}} N_{j,2}(x)dx=\gamma_0/P \tag{5.5}$$

$$\int_{t_{m/2+im}}^{t_{m/2+(i+1)m}} f_n(x)dx=\sum_{j=m/2+im}^{m/2+(i+1)m} C_j\int_{t_{m/2+im}}^{t_{m/2+(i+1)m}} N_{j,2}(x)dx=\gamma_{i+1}/P \qquad i=0,1,\cdots,\alpha-2$$

(5.6)

$$\int_{t_{m/2+(\alpha-1)m}}^{t_n} f_n(x)dx = \sum_{j=m/2+(\alpha-1)m}^{n} C_j \int_{t_{m/2+(\beta-1)m}}^{t_n} N_{j,2}(x)dx = \gamma_\alpha / P \qquad p:\text{任意}$$

(5.7)

を決定方程式に加える。

式(5.5)(5.7)において，γ/P，γ_α/P はそれぞれ範囲$\left[t_0, t_{m/2}\right]$，$\left[t_{m/2+(\alpha-1)m}, t_n\right]$での相対度数であり，また，$\{\gamma_{i+1}/P\}$ は，範囲$\left[t_{m/2+im}, t_{m/2+(i+1)m}\right]$ $(i=0,1,\cdots,\alpha-2)$での相対度数である。式(5.3)から(5.7)では，もし$[n/m]=n/m$ならば，$\alpha=n/m$であり，そうでなければ$\alpha=[n/m]+1$ である。そして，$(2\alpha+1)\times(n+1)$の係数行列を持つ連立方程式が得られる。

それらの連立方程式で，変数の数は方程式の数とは等しくない。$\{C_i\}_{i=0}^{n}$を決定するため一般化逆行列の方法を使って，方程式を解く。

式(5.3)から(5.7)の決定方程式を準一般化逆行列[28]を用いたときの解については非負であるという保証はないけれど， 準一般化逆行列を用いたときの連立方程式の解は２乗ノルムが最小になるという性質と先の決定方程式より，あるC_iが負ならばそれは０と見なしても良いという条件になっている。しかも，この折れ線関数は，Histogram を作りそれを補間，または近似したものと比べると，$\int_{-\infty}^{\infty} f_n(x)dx=1$という条件が存在し，また真の確率密度関数にいくらでも高精度に近似できるという好都合な性質を持っている。

(5.1)によって定義された多角形の確率密度関数は，$x=t_i$というブレイクポイントを持ち，また多くの$\{t_i\}_{i=0}^{n}$，$\{C_i\}_{i=0}^{n}$を持つ。

しかし，この不利な特性は次の節で考える。

5.3 Variation Diminishing Spline 関数による確率密度関数の近似

多くの$\{t_i\}_{i=0}^{n}$，$\{C_i\}_{i=0}^{n}$を持つということは折れ線表示に誤差部分を含む可能性が大きくなる，そこで滑らかな表現ができる V.D. Spline 関数を用いて表現することにより Fisher のいうデータの縮約も行うことを考える。

確率密度関数の推定を Spline 関数によって行う方法については L. I. Boveva, D. Kendal & I. Stefanov [29], I. J. Schoenberg[30]等の研究がある。

折れ線関数表現(5.1)は，多くの$\{C_i\}_{i=0}^{n}$ ， $\{t_i\}_{i=0}^{n}$と曲線を持つので，滑らかにする為に，$\{t_i\}_{i=0}^{n},\{C_i\}_{i=0}^{n}$の数を縮小する必要がある。

このような要求を満たすものとして，V. D. Spline 関数が存在する。

整数$m\geq 2$に対して，２変数を持つ関数を定義する。

$$M_m(t-x)=m\left(t-x\right)_+^{m-1} \tag{5.8}$$

一方，実軸上に数列$\{u_j\}_{j=0}^{k}$: $-\infty\leq u_0\leq u_1\leq\cdots\leq u_k\leq\infty$，$(u_j\leq u_{j+m})$が与えられている。$(k\geq m)$

関数$M_m(x;t)$（xは固定）の変数t（xは固定とする）についての $u_j,u_{j+1},\cdots,u_{j+m}$ 上のm階の差分商

$$M_{j,m}(x)=M_{j,m}\left(x;u_j,u_{j+1},\cdots,u_{j+m}\right)\qquad (j=0,1,\cdots,k-m) \tag{5.9}$$

をm 階の B-Spline という[41][42]。

B-Spline $M_{j,m}(x)$は，$u_j\leq x\leq u_{j+m}$のxに対して正で，それ以外は，０である。

差分商の Peano の定理から，$f(x)\in C$ のとき $f(x)$ の差分商は次のように表現される。

$$f(u_0,u_1,\cdots,u_m)=\frac{1}{m!}\int_{u_0}^{u_m}M_m(x;u_0,u_1,\cdots,u_m)\,f^{(m)}(x)\,dx \tag{5.10}$$

(5.10)で，$f(x)=x^m$とすると，(5.4)となる。

$$\int_{-\infty}^{\infty}M_m(x;u_0,u_1,\cdots,u_m)\,dx=1 \tag{5.11}$$

ここで，標準化 B-Spline $N_{j,m}(x)$を次の式によって定義する。

$$N_{j,m}(x)=\frac{u_{j+m}-u_j}{m}M_{j,m}(x) \tag{5.12}$$

今，knots$\{u_j\}_{j=0}^{k}$が与えられたものとし，p,qを$u_{p+m-1}\leq u_{q+1}$ であるような整数とする。

任意の全ての複素数 z と $x\in[u_{p+m-1},u_{q+1}]$ に対して，次の関係式が成立する。

$$m\left(z-x\right)^{m-1}=\sum_{j=p}^{q}\left(u_{j+m}-u_j\right)(z-u_{j+1})\cdots(z-u_{j+m-1})M_{j,m}(x) \tag{5.13}$$

zに対するこの関係式(5.6)の展開から，次の式が得られる。

$$\sum_{j=0}^{l}N_{j,m}(x)=1\quad(l=k-m) \tag{5.14}$$

更に，この関係式の展開の両辺を z について展開し，z の関数を比較する。

$$\sum_{j=0}^{l} \xi_{j,m} N_{j,m}(x) = x \tag{5.15}$$

$$\sum_{j=0}^{l} \xi_{j,m}^{(2)} N_{j,m}(x) = x^2 \tag{5.16}$$

そして，次が得られる。

$$\xi_{j,m} = \frac{1}{m-1}\left(u_{j+1} + \cdots + u_{j+m}\right) \tag{5.17}$$

$$\xi_{j,m}^{(2)} = \frac{1}{\binom{m-1}{2}} \sum_{i+1 \le s \le r \le j+m-1} u_r u_s \tag{5.18}$$

(5.17)から，

$$u_0 = \xi_{0,m} \le \xi_{1,m} \le \cdots \le \xi_{l,m} = u_k \tag{5.19}$$

が得られる。

次の図に３次の標準化 B-Spline の knots の多重度による変化を示す。縦に並んだ△印の個数はその場所における knots の多重度を示す。

(1) N(x:0,0,0,0,1) (6) N(x:2,3,4,4,5) (11) N(x:5,5,6,6,7)
(2) N(x:0,0,0,1,2) (7) N(x:3,4,4,5,5) (12) N(x:5,6,6,7,8)
(3) N(x:0,0,1,2,3) (8) N(x:4,4,5,5,5) (13) N(x:6,6,7,8,8)
(4) N(x:0,1,2,3,4) (9) N(x:4,5,5,5,6) (14) N(x:6,7,8,8,8)
(5) N(x:1,2,3,4,4) (10) N(x:5,5,5,6,6) (15) N(x:7,8,8,8,8)

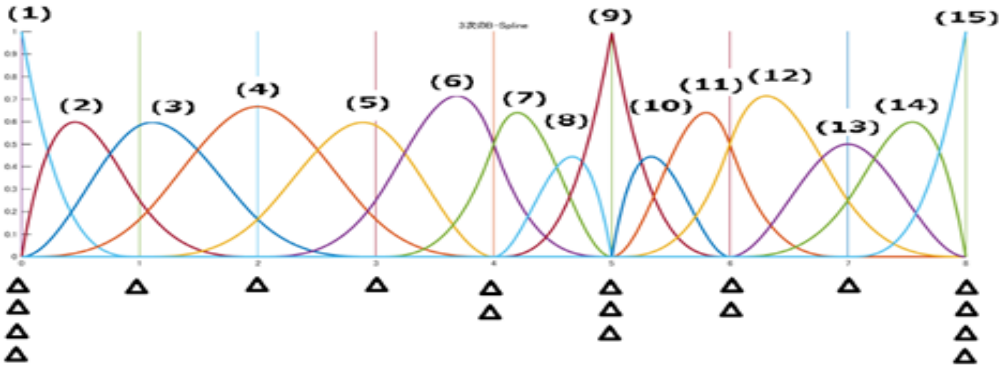


図 5.2 B-Spline の Variation

今，任意の $f \in C[a,b]$ に対して，次の関係式で定義される V.D.Spline 関数の近似を考える。

$$S(x) = S(x; f) = \sum_{j=0}^{l} f\left(\xi_{j,m}\right) N_{j,m}(x) \qquad (a \le x \le b) \tag{5.20}$$

$\left\{\xi_{j,m}\right\}_{j=0}^{l}$ は，近似方法の nodes と呼ばれる。nodes $\xi_{j,m}$ と $N_{j,m}(x)$ は，m 階と knots $\left\{u_j\right\}_{j=0}^{k}$ によって求められる。

このとき，V.D.Spline 関数は，次の特性を持つ。

$$S(x;f)\in C[a,b] \tag{5.21}$$

$$S(x;a_1f_1+a_2f_2)=a_1S(x;f_1)+a_2S(x;f_2) \tag{5.22}$$

$$S(x;A+Bx)=A+Bx \tag{5.23}$$

$$\begin{cases} V(S(x;f))\le V(f) \\ V(S(x;f)-a_1-a_2x)\le V(f-a_1-a_2x) \end{cases} \tag{5.24}$$

(a_1,a_2,A,B は，任意の実数 f_1，$f_2\in C[a,b]$。ここで $V(f)$ ，$V(S(x,f))$ は，任意の $f\in[a,b]$ の開区間 (a,b) における符号変化の数を表す。)

(5.24)から，V.D.Spline 関数が形を保存するという特徴があることが分かる。

これは，$S(x)$ と任意の直線との交差回数が，曲線 $f(x)$ と同じ直線との　交差回数よりも多くはないということである。

次の定理 5.1 を用いることにより，折れ線関数 $f(x)$ の knots $\{t_i\}_{i=0}^{n}$ の中から V.D.Spline 関数の knots $\{x_i\}_{i=0}^{e}$ を選ぶことができる。また，V.D.Spline 関数 $S(x)$ による近似は単調性や凸性を保存するばかりではなく，全変動をも減少させるという性質を持つ。この，V.D. Spline 関数 $S(x)$ を使用することによって，有効に確率密度関数を表現することができる。

また，V.D.Spline 関数近似による誤差限界は M.Marsden ，I.J.Schoenberg [31]，M. Marsden [32] より次のように与えられている。

誤差評価(等間隔 knots の場合) [31]

$f\in[a,b]$,　　$\omega(f;\delta)$ は区間幅 δ に関する連続率

$S(x)$ は

$$\overbrace{a,\cdots,a}^{m\text{ヶ}},a+\frac{b-a}{n},a+\frac{2(b-a)}{n},\cdots,a+\frac{(n-1)(b-a)}{n},\overbrace{b,\cdots,b}^{m\text{ヶ}}$$

を knots に持つ V.D.Spline 関数(order m)とする。

このとき，

i)　　$2<m\le n+2$ ならば

$$\left|f(x)-S(x)\right|\le\left(1+\sqrt{m}\right)\omega\left(f;\frac{b-a}{n}\right)\qquad x\in[a,b] \tag{5.25}$$

ⅱ) $m>n+2$ ならば

$$\left|f(x)-S(x)\right|\le\left\{1+\sqrt{\frac{1}{m-2}\left(\frac{m-1}{4}\right)-\frac{1}{6}\left(n-\frac{1}{n}\right)}\omega\left(f;\frac{b-a}{\sqrt{m-1}}\right)\right\}\qquad x\in[a,b] \tag{5.26}$$

誤差評価(不等間隔 knots の場合) [32]

$a=x_{-m+1}=x_{-m+2}=\cdots=x_0<x_1<\cdots<x_{n-1}<x_n=x_{n+1}=\cdots=x_{n+m-1}=b$ を

knots にもつ V.D.Spline 関数とし

$\alpha^2=\max_{-m+1\le i\le n-1}\left(x_{i+m-1}-x_i\right)^2\Big/(m-1)$ とすると

$$\left|f(x)-S(x)\right|\le 2\omega(f;\alpha)\qquad x\in[a,b] \tag{5.27}$$

この, V.D.Spline 関数 $S(x)$ を使用することによって, 次の定理を用いることで上記の誤差限界を保ちつつ $S(x)$ の knots を決定する方法を提案できる。 この論文で提案する, knots の決定法により, R.A.Fisher[1]のいう統計学の問題の分類(Ⅲ データの簡約方法に関する研究)への貢献が期待される。

定理 5.1(塚越[33],[34])

$f_n(x)$ は, 式(5.1)によって与えられ, ν は自然数である。そのとき, δ は次のように与えられる。

$$\delta=\max\left\{t_{i-1}-t_{i-1}\,;\ i=0,1,\cdots,n-1\right\}$$

$\omega(f_n;\delta)$ は, 区間幅 δ に関する $f_n(x)$ の連続率である。そのとき, 次のように与えられる。

$$\begin{cases}x_0=t_0\\ x_j=\max\left\{t_k:x_{i-1}\le t\le x_i,\left|f_n(t)-f_n\left(x_{i-1}\right)\right|\le\nu\varpi\left(f_n;\delta\right)\right\}\end{cases}\qquad j=1,2,\cdots,e \tag{5.28}$$

(ただし, e は式(5.26)が意味をもつ最大の添字)

$S(x;f_n)$ は, 式(5.26)によって定義された V.D.Spline 関数とする。このとき誤差限界は次のように表される。

$$\left|f_n(x)-S(x;f_n)\right|\le(m-1)\nu\omega(f_n;\delta)\quad for\quad all\quad x\in[a,b] \tag{5.29}$$

定理 5.1 の式(5.26)の方法で$\{x_i\}_{i=0}^{e}$ を決め，これをそのまま式(5.27)の$S(x;f_n)$の knots $\{u_i\}_{j=0}^{k}$にしてもよい。しかし，注意すべき点が２つある。

１つは，連続関数$f_n(x)$は，ある特定の区間でのみ急激に変動して，そこで $\omega(f_n;\delta)$を与え，他の大部分の区間では緩やかに変動するという現象がしばしばみられる。このような関数$f_n(x)$に対しては，式(5.26)の knots$\{x_i\}_{i=0}^{e}$を決める条件は弱すぎるため次の工夫をする。

連続率$\omega(f_n;\delta)$を N 等分する。

$|f_n(t_{i+1})-f_n(t_i)|, i=0,1,2,\cdots,n-1$のうち，

$(n-1)\varepsilon$個$(0\le\varepsilon\le1)$が$\gamma\omega(f_n;\delta)$より小さくなる最小の$\gamma(1\le\gamma\le N)$ を決定する。

この γを用いて， $x_0=t_0,\cdots,t_\lambda=\min\{t_h: x_{j-1}=t_i\le t_h\le t_n, |f_n(t_h)-f_n(t_i)|\ge\gamma\omega(f_n;\delta)/N\}$ とし，

$$x_j=\begin{cases} t_\lambda & when \\ t_{\lambda-1} & otherwise \end{cases} \qquad |f_n(t_\lambda)-f_n(t_i)|\le\omega(f_n;\delta) \qquad j=1,2,\cdots,e \tag{5.30}$$

このとき，定理 5.1 の式(5.27)の誤差限界を満足するような knots$\{x_j\}_{j=0}^{e}$が選択される。(塚越[33])

２つ目は，閉区間$[a,b]$の両端a,b の近傍では情報が不足である。したがって，両端での情報不足を補い，かつ両端での近似精度を高めるために次のような工夫を行う。今，knots $\{x_j\}_{j=0}^{e}$の両端においてm個を多重させた次の記法を用いる。

$$u_j=\begin{cases} x_0 & if \quad j=0,1,...,m-1 \\ x_{j-m+1} & if \quad j=m,m+1,...,e+m+2 \\ x_e & if \quad j=e+m-1,e+m,...,e+2(m-1) \end{cases} \tag{5.31}$$

$\{u_j\}_{j=0}^{e+2(m-1)}$より，$\xi_{j,m}$，　$f_n(\xi_{j,m})$，を計算することにより V.D.Spline 関数表現された確率密度関数$S(x)$が求められる。$(k=e+2(m-1),\ l=k-m=e+m-2)$ (塚越[33])

5.4 各特性値の計算

母集団の確率密度関数が，V.D.Spline 関数 $S(x)=S(x;f_n)$ で近似的に表現されたとき，この母集団の平均と分散は次のように与えられる。

$\mu=\int_{-\infty}^{\infty}xS(x)dx$ ， $\sigma^2=\int_{-\infty}^{\infty}(x-\mu)^2S(x)dx$

式(5.3)より，B-Spline の平均と分散は次のように与えられる。

$$\mu_{j,m}=\int_{-\infty}^{\infty}xM_{j,m}(x)dx=\frac{1}{m+1}\sum_{\nu=0}^{m}u_{j+\nu} \tag{5.32}$$

$$\sigma_{j,m}^2=\int_{-\infty}^{\infty}\left(x-\mu_{j,m}\right)^2M_{j,m}(x)dx=\frac{1}{(m+1)^2(m+2)}\sum_{\gamma\le s}\left(u_{j+\gamma}-u_{j+s}\right)^2 \tag{5.33}$$

式(5.32) (5.33)より，平均μ と分散σ^2は，knots によってだけ計算される。

定理 5.2(塚越[35],[36])

母集団確率密度関数が，式(5.20)の V.D.Spline 関数によって表現されるとき，母平均μと母分散σ^2は，knots $\{u_j\}_{j=0}^{n}$ と nodes から次の式により計算される。

$$\mu=\int_{-\infty}^{\infty}xS(x)dx=\sum_{j=0}^{l}f_n\left(\xi_{j,m}\right)\frac{u_{j+m}-u_j}{m}\mu_{j,m} \tag{5.34}$$

$$\sigma^2=\int_{-\infty}^{\infty}(x-\mu)^2S(x)dx=\sum_{j=0}^{l}f_n\left(\xi_{j,m}\right)\frac{u_{j+m}-u_j}{m}\left[\sigma_{j,m}^2+\left(\mu_{j,m}-\mu\right)^2\right] \tag{5.35}$$

［証明］

$$\mu=\int_{-\infty}^{\infty}xS(x)dx=\int_{-\infty}^{\infty}\sum_{j=0}^{l}fn\left(\xi_{j,m}\right)N_{j,m}(x)dx$$

$$=\sum_{j=0}^{l}fn\left(\xi_{j,m}\right)\frac{u_{j+m}-u_j}{m}\int_{-\infty}^{\infty}xM_{j,m}(x)dx=\sum_{j=0}^{l}fn\left(\xi_{j,m}\right)\frac{u_{j+m}-u_j}{m}\left[\sigma_{j,m}^2+\left(\mu_{j,m}-\mu\right)\right]$$

$$\sigma^2=\int_{-\infty}^{\infty}\left(x-\mu\right)^2S(x)dx=\int_{-\infty}^{\infty}\left(x-\mu\right)^2\sum_{j=0}^{l}f\left(\xi_{j,m}\right)N_{j,m}(x)dx=\sum_{j=0}^{l}\frac{u_{j+m}-u_j}{m}f_n\left(\xi_{j,m}\right)\int_{-\infty}^{\infty}\left(x-\mu\right)^2M_{j,m}(x)dx$$

$$= \sum_{j=0}^{l} \frac{u_{j+m} - u_j}{m} f_n\left(\xi_{j,m}\right)\left[\int_{-\infty}^{\infty} x M_{j,m}(x)dx - 2\mu\mu_{j+m} + \mu^2\right] = \sum_{j=0}^{l} \frac{u_{j+m} - u_j}{m} f_n\left(\xi_{j,m}\right)\left[\sigma_{j,m}^2 + \mu_{j,m}^2 - 2\mu\mu_{j+m} + \mu^2\right]$$

$$= \sum_{j=0}^{l} \frac{u_{j+m} - u_j}{m} f_n\left(\xi_{j,m}\right)\left[\sigma_{j,m}^2 + \left(\mu_{j,m} - \mu\right)\right]$$

Q.E.D

5.5 V.D.Spline 関数によって近似された確率密度関数の特性関数

以前の節では，確率密度関数の特性は，差分商の Peano の定理から knots と nodes によって表現される。

この章では，確率密度関数 $S(x)$ から特性関数を得る。

V.D.Spline 関数は，特性関数を得るのにとても効果的である。なぜなら，V.D.Spline 関数は B-Spline の線結合により表現され，この B-Spline の特性は正となり，その数直線上の積分は１に等しい。そのとき，式(5.9)の B-Spline $M(x)$は確率密度関数と見なすことができる。

B-Spline $F(x)$

$$F_m(x) = F_m\left(x; u_0, u_1, \cdots, u_m\right) = \int_{-\infty}^{\infty} M_m\left(t; u_0, \cdots, u_m\right)dt \tag{5.36}$$

の分布関数は，次の特性を持つ。

$$\lim_{m\to\infty} F_m(x) = F(x) \tag{5.37}$$

このとき，$F(x)$はポリヤ分布関数である。(H.B.Curry, I.J. Schoenberg[37])

一方，式(5.33)より次が得られる。

$$\lim_{m\to\infty} \int_{-\infty}^{\infty} \left(1 + \frac{itx}{m}\right)^{-m-1} dx = \int_{-\infty}^{\infty} e^{-itx} dF(x) \tag{5.38}$$

B-Spline の特性関数は，次の関係によって表現される。

$$\int_{-\infty}^{\infty} \left(1 + \frac{itx}{m+1}\right)^{-m-1} M_m(x)dx = \prod_{\nu=0}^{m} \frac{1}{\left(1 + \frac{itu_\nu}{m+1}\right)} \tag{5.39}$$

一般的に，特性関数$\Phi(x)$はこのモーメント

$\mu_h \quad (h=1,2,\cdots),\ (\mu_0=1)$　　によって次のように表現される。

$$\varphi(x)=\int_{-\infty}^{\infty}e^{-it}dF(x)=\sum_{h=0}^{\infty}(-1)\frac{\mu_h}{h!}t^h \tag{5.40}$$

今，B-Spline の特性関数からこの関係

$$\phi(x)=\int_{-\infty}^{\infty}\left(1+\frac{itx}{m+1}\right)^{-m-1}M_m(x)dx=\prod_{\nu=0}^{m}\frac{1}{\left(1+\frac{itu_\nu}{m+1}\right)}$$

を導く。

$$\phi'(0)=-\frac{\sum_{\nu=0}^{m}u_\nu}{-m+1} \tag{5.41}$$

$$\phi''(0)=\frac{2\sum_{\gamma\le s}^{m}u_\gamma u_s}{(m+1)^2} \tag{5.42}$$

このとき，確率密度関数の $M_m(x)$ について１，２次のモーメントは式(5.9)から導かれる。

$$\mu_1^{(m)}=\int_{-\infty}^{\infty}xM_m(x)dx=\frac{\sum_{\nu=0}^{m}u_\nu}{m+1}=(-1)\phi'(0) \tag{5.43}$$

$$\mu_2^{(m)}=\int_{-\infty}^{\infty}x^2M_m(x)dx=\frac{2\sum_{\gamma\le s}^{m}u_\gamma u_s}{(m+1)(m+2)}=(-1)^2\frac{(m+1)^2}{(m+1)(m+2)}\phi''(0) \tag{5.44}$$

その結果，B-Spline の要素で $\mu_h^{(m)}$ が得られ，次のようになる。

$$\int_{-\infty}^{\infty}\left(1+\frac{itx}{m+1}\right)^{-m-1}M_m(x)dx=\sum_{h=0}^{\infty}(-1)^h\frac{\prod_{\nu=0}^{h}(m+\nu)}{(m+1)}\frac{\mu_h^{(m)}}{m!}(it)^h \tag{5.45}$$

そして，式(5.20)より V.D.Spline 関数の確率密度関数 $s(x)$ の特性関数は，次の関数によって表される。

$$\int_{-\infty}^{\infty}e^{-itx}S(x)dx=\int_{-\infty}^{\infty}e^{-itx}\sum_{j=0}^{l}f(\xi_{j,m})N_{j,m}(x)dx=\sum_{j=0}^{l}f(\xi_{j,m})\frac{u_{j+m}-u_j}{m}\int_{-\infty}^{\infty}e^{-itx}M_{j,m}(x)dx \tag{5.46}$$

式(5.35)(5.36)より，次が得られる。

$$\int_{-\infty}^{\infty} e^{-itx} S(x)dx \cong \sum_{j=0}^{l} f\left(\xi_{j,m}\right)\frac{u_{j+m}-u_j}{m}\int_{-\infty}^{\infty}\left(1+\frac{itx}{m+1}\right)M_{j,m}(x)dx$$
$$= \sum_{j=0}^{l} f\left(\xi_{j,m}\right)\frac{u_{j+m}-u_j}{m}\prod_{\nu=0}^{m}\frac{1}{\left(1+\frac{itu_{j+\nu}}{m+1}\right)} \quad (5.47)$$

そして,

$$\Phi(t)=\int_{-\infty}^{\infty} e^{-itx} S(x)dx = \sum_{j=0}^{l} f\left(\xi_{j,m}\right)\frac{u_{j+m}-u_j}{m}\prod_{\nu=0}^{m}\frac{1}{\left(1+\frac{itu_{j+\nu}}{m+1}\right)}$$

と与えられたとき，次が得られる。

$$\Phi'(0)=-\sum_{j=0}^{l} f\left(\xi_{j,m}\right)\frac{u_{j+m}-u_j}{m}\frac{\sum_{\nu=0}^{m}u_{j+\nu}}{m+1} \quad (5.48)$$

$$\Phi''(0)=\sum_{j=0}^{l} f\left(\xi_{j,m}\right)\frac{u_{j+m}-u_j}{m}\frac{2\sum_{\gamma\le s}u_{j+\gamma}u_{j+s}}{(m+1)} \quad (5.49)$$

そして，$S(x)$の原点周りの１，２次のモーメントは次のようになる。

$$\mu_1=\int_{-\infty}^{\infty} xS(x)dx = \sum_{j=0}^{l} f\left(\xi_{j,m}\right)\frac{u_{j+m}-u_j}{m}\frac{\sum_{\nu=0}^{m}u_{j+\nu}}{m+1}=(-1)\Phi'(0) \quad (5.50)$$

$$\mu_2=\int_{-\infty}^{\infty} x^2S(x)dx = \sum_{j=0}^{l} f\left(\xi_{j,m}\right)\frac{u_{j+m}-u_j}{m}\frac{2\sum_{\gamma\le s}u_{j+\gamma}u_{j+s}}{(m+1)(m+2)}=(-1)^2\frac{(m+1)^2}{(m+1)(m+2)}\Phi''(0) \quad (5.51)$$

その結果，$S(x)$のモーメントがμ_hとすると次のようになる。

$$\Phi(t)=\int_{-\infty}^{\infty}\left(1+\frac{itx}{m+1}\right)^{-m-1} S(x)dx = \sum_{j=0}^{l} f\left(\xi_{j,m}\right)\frac{u_{j+m}-u_j}{m}\sum_{-\infty}^{\infty}(-1)^h\frac{\prod_{\nu=1}^{h}(m+\nu)}{(m+1)^h}\frac{\mu_h}{m!}(it)^h \quad (5.52)$$

特性関数は少しの knots と nodes によって計算される。

従って，今母集団の V.D.Spline 関数の確率密度関数の特性が $S(x)$ の knots と nodes から計算されるのは明白である。

5.6 数値実験

本章では，V.D.Spline 関数によって近似折れ線関数 $f(x)$ を得た確率密度関数を求めるために次の 3 点を調べる。

１） $S(x)$ による母集団確率密度関数の再生の水準

２） $S(x)$ による母平均，母分散の近似

３） $S(x)$ が必要とする knots と nodes の数

例として，５００個の正規分布に従う乱数から V.D.Spline 関数により表現された確率密度関数を求める。

そして，母集団確率密度関数と V.D.Spline 関数による推定 $S(x)$ を図で比較する。

更に，数少ないデータの Histogram と V.D.Spline 関数 $S(x)$ との比較を表す。

実験 5.1

標本は，正規分布 $N(45,16)$ に従う乱数を用いる。図 5.3 は母集団確率密度関数と $S(x)$ を表し，図 5.4 は Histogram と $S(x)$ を表す。

そのとき，式(5.32)(5.33)より求めた平均と分散はそれぞれ $\mu = 45.15$, $\sigma^2 = 18.74$ であり，knots の数は 18 個である。

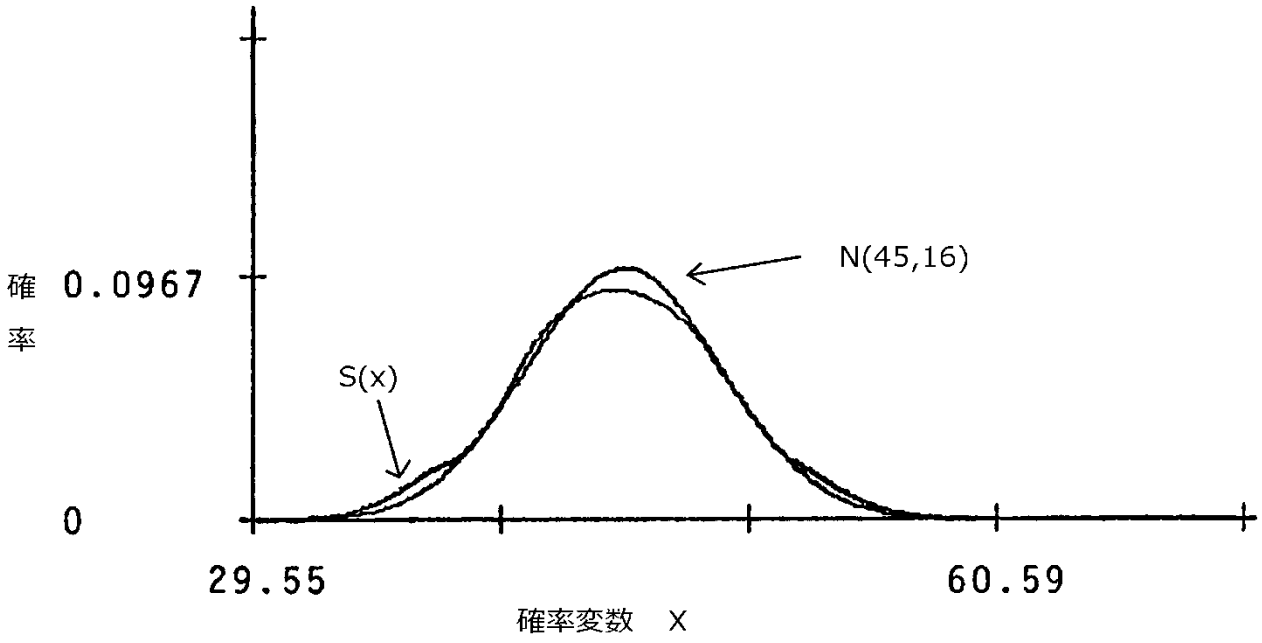


図 5.3 母集団確率密度関数と $S(x)$

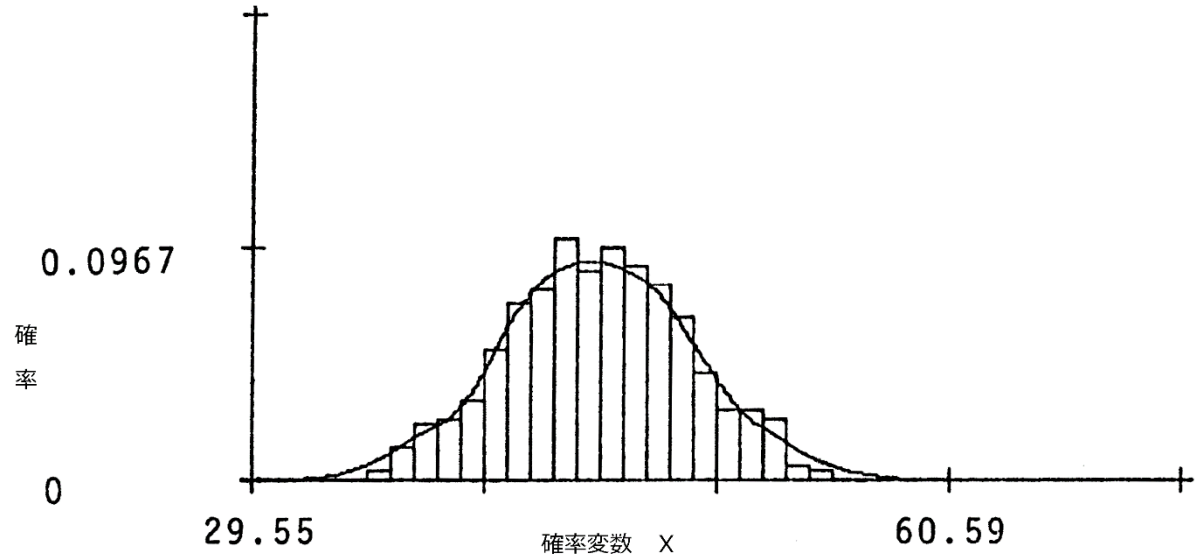


図 5.4 Histogram と $S(x)$

図 5.3, 図 5.4 より $\frac{1}{\sqrt{2\pi}4}e^{-\frac{(x-45)^2}{2\times16}}$ にも Histogram にもどちらにも適応している。

実験 5.2

２つの標本は，等しい平均と異なる分散を持つ正規分布にしたがって得られる。

$\left(N(35,12.25),\ N(35,25)\right)$

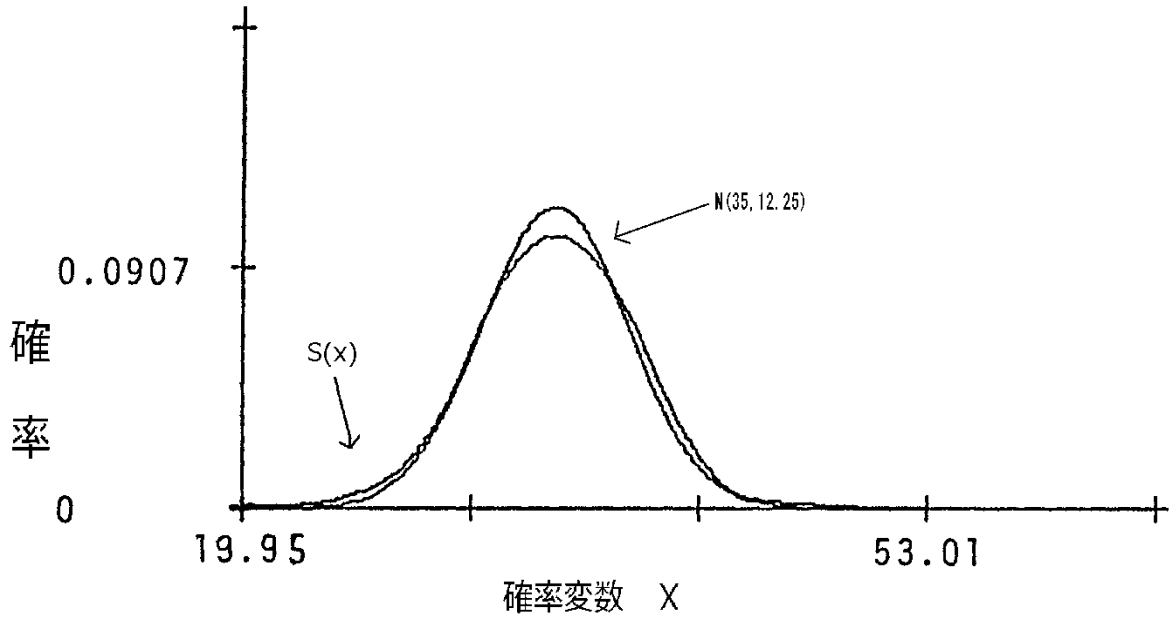


図 5.5 $N(35,12.25)$

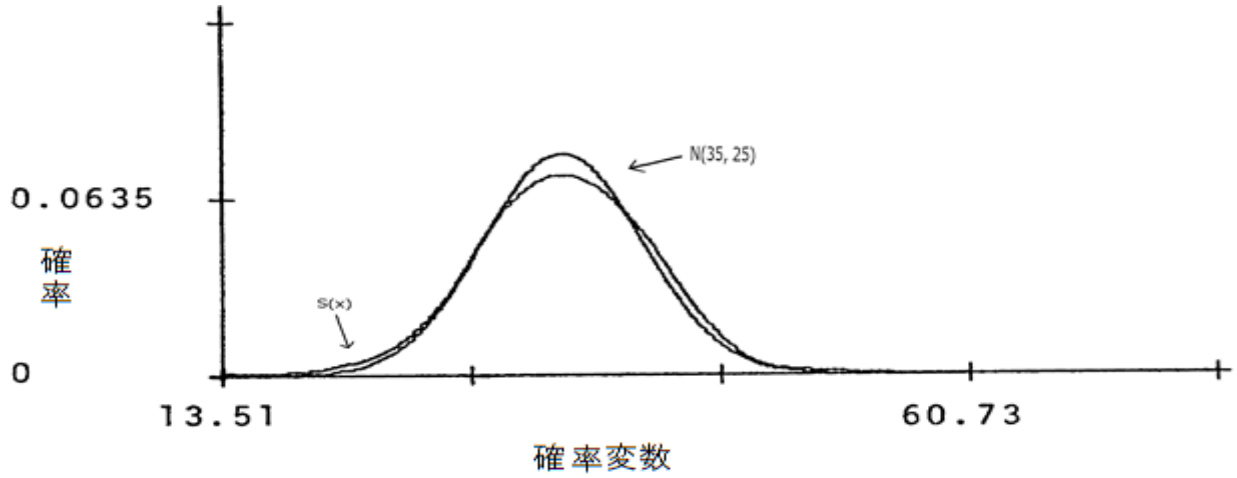


図 5.6 $N(35,25)$

図 5.5, 図 5.6 より $\frac{1}{\sqrt{2\pi}5}e^{-\frac{(x-35)^2}{2\times25}}$ にも $\frac{1}{\sqrt{2\pi\times12.25}}e^{-\frac{(x-35)^2}{2\times12.25}}$ にもどちらにも適応している。

図 5.5 では平均 35.58, 分散 15.91 図 5.6 では平均 35.61, 分散 32.16 どちらも knots 数は 18 となった。

実験 5.3

確率密度関数 $S(x)$ は, 図 5.7 の Histogram によって表現されたデータによって求められる。そして, 母平均と母分散はそれぞれ $\mu=49.98$, $\sigma^2=9.34$ である。また, $S(x)$ による平均と分散はそれぞれ $\hat{\mu}=50.47$, $\hat{\sigma}^2=10.75$ である。

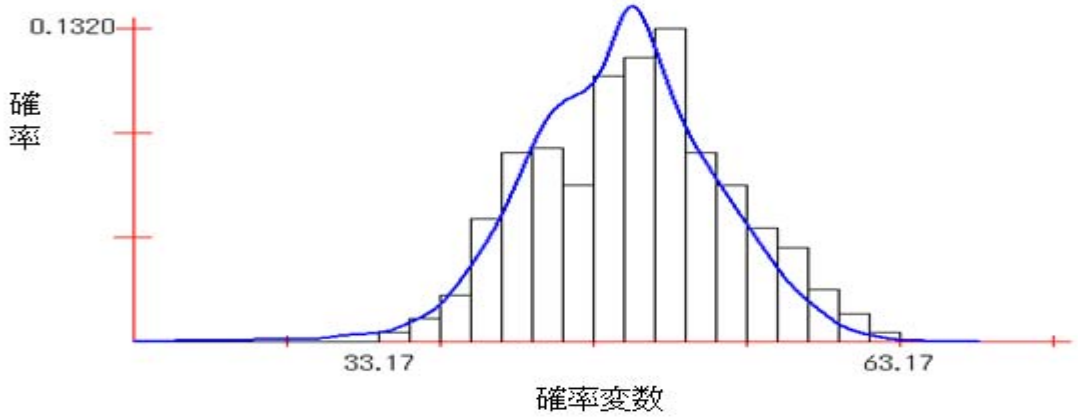


図 5.7

Histogram への適応も, 特性値の推定もうまく表現されている。

実験 5.4

3 章, Histogram のところで用いた, サンプル数 1041 の図 3.2 のデータについて V.D.Spline 関数 $s(x)$によって表現された確率密度関数を表示する。

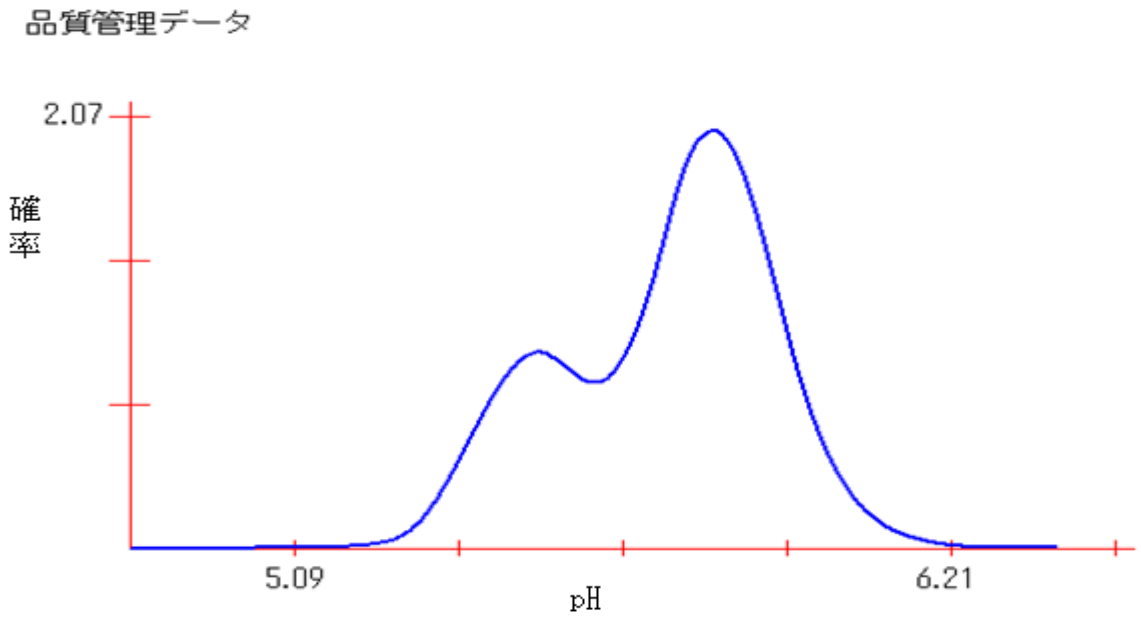


図 5.8　図 3.2 のデータへの V.D.Spline 関数 $S(x)$による推定

Sturges の規則, Scott の選択に近い形である。

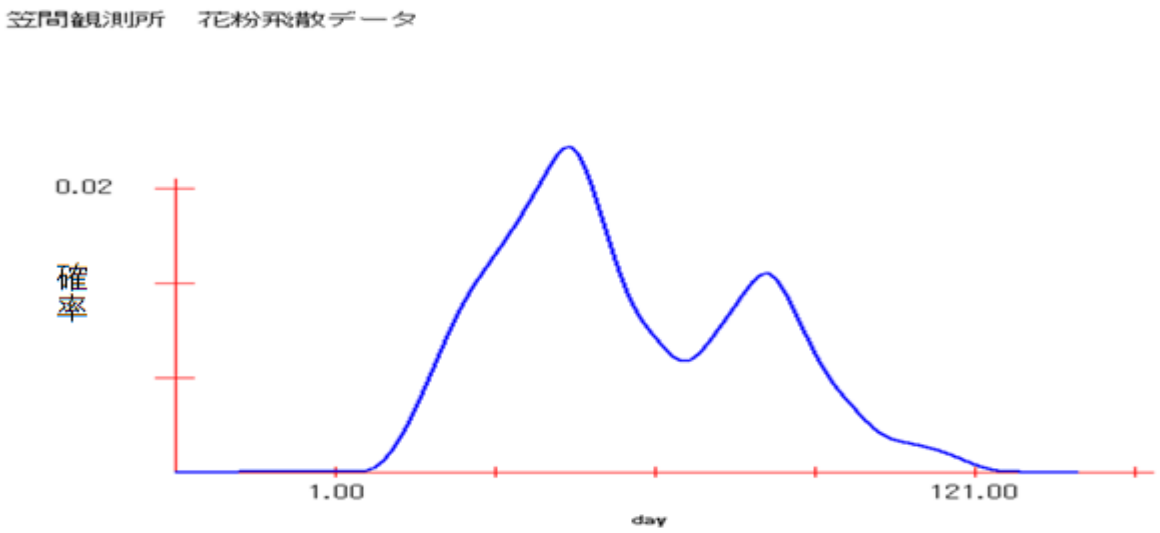


図 5.9 図 3.10, 図 4.6 の笠間観測所花粉飛散データの
V.D. Spline 関数による密度関数の推定

31 個の knots と nodes で表現できる。

確認

分布型が知られているとき，χ^2による適合の慣習的な良さが使われている。しかし，適合の良さのχ^2分析は，既に成し遂げられた分布に対して標本分布を分析することである。したがって，母集団分布型で情報なしでこのχ^2分析を適合させることはできない。

多くの評価方法は，確率密度関数から展開された。しかし，V.D.Spline 関数$s(x)$によって表現された確率密度関数を使用する有力な Nonparametric な分析方法を立証することが望ましい。

ここでは，knots と nodes を用いた B-Spline 関数による確率密度関数の推定を行った。これは，Kernel 関数を B-Spline 関数とした確率密度関数の推定を行っていることである。それに伴って B-Spline 関数に必要に knots と nodes をどう決めるべきかを述べた。(5.27)，(5.28)，(5.29)により knots を決めることは(変動が少なければ knots の間隔が長くなる。)，データの変動の大きさにより Band 幅が可変になり，より表現の自由度が増していることを示し，更にその多重度を決めるということは図 5.2 の V.D.Spline 関数の variation に見られる様にデータに対する適応が敏感になる。

それにより nodes を決め確率密度関数の推定することによりデータの変動の激しいところには敏感に，変動が激しくないところには穏やかに適応する確率密度関数の推定方法ができた。

しかも，核関数の数が通常の Kernel 関数を用いた確率密度関数の推定よりも少なく再計算の際の効率化が図れている。

通常の Kernel 関数を用いた確率密度関数の推定では確率密度関数を表現するときには全てのデータが必要だけれど(通常，数百の単位以上)，B-Spline 関数による表現は knots と nodes だけで表現できるから(多くて，二十前後～数十の単位)，保存データ数も少なくて済む。

確率密度関数推定への従来手法

1. Kernel 関数を用いた, Kernel 法

$$\hat{f}_h(x)=\frac{1}{n}\sum_{i=1}^{n}K_h(x-x_i)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right)$$

$K(\cdot)$はKernel　　$K_h(x) = 1/h\,K(x/h)$.

確率密度関数推定への本章で提案した手法

2. Variation Diminishing Spline 関数を用いた提案する方法

$$S(x)=S(x;f)=\sum_{j=0}^{l}f(\xi_{j,m})N_{j,m}(x) \qquad x\in[x_0,x_n]$$
$$f(x)\in C[x_0,x_n]$$

6.　提案する正規混合分布の解析方法１（非線形最適化手法を用いる方法）

この節では，混合分布の要素分布の推定法の１っとして非線形最適化手法を用いた，方法を提案する。[38]

従って，今回は

$$l^2=\sum_{i=1}^{k}\left\{P(x_i)-\sum_{j=1}^{3}\omega_j\frac{1}{\sqrt{2\pi}\sigma_j}e^{-\frac{(x_i-\mu_j)^2}{\sigma_j^2}}\right\}^2\rightarrow\min \tag{6.1}$$

$\left(P(x_i)\quad は x_i における母集団確率\right)$

とする問題としてとらえていく。

この非線形最小化問題の解法として以下のような Fletcher-Powel 法を用いる。

6.1 Fletcher-Powell 法

非線形最小化問題の解法としての Fletcher-Powell 法を簡単に説明しておく。

評価関数 $f(x)$ が与えられたとき，任意の点 $\mathbf{x}$ と最小点 $\mathbf{x}^*$ の差 $\mathbf{h}$ を求めてみよう。$\mathbf{x}$ は $\mathbf{x}^*$ の近似であるとし，$\mathbf{h}=\mathbf{x}-\mathbf{x}^*$ を $f(\mathbf{x})$ に代入する。そして，$f(\mathbf{x}^*)$ を $\mathbf{x}=\mathbf{x}^*$ の近傍でティラー展開し，$(\mathbf{x}^*-\mathbf{x})$ の 3 次以上の項を無視すると，次の 2 次形式を得る。

$$f(\mathbf{x}^*)=f(\mathbf{x}+\mathbf{h})\approx f(\mathbf{x})+f_x^T(\mathbf{x})\mathbf{h}+\frac{1}{2}\mathbf{h}^T f_{xx}(\mathbf{x})\mathbf{h} \tag{6.2}$$

ただし，

$$f_x^T(\mathbf{x})=\left(\frac{\partial f}{\partial x}\right)^T=\mathbf{g}\qquad f_{xx}(\mathbf{x})=\left[\frac{\partial^2 f}{\partial x_i\partial x_j}\right]=\mathbf{G}$$

であり，$\mathbf{G}$ は $n\times n$ 次の正則な対称行列（Hesse 行列）であるとする。

$f(\mathbf{x}^*)$ は最小値であるから，(6.1) 式より

$$\frac{\partial f}{\partial x} = \mathbf{o} = \mathbf{g} + \mathbf{G}\mathbf{h} \tag{6.3}$$

となる。

よって，$\mathbf{h} = -\mathbf{G}^{-1}\mathbf{g}$ を得る。

Fletcher-Powell 法では $\mathbf{G}^{-1}$ を直接計算しないで，勾配ベクトル$\mathbf{g}$を用いてそれを求める。

行列 $\boldsymbol{H}^{(k)}$ は$k-1$回の繰返しによって得られた正定値対称行列であるとする。

そして，$\boldsymbol{x}^{(k)}$における方向ベクトルは　$\mathbf{d}^{(k)} = -\mathbf{H}^{(k)} gradf\left(\mathbf{x}^{(k)}\right) = -\mathbf{H}^{(k)}\mathbf{g}^{(k)}$

で，かつ行列 $\mathbf{G}$ に対していままでの繰返しで得られた$k-1$個のベクトルが $\mathbf{d}^{(1)}$，$\mathbf{d}^{(2)}$，・・・，$\mathbf{d}^{(k-1)}$ に対して，$\mathbf{d}^{(k)T}\mathbf{G}\mathbf{d}^{(l)} = 0$　　$\left(l = 1,2,\cdots,k-1\right)$ を満たすように，すなわち，互いに共役になるように $\mathbf{d}^{(k)}$ を選ぶ。

この方向ベクトル $\mathbf{d}^{(k)}$ の方向に降下したとき直線

$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha\mathbf{d}^{(k)}$　上にある最小点へ移行するときのベクトル $\mathbf{d}^{(0)}$，$\mathbf{d}^{(1)}$，$\mathbf{d}^{(2)}$，・・・，$\mathbf{d}^{(k)}$ が行列 $\mathbf{H}^{(k+1)}\mathbf{G}$ の固有値１をもつ固有ベクトルになるように $\mathbf{H}^{(k)}$ を修正して$\mathbf{H}^{(k+1)}$を求める。

すなわち $\mathbf{H}^{(k)}$ から $\mathbf{H}^{(k+1)}$ を求めるとき，近似的に $\mathbf{G}\mathbf{H}^{(k+1)} = \mathbf{I}$ になるように $\mathbf{H}^{(k+1)}$ を求めれば，収束したときには $\mathbf{H}^{(k+1)}$ は$\mathbf{G}$ の逆行列になる。

$\mathbf{H}^{(k+1)}$の求め方：次の行列を考える。

$$\sum_{i=1}^{p} \alpha_i \mathbf{d}^{(i)}\mathbf{d}^{(i)T} \tag{6.4}$$

ただし，$\alpha_i = 1/\mathbf{d}^{(i)T}\mathbf{G}\mathbf{d}^{(i)}$　$\mathbf{d}^{(i)}\left(i = 1,2,\cdots,p\right)$ は p 個の共役な方向ベクトルである。

$\left(l = 1,2,\cdots,p\right)$　において

$$\left(\sum_{i=1}^{p} \alpha_i \mathbf{d}^{(i)}\mathbf{d}^{(i)T}\right)\mathbf{G}\mathbf{d}^{(i)} = \alpha_l \mathbf{d}^{(l)}\mathbf{d}^{(l)T}\mathbf{G}\mathbf{d}^{(l)} = \mathbf{d}^{(l)}$$

が成り立つ。

特に，（4.5）式の 2 次形式に対して，$p = n$の場合，上式から次式が得られる。

$$\mathbf{G}^{-1} = \sum_{i=1}^{n} \mathbf{A}^{(i)}$$

$$\mathbf{A}^{(i)} = \frac{\mathbf{d}^{(i)}\mathbf{d}^{(i)T}}{\mathbf{d}^{(i)T}\mathbf{G}\mathbf{d}^{(i)}} \tag{6.5}$$

この部分和は（6.4）で表される意味において，逆行列の近似値として利用することができる。

いま，k 回目の探索における逆行列の最良の近似値を$\mathbf{H}^{(k+1)}$とすると，次の（k＋1）回目の方向ベクトルとして　$\boldsymbol{d}^{(k+1)}=-\boldsymbol{H}^{(k+1)}\boldsymbol{g}^{(k+1)}$

を用い，その探索の結果を用いて，さらに逆行列の近似値を改良するという方法を考える。

k 回目の方向ベクトル$\mathbf{d}^{(k)}$，および（k+1）回目の近似の最小点$x^{(k)}$，G^{-1}の近似の行列$\mathbf{H}^{(k+1)}$は

$$\mathbf{d}^{(k)}=-\mathbf{H}^{(k)}\mathbf{g}^{(k)} \tag{6.6}$$

$$\mathbf{x}^{(k+1)}=\mathbf{x}^{(k)}+\alpha_k\mathbf{d}^{(k)} \tag{6.7}$$

$$\mathbf{H}^{(k+1)}=\mathbf{H}^{(k)}+\mathbf{A}^{(k)}+\mathbf{B}^{(k)} \tag{6.8}$$

となる。

ここに，$\mathbf{H}^{(k)}$を修正するとき，（6.5）の$\mathbf{A}^{(k)}$を 1 次近似として使い，その修正として行列$\mathbf{B}^{(k)}$を加える。

これまでの方向ベクトル$\mathbf{d}^{(1)}$，$\mathbf{d}^{(2)}$，・・・，$\mathbf{d}^{(k)}$は$\boldsymbol{G}^{-1}$に関して互いに共役であり，方向ベクトル$\boldsymbol{d}^{(k)}$を共役ベクトルとして選ぶことができる。

というのは，　$\mathbf{d}^{(l)T}\mathbf{G}\mathbf{H}^{(k)}=\mathbf{d}^{(l)T}\quad(l=1,2,\cdots,k-1)$であれば，（6.7）と直交関係

$\mathbf{d}^{(l)T}\mathbf{g}^{(k)}=0\quad(l=1,2,\cdots,k-1)$　　によって

$$-\mathbf{d}^{(l)T}\mathbf{G}\mathbf{d}^{(k)}=\mathbf{d}^{(l)T}\mathbf{G}\mathbf{H}^{(k)}\mathbf{g}^{(k)}=\mathbf{d}^{(l)T}\mathbf{g}^{(k)}=0\quad(l=1,2,\cdots,k-1)$$

となるので $\mathbf{d}^{(0)}$，$\mathbf{d}^{(1)}$，　$\mathbf{d}^{(2)}$，・・・，$\mathbf{d}^{(k)}$が互いに共役となる。

したがって，

$\mathbf{d}^{(l)T}\mathbf{G}\mathbf{H}^{(k)}=\mathbf{d}^{(l)T}\quad(l=1,2,\cdots,k)$を満足するように$\mathbf{B}^{(k)}$を選べばよい。

上式で$i=k$と置き，式（6.7），（6.8）を用いると

$$\mathbf{d}^{(k)T}\mathbf{G}\mathbf{H}^{(k+1)}=\mathbf{d}^{(k)T}\mathbf{G}\left(\mathbf{H}^{(k)}+\frac{\mathbf{d}^{(k)}\mathbf{d}^{(k)T}}{\mathbf{d}^{(k)T}\mathbf{G}\mathbf{d}^{(k)}}+\mathbf{B}^{(k)}\right)=\mathbf{d}^{(k)T}\mathbf{G}\left(\mathbf{H}^{(k)}+\mathbf{B}^{(k)}\right)+\frac{\mathbf{d}^{(k)T}\mathbf{G}\mathbf{d}^{(k)}\mathbf{d}^{(k)T}}{\mathbf{d}^{(k)T}\mathbf{G}\mathbf{d}^{(k)}}=\mathbf{d}^{(k)T} \tag{6.9}$$

となる。よって，$\mathbf{d}^{(k)T}\mathbf{G}\left(\mathbf{H}^{(k)}+\mathbf{B}^{(k)}\right)=0$

が得られる。式(6.5)および式(6.3)より

$$\mathbf{d}^{(k)T}\mathbf{G}=\frac{\mathbf{x}^{(k+1)}-\mathbf{x}^{(k)}}{a_k}\mathbf{G}=\frac{1}{a_k}\left(\mathbf{x}^{(k+1)T}\mathbf{G}-\mathbf{x}^{(k)T}\mathbf{G}\right)=\frac{1}{a_k}\left(\mathbf{G}\mathbf{x}^{(k+1)}-\mathbf{G}\mathbf{x}^{(k)}\right)^T=\frac{1}{a_k}\left(\mathbf{g}^{(k+1)}-\mathbf{g}^{(k)}\right)^T \tag{6.10}$$

であるから，これを式(6.9)に代入すると，次式の直交式

$$\left(\mathbf{g}^{(k+1)}-\mathbf{g}^{(k)}\right)^T\left(\mathbf{H}^{(k)}+\mathbf{B}^{(k)}\right)=0 \tag{6.11}$$

を得る。この式の最も簡単な解は

$$\mathbf{B}^{(k)} = -\frac{\mathbf{H}^{(k)}\mathbf{y}^{(k)}\mathbf{y}^{(k1)T}\mathbf{H}^{(k)}}{\mathbf{y}^{(k)T}\mathbf{H}^{(k)}\mathbf{y}^{(k1)}} \tag{6.12}$$

である。ただし，

$\mathbf{y}^{(k)} = \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}$　と 置いた。というのは，式（6.12)の両辺に右から$\mathbf{y}^{(k)}$を乗ずると

$$\mathbf{B}^{(k)}\mathbf{y}^{(k)} = -\frac{\mathbf{H}^{(k)}\mathbf{y}^{(k)}\mathbf{y}^{(k1)T}\mathbf{H}^{(k)}\mathbf{y}^{(k)}}{\mathbf{y}^{(k)T}\mathbf{H}^{(k)}\mathbf{y}^{(k)}} = -\mathbf{H}^{(k)}\mathbf{y}^{(k)}$$

となる。これより次式を得る。

$$\left(\mathbf{B}^{(k)} + \mathbf{H}^{(k)}\right)\mathbf{y}^{(k)} = 0$$

$\mathbf{B}^{(k)}$，$\mathbf{H}^{(k)}$は対称行列であるから，両辺の転置をとると式(6.9)が得られた。

一般の評価関数 $f(\boldsymbol{x})$ は（6.2）式のような 2 次形式でない場合が多い。このような場合，上述の手法を適用できるようにするため $\mathbf{H}^{(k)}$ の修正式（6.7）の $\mathbf{A}^{(k)}$ から $\mathbf{d}^{(k)T}\mathbf{G}\mathbf{d}^{(k)}$ を除いておくことが必要である。

式(6.10)を用いると，

$$\mathbf{d}^{(k)T}\mathbf{G}\mathbf{d}^{(k)} = \frac{\left(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right)^T}{a_k}\mathbf{G}\mathbf{d}^{(k)} = \frac{1}{a_k}\left(\mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}\right)^T\mathbf{d}^{(k)} = \frac{1}{a_k}y^{(k)T}\mathbf{d}^{(k)}$$

これより

$$\mathbf{A}^{(k)} = \frac{\mathbf{d}^{(k)}\mathbf{d}^{(k)T}}{\mathbf{d}^{(k)T}\mathbf{G}\mathbf{d}^{(k)}} = \frac{\alpha_k\mathbf{d}^{(k)}\mathbf{d}^{(k)T}}{\mathbf{y}^{(k)T}\mathbf{d}^{(k)}} = \frac{\left(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right)\left(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right)^T}{\mathbf{y}^{(k)T}\left(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right)}$$

を得る。

6.2 Kolmogorov-Smirnov　検定

データから各 Parameter と混合率を計算して求めた確率密度関数が適切であるかどうかを検定する方法は様々あると思われるが，ここでは Kolmogorov-Smirnov 検定を用いる。

$F_n(x)=\frac{1}{n}\left(X_1,X_2,\cdots X_n\text{の}x\text{を越えないものの個数}\right)$

を経験分布関数（emperical distribution function）という。

経験分布関数はとくに分布関数について未知母数を含まない仮説

$P(X_i\le x)=F_0(x)$　　を検定するために用いられる。

仮説が正しいとき

$S_n(x)=\sqrt{n}(F_n(x)-F_0(x))$ が漸近的に Gauss（正規）過程になり

$E(S_n(x))=0$

$Cov(S_n(x),S_n(y)=F_0(x)(1-F_0(y))$

x≦ｙ となる。このことを用いて統計量の漸近分布が求められる。

しばしば用いられる統計量として Kolmogorov-Smirnov　統計量

$D_n=\sup_x\left|F_n(x)-F_0(x)\right|$ がある。

ただし $X_{(r)}$ は，X_1，・・・，X_n の第ｒ順位統計量である。

Kolmogorov-Smirnov 統計量については，仮説のもとでの漸近分布は

$$\lim_{n\to\infty}P\left(D_n>d/\sqrt{n}\right)=2\sum_{r=1}^{\infty}(-1)^{r-1}\exp(-2r^2z^2)$$

になる。

$D_\alpha^*(n)\approx\sqrt{\frac{-\ln\alpha}{2n}}-\frac{0.18}{n}$ 2)

6.3 耐性菌についての解析
（提案する非線形最適化手法を用いた解析）

抗生物質，化学療法剤，紫外線あるいはバクテリオファージなど，細菌が本来生育を阻害されるはずの要因が存在する環境下でも，細菌が生育を続けるようになることがある。このように,抗生物質や代謝阻害剤など細菌に対して発育を阻害したり殺菌する作用をもつ薬剤に抵抗性を示す細菌が発生することがある。

このような，同類の細菌に対して有効な薬剤がまったく無効である細菌を耐性菌という。耐性菌とは,抗生物質などの抗菌剤に対する抵抗性が著しく高くなった細菌。(向島 達 ，他[39]） 耐性菌の出現には次の 2 つの機構があると考えられている。

第 1 は抗菌剤が標的とする細菌の酵素あるいは蛋白質に突然変異が起き,抗菌剤がきかなくなる場合である。(癌等)

もう一つはある細菌が獲得した耐性が､別の細菌に伝達されて新たな耐性細菌が生じる場合がある。(インフルエンザ等)

細菌や真菌など培養可能な微生物については,検査する薬剤を一定の濃度になるよう加えた培地でその微生物が生育可能かどうかの検査（生育阻止試験）が行われる。

それぞれ完全に生育阻止または殺菌が可能であった最低の濃度を，最小発育阻止濃度（minimal inhibitory concentration, MIC）として，その微生物に対する薬剤の効果の指標とする。MIC が小さいほど，薬剤の効果が高い，あるいはその微生物の感受性が高いことを表し，指標値よりも MIC が大きければ，微生物のその薬剤に対する感受性が低い，すなわち薬剤耐性であることになる。

この時,計測した阻止円の直径を判定基準と照合して,抗菌性物質に対する感受性の程度,を感性（S），中間（I），耐性（R）のカテゴリーで判定する。

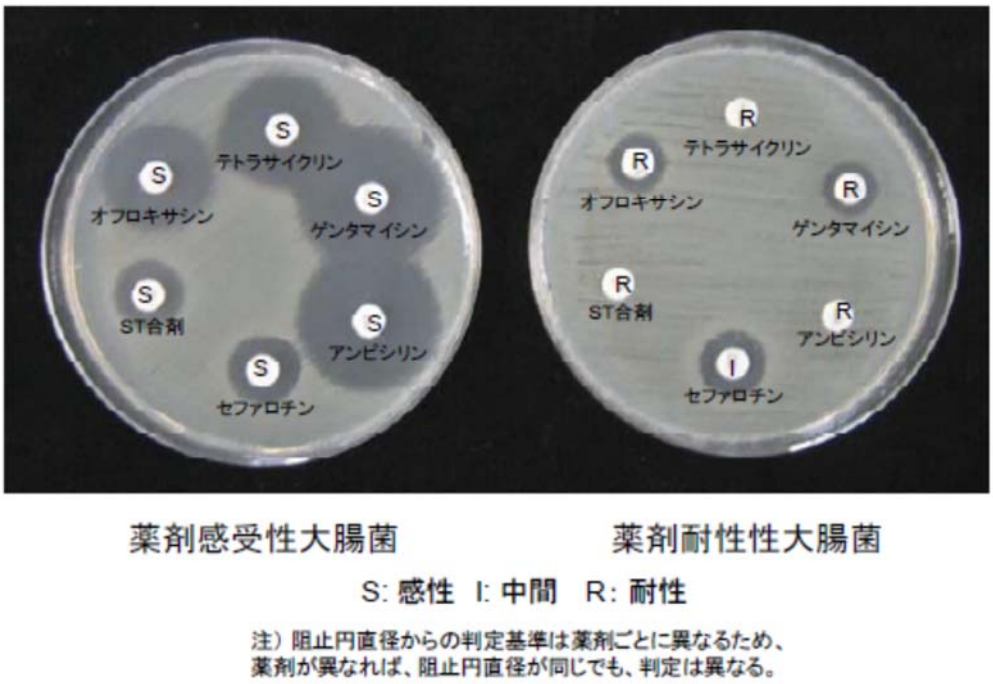


図 6.1　耐性菌の状態（農林水産省ホームページより[40]）

ここでは数値例として臨床細菌検査の阻止円直径の分布のデータを用いている。
データは

ペニシリン系注射剤　　CBPC ｶﾍﾞﾆｼﾘﾝ　　SBPC ｽﾍﾞﾆｼﾘﾝ

タンパク質合成阻害剤（ｱﾐﾉｸﾞﾘｺｼﾄﾞ系）　　SM ｽﾄﾚﾌﾟﾄﾏｲｼﾝ

テトラサイクリン系　　TC ﾃﾄﾗｻｲｸﾘﾝ

の 4 つについて各 552 の阻止円直径データ(mm)である。

実験 6.1　CBPC ｶﾍﾞﾆｼﾘﾝの解析

カベニシリンの阻止円直径の分布のデータの解析，計算結果の表とそのグラフを下に示す。

表 6.1 CBPC の計算結果

CBPC	第一分布	第二分布
平均	23.624	24.086
標準偏差	6.842	2.416
混合率	0.263	0.736

反復回数:48　l^2 -norm の極小値:0.002341

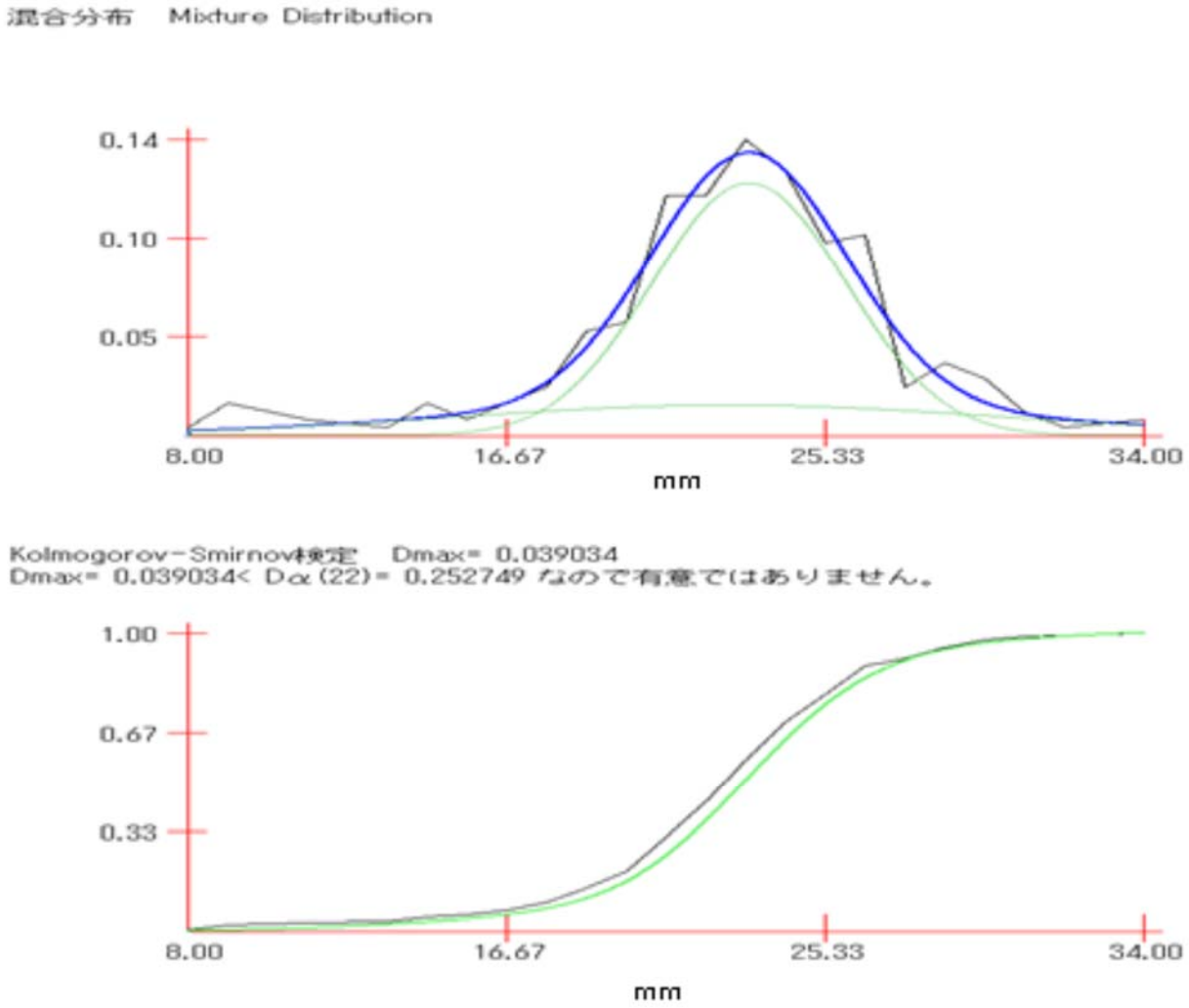


図 6.2 CBPC

青の折れ線が 1 次の Spline 関数表現した入力信号である。青の滑らかな曲線が推定した混合分布である。緑の曲線がそれぞれの要素分布である。

このデータでは，2 つの要素分布の平均値は近い値をとり，標準偏差の異なる 2 つの分布であることが理解される。Kolmogorov-Smirnov 検定の結果差が 0.0390 で漸近分布の値は 0.2527 であるので経験分布と推定した理論分布の有意差は認められない。ただし，この場合，単一分布と判断してもおかしくはない。

実験 6.2　SBPC ｽﾍﾞﾆｼﾘﾝの解析

スベニシリン阻止円直径の分布のデータの解析，計算結果の表とそのグラフを下に示す。

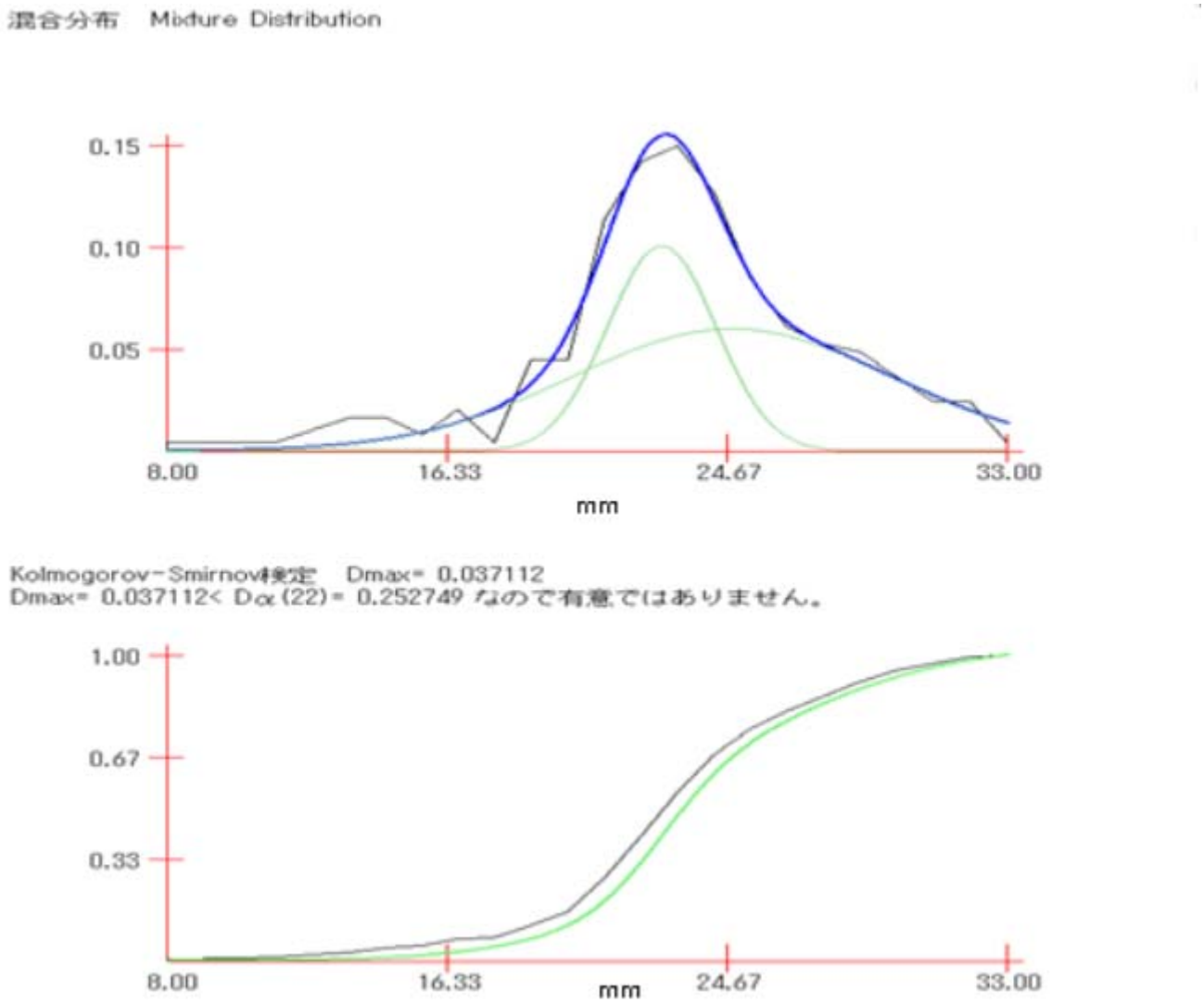


図 6.3 SBPC

表 6.2 SBPC の計算結果

SBPC	第一分布	第二分布
平均	25.472	23.548
標準偏差	4.393	1.487
混合率	0.638	0.362

反復回数:27　l^2 -norm の極小値:0.001196

このデータでは平均値は約２の差があり標準偏差も異なる２つの分布の混合していることが理解される。Kolmogorov-Smirnov　検定の結果も差が 0.0371 で漸近分布の値は 0.2527 であるので経験分布と推定した理論分布の有意差は認められない。
収束への反復回数は CBPC の約半分である。

実験 6.3　　SM ｽﾄﾚﾌﾟﾄﾏｲｼﾝの解析

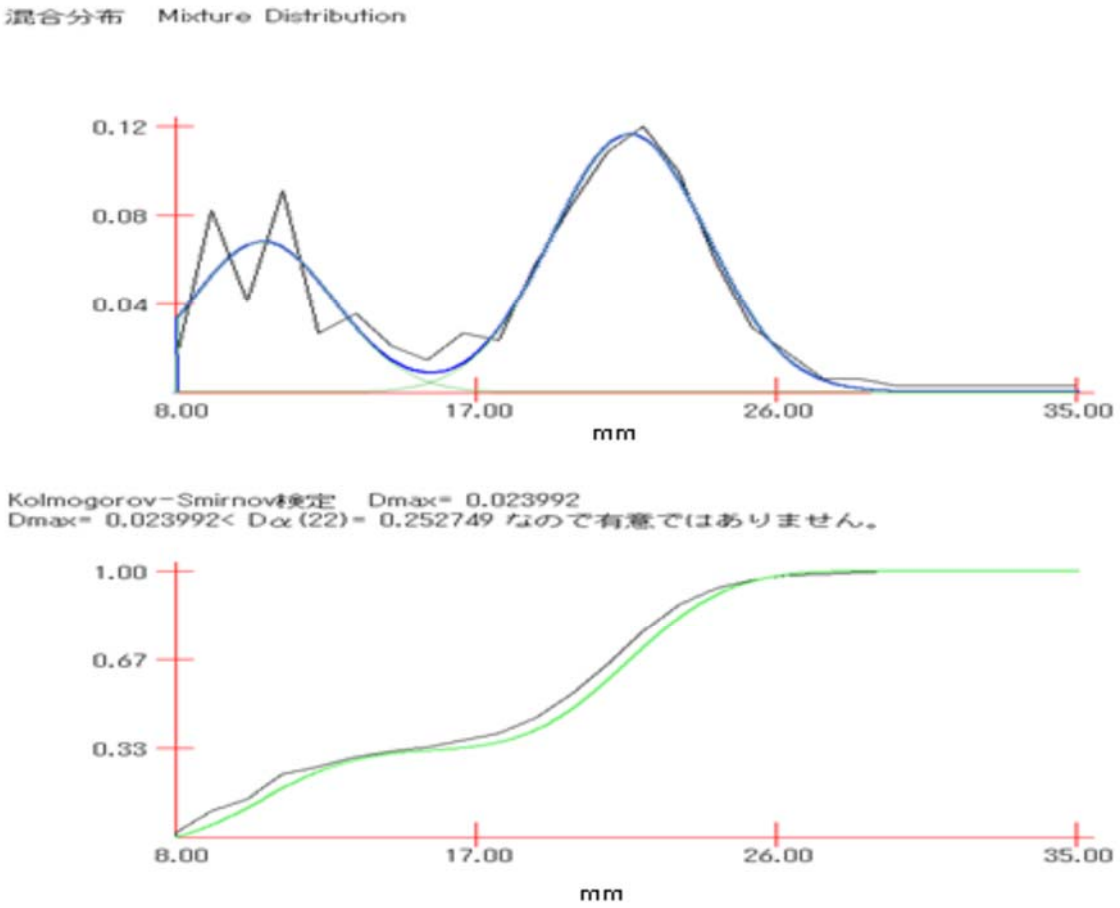


図 6.4 SM

表 6.3 SM の計算結果

SM	第一分布	第二分布
平均	12.417	22.607
標準偏差	1.984	2.165
混合率	0.349	0.651

反復回数:17　l^2 -norm の極小値:0.003722

このデータでは平均値は大きく異なり，標準偏差は近い値をもつ 2 つの分布が混合している，ただし標準偏差が小さい第一分布のピークの高さが低いのは第一分布の混合率の方が小さいためであることが理解される。Kolmogorov-Smirnov　検定の結果差が 0.0239 で漸近分布の値は 0.2527 であるのでも経験分布と推定した理論分布の有意差は認められない。収束への反復回数は 17 回と少ない回数で収束している。

実験 6.4　　TC ﾃﾄﾗｻｲｸﾘﾝﾝの解析

表 6.4 TC の計算結果

TC	第一分布	第二分布	第三分布
平均	13.282	19.969	29.415
標準偏差	3.3	0.467	2.272
混合率	0.537	0.028	0.435

復回数:86　　l^2 -norm の極小値:0.003365

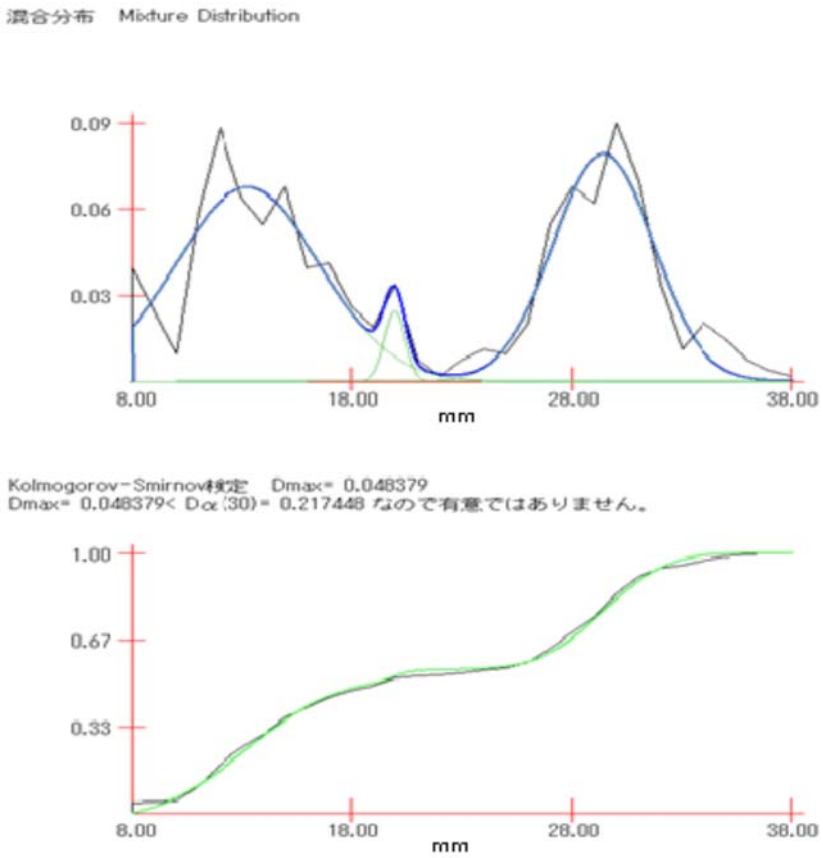


図 6.5 TC

このデータでは平均値は大きく異なり，標準偏差も異なる値をもつ 3 つの分布が混合している，ただしピークの高さの違いは，混合率の違いであることが理解される。

Kolmogorov-Smirnov　検定の結果差が 0.0483 で漸近分布の値は 0.2174 であるのでも経験分布と推定した理論分布の有意差は認められない。

収束への反復回数は 86 回と Parameter の数が増えたので反復回数も多い。

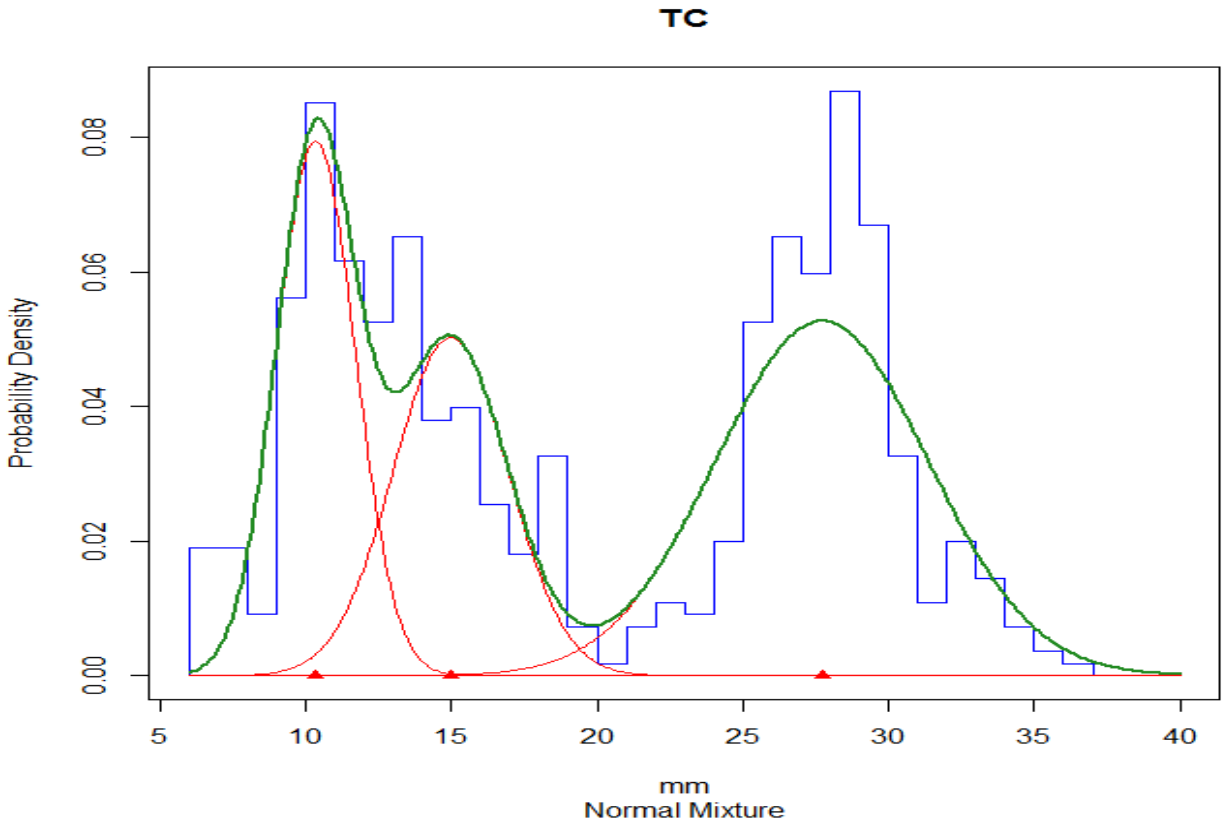


図 6.6 EM アルゴリズムによる TC の解析

表 6.5 TC の計算結果

TC	第三分布	第二分布	第一分布
平均	10.35	14.98	27.73
標準偏差	1.359	1.967	3.641
混合率	0.2708	0.2477	0.4815

図 6.5，図 6.6　ともに 3 つの分布が確認できる

実験 6.1 CBPC ， 実験 6.2 SBPC，実験 6.3 SM については，阻止円直径が 8mm のデータが 552 の四分の一以上であったので耐性菌の存在があるものとして阻止円直径が 8mm のデータを除外して解析を行った。

6.4 品質管理問題への応用

ある医薬品関連の企業の製品特性の管理への混合分布の解析の適用について述べる。

この企業のある製品の pH 強度は中世から弱酸性を目標値としている。このとき製品 1041 個の統計量を計算すると平均 5.704　標準偏差 0.2085 であった。どちらも，規格内であったが，5.4 を中心とする山が気になった。そこで，pH の検査結果を混合分布で解析すると図 6.5，表 6.5 のようになった。

表 6.6 pH の計算結果

	第一分布	第二分布
平均	5.399	5.8
標準偏差	0.082	0.127
混合率	0.25	0.75

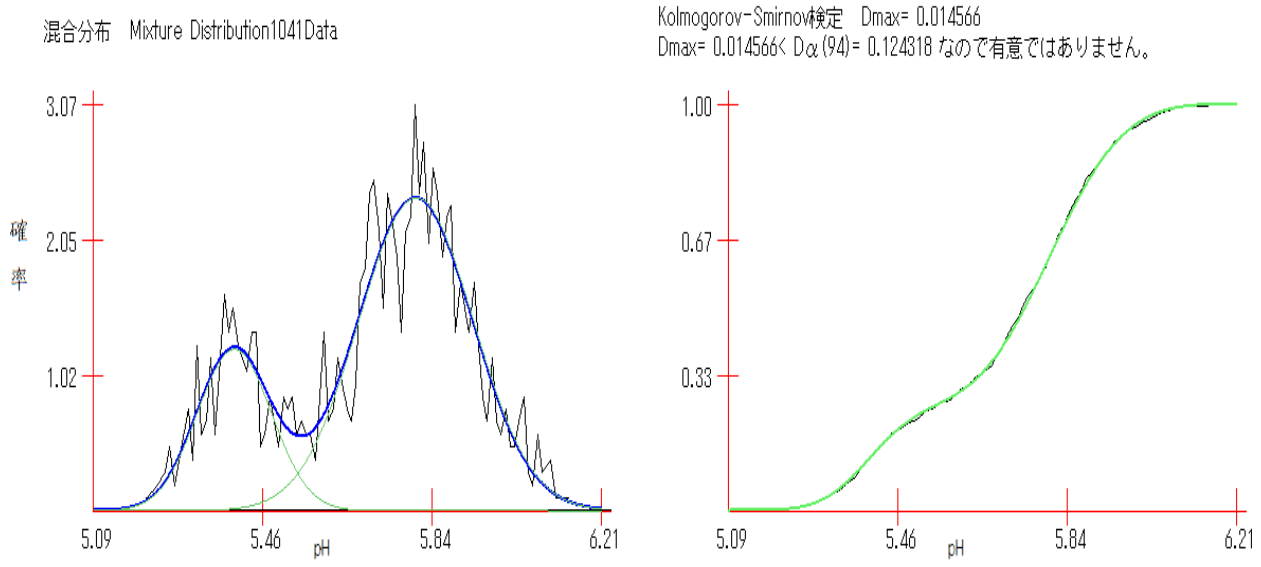


図 6.7　pH 強度

その結果，製造ライン，ロット番号，検査工程，材質仕入先等の各種の項目を調べた。結果として材質仕入先が 2 社あり，仕入れ量が約 1:3 であった。仕入れ量の多い方の取引先企業に限定すれば標準偏差は 0.606 倍になり小さくなる。したがって，品質工学[41]では品質は損失関数 $L=k\sigma^2$ で評価する。 k は機能限界と製品が機能限界外になった時の平均損失との比で表されるので，品質評価は 2.716 倍にあがる。

6.5 まとめ

耐性菌の分布状況を調べることにより，薬が細菌に有効でなりなってきている状況が解り，薬剤濃度を濃くすることにより薬剤効果を高めたり，量を増すことにより薬剤効果を高めたりすることにより院内感染等を防ぐための予防処置をとることができる。
また，新たな抗薬剤細菌の出現等が予測できる。（インフルエンザの菌） さらに，6.4 品質管理問題への応用で述べたように品質評価の向上につながる解析を行った。

今回， l^2 ノルムを最小化する方法を用いて混合分布の構成要素の分布の parameter を推定した。ここで用いた手法は，無理やり，要素数とその初期値を与えて，非線形最適化手法という腕ずくの方法を用いての解法である。この方法は，場合によっては結果が不安定になる場合がある。

尤度関数を用いた E-M Algorithm はよく用いられる手法だが，非線形最適化手法を用いた方法は，最小２乗誤差を与えるという，統計量の性質が，統計学者のプライドが許さないのか文献を見かけない。

7. 提案する正規混合分布の解析方法 2 (Wavelet 解析による正規混合分布の解析方法)

この章では，Wavelet 解析を用いて，混合分布の要素分布の推定する方法を提案する。ここで提案する方法は，4 章で解説した E-M Algorithm による方法，６章で提案した非線形最適化手法を用いた解析方法と異なり,初期値の入力が必要ない方法であることが特徴である。

7.1 Wavelet 解析について

連続 Wavelet 変換は，信号 $f(t)$ に基底関数として Wavelet 関数(Mother-Wavelet)を乗算し,対象時間全体に渡る和を計算する[41]。

$$CWT(b,a)=\int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right)dt \tag{7.1}$$

上式で得られる CWT を Wavelet 係数と呼ぶが，Wavelet 係数は，Scale Factor (a)と Translating (Shift 係数)(b)の関数として得られる。

適切な Scale や位置の Wavelet 関数を掛けることで得られる Wavelet 係数は,信号 $f(t)$ にこれらの Wavelet の成分がどのくらい含まれているかを相対的に示す。

適切な Wavelet を選択すれば，ある Scale と位置に対して,高い類似性を示す Wavelet 係数を得ることができる。

Scaling

Wavelet 関数を「Scaling する」とは,Wavelet 関数を時間軸方向に引き伸ばす(または縮める)ことを意味する。

Wavelet 関数の Scaling は，Scale Factor (a)と呼ばれる係数を使っておこないる。

これは CWT の式の τ に相当し,Scale Factor の値を大きくすれば Wavelet は時間軸方向に引

き伸ばされ，値を小さくすれば時間軸方向に縮まる。

Scale Factor の値が小さな Wavelet により信号の高周波数成分を解析することができ，Scale Factor の値が大きな Wavelet により信号の低周波数成分を解析することができる。

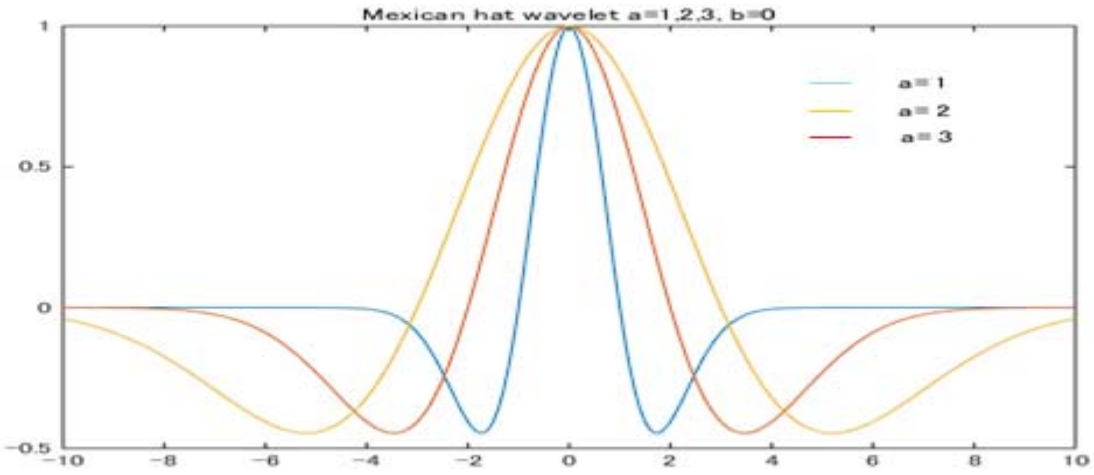


図 7.1　Scaling

Translating (Shift)

Wavelet を「Translating (Shift)する」とは，その波形の立ち上がりを時間的に遅らせる(あるいは早める)ことを意味する。関数 $f(t)$ を k サンプル遅らせることは，数学的に $f(t-k)$ と表現することができる。これは，CWT の式の b に相当する。

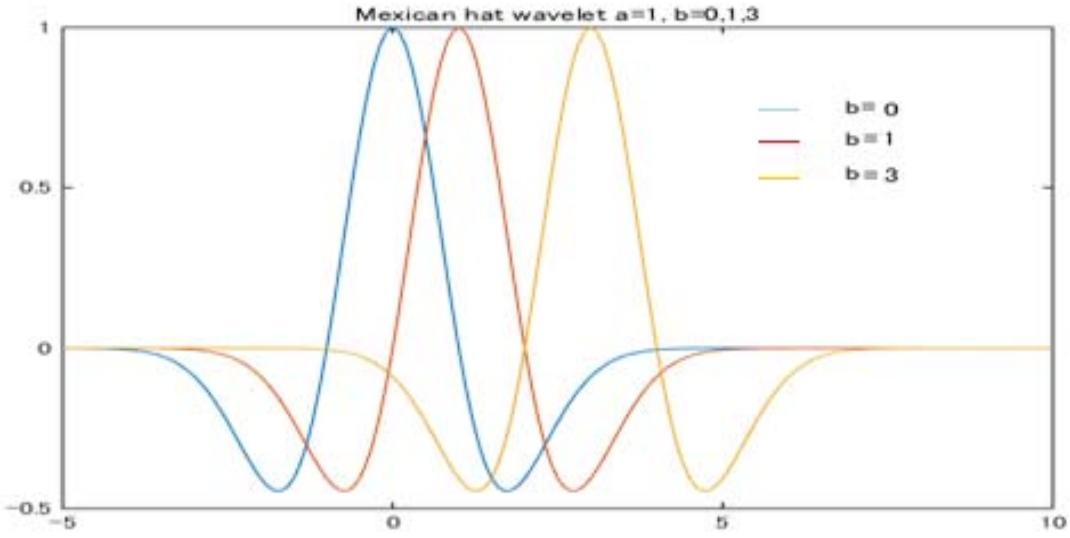


図 7.2　Translating

$\psi(x)$: **マザーウェーブレット**

$\psi\left(\frac{(x-b)}{a}\right)$

a:**スケールパラメータ（時間軸方向に拡大）**

b:**トランスレート　　（時間軸方向に平行移動）**

Wavelet 解析は,Wavelet 関数を用いて観測信号との関係を調べることである。

Scale Parameter：伸縮 拡大 ダイレーション

Translate(Shift)：平行移動,

Mother Wavelet(今回は Gauss 系の Mother Wavelet をもちいる。)：Shift 時間軸上での移動基底関数

$f(x)$を信号関数, $\psi\left(\frac{(x-b)}{a}\right)$を Wavelet としたとき Wavelet 変換は次のようになる。

$$CWT(a,b)=\int_{-\infty}^{\infty} f(x)\frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right)dx \tag{7.2}$$

(7.2)式における$f(x)$に(7.1)式を代入する。また，用いる Wavelet 関数は次の 1 階の Gauss((4)式)あるいは 2 階の Gauss(メキシカンハット)Wavelet 関数((5)式)とする。

$$\psi\left(\frac{x-b}{a}\right)=-\frac{\sqrt{2}e^{-\frac{\left(\frac{x-b}{a}\right)^2}{2}}\left(\frac{x-b}{a}\right)}{\pi^{1/4}} \tag{7.3}$$

$$\psi\left(\frac{x-b}{a}\right)=\frac{2}{\sqrt{3a}}\pi^{-\frac{1}{4}}\left(1-\frac{(x-b)^2}{a^2}\right)e^{-\frac{(x-b)^2}{2a^2}} \tag{7.4}$$

これらは,Gauss Wavelet 関数の 1 階，2 階導関数になる。([43])

この，Wavelet 技法は,ファジィシステムやニューラルネットワークや進化計算といったヒューリスティック的 Algorithm を中心とする計算知能における一つの技法として用いられている。

7.1.1 Wavelet 変換における諸条件

Wavelet 変換に対して微分演算を行うに当たって事前に検討しておくべき諸条件について調べるものとする。

以下では，連続 Wavelet 変換について考える。

Ⅰ Wavelet の条件

Wavelet とは零平均(zero sum) $\int_{-\infty}^{\infty}\psi(t)dt=0$ を満たす $\psi\in L^2(R)$ である。

多くの場合, $\|\psi\|=1$ となるように正規化し, 中心は $t=0$ の近傍におかれる。

‘time-frequency atom ’ $\psi_{b,a}(t)$ の族

a によって ϕ を scaling し, b によって ϕ を translating することにより得られる。

$$\psi_{b,a}(t)=\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{7.5}$$

これらの atom も正規化される。

$$\|\psi_{b,a}\|=1$$

アナライズインク Wavelet とは次の許容条件を満たす Wavelet のことである。

$$\hat{C}_{\psi}\equiv 2\pi\int_{-\infty}^{\infty}\frac{|\hat{\psi}(\xi)|^2}{|\xi|}d\xi<\infty \tag{7.6}$$

ここで，$\hat{\psi}(\xi)$ は ψ のフーリェ変換である。

さらに，ψ をマザーWavelet，$\psi_{b,a}$ を Wavelet と呼ばれる。

許容条件とは次のように定義される。

$$C_{\psi}\equiv 2\pi\int_{0}^{\infty}\frac{|\hat{\psi}(\xi)|^2}{|\xi|}d\xi<\infty \tag{7.7}$$

Ⅱ Wavelet 変換の条件

$f\in L^2(R)$ の ’ time b, scale a’ における Wavelet 変換 $CWf(b,a)$ とは

$$CWf(b,a)=\langle f,\psi_{b,a}\rangle=\int f(t)\frac{1}{\sqrt{a}}\psi^*\left(\frac{t-b}{a}\right)dt \quad 。$$

ここで，$\psi^*(\bullet)=\bar{\psi}(\bullet)$（共役複素関数)。

この定義より次の不等式が成立する。

$$|CWf(b,a)|\le\|f\| \tag{7.8}$$

さらに次の命題が成り立つ

命題１　$\left(\hat{C}_{\psi}<\infty\ \text{を仮定}\right)$

任意の $f\in L^2(R)$ にたいして

$$f(t)=\frac{1}{C_{\psi}}\int_0^{\infty}\int_{-\infty}^{\infty}CWf(b,a)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)db\frac{da}{a^2} \tag{7.9}$$

$$\int_{-\infty}^{+\infty}\left|f(t)\right|^2dt=\frac{1}{C_{\psi}}\int_0^{\infty}\int_{-\infty}^{\infty}\left|CWf(b,a)\right|^2db\frac{da}{a^2} \tag{7.10}$$

7.1.2 Wavelet 変換におけるｂ＝0 点を取る理由

Wavelet 変換においては，低い周波数では window 幅を広く，高い周波数では window 幅を狭くすることができる。これは，Spline 関数表示による確率密度関数が Kernel 関数における可変の Band 幅と同一であるように，確率密度関数の推定における可変の Band 幅を取ることである。いま，信号関数として $N(3.75,0.75^2),N(6,0.5^2)$ の和の分布（$N(3.75,0.75^2)+N(6,0.5^2)$）を考える。

全ての場合に，成立するわけではないが，図 7.3 に見るように正規分布では，１次の導関数が 0 の与える点，２次の導関数が 0 の点は，それぞれ平均と標準偏差に関係している。

よって，ここでは統計学で用いられる正規分布に基底関数として Wavelet 関数(Mother-Wavelet)を以下に示す，Gauss 系の Wavelet 関数を用いる。このことが正規混合分布では，どのような条件で成立するかを考察する。以下に，各次数の Gauss Wavelte 関数を示す。

0 次の Wavelet 関数　$f(x)=e^{-\frac{x^2}{2}}$　　1 次の Wavelet 関数　$f'(x)=(-x)e^{-\frac{x^2}{2}}$

2 次の Wavelet 関数　$f''(x)=(1-x^2)e^{-\frac{x^2}{2}}$　　3 次の Wavelet 関数　$f'''(x)=(-x^3+3x)e^{-\frac{x^2}{2}}$

4 次の Wavelet 関数　$f^{(4)}(x)=(x^4-6x^2+3)e^{-\frac{x^2}{2}}$

$$CWT(a,b)=\int_{-\infty}^{\infty}f(x)\psi\left(\frac{x-b}{a}\right)dx\qquad\psi\left(\frac{x-b_0}{a_0}\right)=0\Rightarrow CWT(a_0,b_0)=0 \tag{7.11}$$

と推測する。

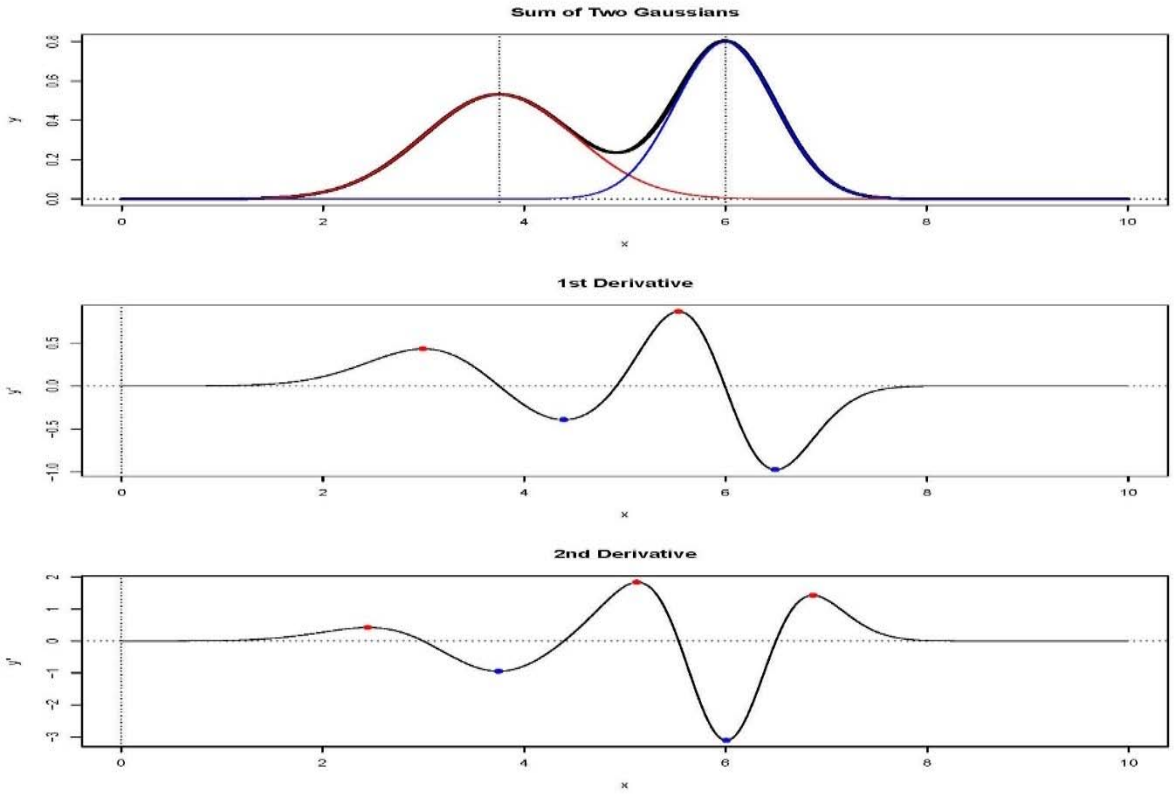


図 7.3　$f(x)=\frac{1}{\sqrt{2\pi}\times 0.75}e^{-\frac{(x-3.5)^2}{2\times 0.75^2}}+\frac{1}{\sqrt{2\pi}\times 0.5}e^{-\frac{(x-6)^2}{2\times 0.5^2}}$ とその 1 階・2 階導関数

このとき,

$$\left|\psi\left(\frac{x-b_0}{a_0}\right)\right|\le\varepsilon \qquad \int_{x_0-\delta}^{x_0+\delta} f(x)\psi\left(\frac{x-b}{a}\right)dx \tag{7.12}$$

$$0\leftarrow\int_{x_0-\delta}^{x_0+\delta}\left|f(x)\psi\left(\frac{x-b_0}{a_0}\right)\right|dx\le S_{\max}\cdot\varepsilon^k\cdot 2\delta=S_{\max}\cdot\varepsilon^{k-1}\cdot 2(\varepsilon\delta) \tag{7.13}$$

$$\left|\psi\left(\frac{x-b_0}{a_0}\right)\right|<\varepsilon^k \qquad where \qquad x_0-\delta\le x\le x_0+\delta \tag{7.14}$$

$\mathrm{k}=1$ の時 $\mathrm{CWT}(\mathrm{a}_0,\mathrm{b}_0)=0$ となるのは

1 次の Gauss 関数では $x=\mu$ が必要条件として与えられる。

$\mathrm{k}=2$ 以上の時 $\mathrm{CWT}(\mathrm{a}_0,\mathrm{b}_0)=0$ となるのは

2 次の Gauss 関数である Mexcanhat 関数では $x=\mu\pm\sigma$

従って, $x_1=\mu+\sigma$, $x_2=\mu-\sigma$ とおくと $\mu=\frac{x_1+x_2}{2}$, $\sigma=\frac{x_1-x_2}{2}$

4 次の Gauss 関数では $x=\mu\pm\sqrt{3\pm\sqrt{6}}\sigma$

今，信号を　$Sig(x)=\alpha\varphi_1(\mu_1,\sigma_1|x)+\beta\varphi_2(\mu_2,\sigma_2|x)$ とおくと

$$CWT(a,b)=\int_{-\infty}^{\infty} f(x)\psi\left(\frac{x-b}{a}\right)dx=\alpha\int_{-\infty}^{\infty} f(x)\varphi_1(\mu_1,\sigma_1|x)dx+\beta\int_{-\infty}^{\infty} f(x)\varphi_2(\mu_2,\sigma_2|x)dx \qquad (7.15)$$

$$\begin{aligned} CWT(a_0,b_0)&=\int_{-\infty}^{\infty} f(x)\psi\left(\frac{x-b_0}{a_0}\right)dx \\ &=\alpha\int_{-\infty}^{\infty}\varphi_1(\mu_1,\sigma_1|x)\psi\left(\frac{x-b_0}{a_0}\right)dx+\beta\int_{-\infty}^{\infty}\varphi_2(\mu_2,\sigma_2|x)\psi\left(\frac{x-b_0}{a_0}\right)dx \end{aligned} \qquad (7.16)$$

上式の第三式の第一項は 0 となるので第二項が≈0 となるようにはどうすれば良いだろうか。

$a_0=\sigma_1, b_0=\mu_1$ とおいて $\frac{(x-\mu_2)^2}{\sigma_2^2}+\frac{(x-b_0)^2}{a_0^2}$ を

$e^{\frac{(x-\mu_2)^2}{2\sigma_2^2}}\cdot e^{\frac{(x-b_0)^2}{2a_0^2}}=e^{\frac{(x-\mu_2)^2}{2\sigma_2^2}-\frac{(x-b_0)^2}{2a_0^2}}=e^{-\frac{1}{2}\left(\frac{(x-\mu_2)^2}{\sigma_2^2}+\frac{(x-b_0)^2}{a_0^2}\right)}=0$　となるように選べばよい。

$$\frac{(x-\mu_2)^2}{\sigma_2^2}+\frac{(x-b_0)^2}{a_0^2}\geq 3 \qquad (7.17)$$

とすれば第二項がほぼ 0 となる。([44])

このような条件を満たせば，平均・分散が求められ，要素分布の分解ができる[32, 33, 34, 35, 37.]。

これを，Wavelet 変換を用いて表現すると図 7.4 のようになる。この図では上段が混合分布の信号関数，中段が 1 次の Gaussian Wavelets 解析，下段が 2 次の Gaussian Waveltes 解析である。

図 7.4 から 0 の等高線の場所の translate 値(横軸の値)を選べばよい。しかし図 7.4 中段，下段の 0 の等高線は曲がっている。そこで，何らかの意味で scale 値(縦軸の値)を選択し横座標に反映させることを考える。このことは 7.3 節で説明する。

図 7.4 から見て分かるように，ほぼ $f(x)=\frac{1}{\sqrt{2\pi}\times 0.75}e^{-\frac{(x-3.5)^2}{2\times 0.75^2}}+\frac{1}{\sqrt{2\pi}\times 0.5}e^{-\frac{(x-6)^2}{2\times 0.5^2}}$ であることが理解できる。

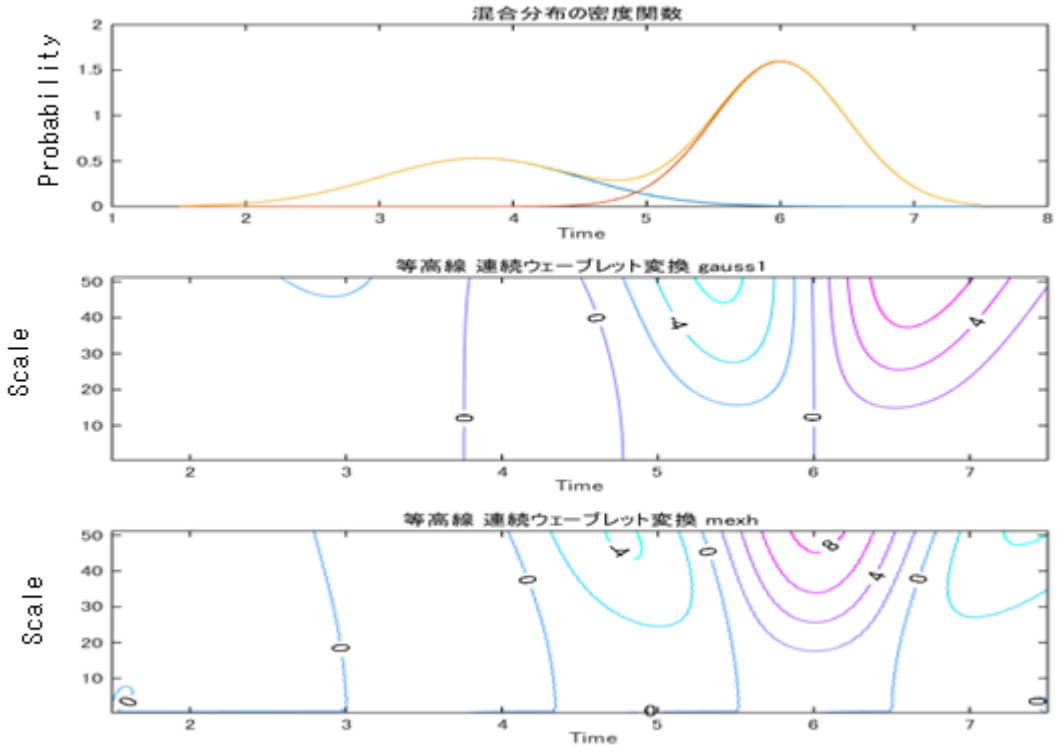


図 7.4 $f(x)=\frac{1}{\sqrt{2\pi}\times 0.75}e^{-\frac{(x-3.5)^2}{2\times 0.75^2}}+\frac{1}{\sqrt{2\pi}\times 0.5}e^{-\frac{(x-6)^2}{2\times 0.5^2}}$ の解析

7.2 連続 Wavelet 変換曲面上の等高線描画 Algorithm

7.1 節で述べたように，連続 Wavelet 変換の等高線が 0 の点を特性値として，取るためには，連続 Wavelet 変換曲面上の等高線の存在が必要になる。そこで，連続 Wavelet 変換曲面上の等高線探索 Algorithm のために，積分記号の元での微分法について述べる。

定理 I

測度空間(X,B,μ)があり， 関数： $f(x,\alpha)$ $(x\in X, a<\alpha<b)$

はxの関数としてはXの上で可積分，αの関数としては微分可能とし，またxの上で積分可能な関数 $\varphi(x)$ ： $|f_\alpha(x,\alpha)|\le\varphi(x)$ on $XX(a,b)$

が存在するものとする。このとき，関数 $\tilde{f}(\alpha)\equiv\int_X f(x,\alpha)d\mu(x)=\int_X\frac{\partial}{\partial\alpha}f(x,\alpha)d\mu(x)$

例 1

公式 $\int_0^\infty e^{-\alpha x^2}dx=\frac{\pi}{2}$ において， x を $\sqrt{\alpha}x(\alpha>0)$ に変換して，

$$\int_0^\infty e^{-\alpha x^2}dx=\frac{1}{2}\sqrt{\frac{\pi}{\alpha}} \tag{7.18}$$

$\bar{a}$ を 1 より小さい正の数とすると上式の左辺の被積分関数を α で n 回偏微分したものについて $\bar{a}<\alpha<\infty$ において

$$\left|e^{-\alpha x^2}\left(-x^2\right)^n\right|\le e^{-\alpha x^2}\cdot x^{2n},\quad \int_0^\infty e^{-\alpha x^2}x^{2n}dx<0 \tag{7.19}$$

が成立するから $n=1,2,\cdots$ として定理Ⅰを逐次適用できる。

従って $\int_0^\infty e^{-\alpha x^2}dx=\frac{1}{2}\sqrt{\frac{\pi}{\alpha}}$ の左辺は α で何回も微分できて，

$$\int_0^\infty e^{-\alpha x^2}\left(-x^2\right)^n dx=\frac{1}{2}\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(-\frac{5}{2}\right)\cdots\left(-\frac{2n-1}{2}\right)\sqrt{\pi}\alpha^{-\frac{1}{2}-n} \tag{7.20}$$

ここでは， $\bar{a}$ は任意に小さくとれるから，任意の $\alpha>0$ に対して

$$\int_0^\infty e^{-\alpha x^2}\left(-x^2\right)^n dx=\frac{1\cdot3\cdot5\cdots(2n-1)}{2^{n+1}}\sqrt{\pi}\alpha^{-\frac{1}{2}-n} \tag{7.21}$$

特に $\alpha=1$ とおけば $\int_0^\infty e^{-\alpha x^2}x^{2n}dx=\frac{1\cdot3\cdot5\cdots(2n-1)}{2^{n+1}}\sqrt{\pi}$

同様にして次の等式が証明される。

$$\int_0^\infty e^{-x^2}x^{2n+1}dx=\frac{n!}{2}\quad\left(\sqrt{\pi}\quad\text{はない}\right) \tag{7.22}$$

例 2

$t>0,\quad x=\left(x_1,x_2,\cdots,x_n\right)\in R^n$ にたいし， $K(t,x)$ を次のように定義する。

$K(t,x)=(4\pi t)^{-\frac{N}{2}}e^{\frac{-\|x\|}{4t}}$　ただし， $\|x\|=\left(x_1^2+x_2^2+x_3^2+\cdots x_{n1}^2\right)^{\frac{1}{2}}$

$f(x)$ を R^n で Lebersgue 可測であって積分可能な，または有界な任意の関数とし

$u(t,x)=\int_{R^n}K(t,x-y)f(y)dy$ とおくと， $t>0,\quad x\in R^n$ において次の各式が成立する。

$$\begin{cases}\dfrac{\partial u(t,x)}{\partial x_j}=\int_{R^n}\dfrac{\partial K(t,x-y)}{\partial x_j}f(y)dy\\ \dfrac{\partial u(t,x)}{\partial x_j\partial x_k}=\int_{R^n}\dfrac{K(t,x-y)}{\partial x_j\partial x_k}f(y)dy\end{cases}\qquad(j,k=1,2,\cdots N) \tag{7.23}$$

$$\frac{\partial u(t,x)}{\partial t}=\left[\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}+\cdots+\frac{\partial^2}{\partial x_n^2}+\right]u(t,x) \qquad (7.24)$$

7.2.1 Mexican hats

ここで,もう一度Gaussian Waveletsとその二階導関数であるMexican hatを表記しておく。

1)Gaussian

$f(t)=e^{-t^2}$ $\left(C^{\infty}\text{関数}\right)$

$\hat{f}(\omega)=\sqrt{\pi}e^{-\frac{\omega^2}{4}}$ f の*Fourie*変換、*Gaussian*

2）Mexican hats （Gaussianの二階導関数）

‘normalized Mexican hat Wavelet’

$$\psi(t)=\frac{2}{\pi^{\frac{1}{4}}\sqrt{3\sigma}}\left(\frac{t^2}{\sigma^2}-1\right)e^{-\frac{t^2}{2\sigma^2}}$$

$$\hat{\psi}(t)=\frac{-8\sigma^{\frac{5}{2}}\pi^{\frac{1}{4}}}{\sqrt{3}}\omega^2 e^{-\frac{\sigma^2\omega^2}{2}}$$ f の*Fourie*変換、

7.2.2 陰関数定理

ここでは，陰関数定理について述べ，次の連続 Wavelet 変換曲面上の等高線描画 Algorithmの存在につなげる。

定理 ‘Implicit function’ 定理

平面領域 G で $f(x,y)$ が連続, (x_0,y_0) の近傍 U で y について偏微分可能で f_y も連続とする。

$f(x_0,y_0)=0, f_y(x_0,y_0)\neq 0$ ならば $x=x_0$ の十分近くで定義された連続関数 $y=g(x)$

$y_0=g(x_0)$, 恒等的に $f(x,g(x))\equiv 0$ であるものが一意的に定まる。

もし，$f(x,y)$ が, (x_0,y_0) において，x についても偏微分可能ならば $g(x)$ は $x=x_0$ において x の関数として微分可能であって

$$g'(x_0)=-\frac{f_x(x_0,y_0)}{f_y(x_0,y_0)}$$ が成立する。

さらに f が U で C^1 級ならば $g(x)$ も C^1 級であって

$$\frac{d}{dx}g(x)=-\frac{f_x(x,y)}{f_y(x,y)} \tag{7.25}$$

が成立する。

系

同条件下でさらに f で C^r 級ならば(r=1, 2, …, ∞)も C^r 級である。

定義

この定理で定められる関数 $y=g(x)$ を $f(x,y)=0$ から定まる(局所的)陰関数という。

7.2.3 連続 Wavelet 変換曲面上の等高線描画 Algorithm の存在

Wavelet 変換の条件で述べた連続 Wavelet 変換を考える。

$$CWf(b,a)=\langle f,\psi_{b,a}\rangle=\int_{-\infty}^{\infty}f(t)\frac{1}{\sqrt{a}}\psi^*\left(\frac{t-b}{a}\right)dt \tag{7.26}$$

ここで，$f(t)$ および $\psi_{b,a}$ は 7.1 節 及び 7.2 節 において，いままで述べたすべての Wavelet の変換条件を満しているものとする。

以下の便宜のため次の記号を定める。

$F(b,a)=CWf(b,a)\quad(-\infty<b<\infty,0<a<+\infty)$

いま，定数 C(=0)に基づく等高線を考える。

$F(b,a)=C$

ただし，次の条件(a), (b)のいずれかを満たす関数 g が存在するものとする。

$$(a)\quad \begin{cases} F(b,a)=C,\quad F_a(b_0,sa_0)\neq 0 \\ a_0=g(b_0),\text{恒等的に}\quad F\big(b,g(b)\big)=C\quad on\quad U(b_0) \\ \dfrac{d}{du}g(b)=-\dfrac{F_u(b,a)}{F_a(b,a)}\quad on\quad U(b_0) \end{cases} \tag{7.27}$$

$$(b)\quad \begin{cases} F(b,a)=C,\quad F_b(b_0,a_0)\neq 0 \\ b_0=g(a_0),\text{恒等的に}\quad F\big(g(a),a\big)=C\quad on\quad U(a_0) \\ \dfrac{d}{da}g(a)=-\dfrac{F_a(b,a)}{F_u(b,a)}\quad on\quad U(a_0) \end{cases} \tag{7.28}$$

以下では，条件(b)が成り立つものとして微分方程式(7.6)の数値解法について考える。更に，以下では ψ は実数値関数とする。上記の記号設定より，

$$F(b,a)=\int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{a}}\psi\left(\frac{t-a}{a}\right)dt,\quad F(b_0,a_0)=0 \tag{7.29}$$

$$\frac{\partial}{\partial a}F(b,a)=\int_{-\infty}^{\infty} f(t)\frac{\partial}{\partial b}\left\{\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)\right\}dt \tag{7.30}$$

$$\therefore \frac{\partial}{\partial u}F(b,a)=-\int_{-\infty}^{\infty} f(t)a^{-\frac{3}{2}}\psi'\left(\frac{t-b}{a}\right)dt$$

$$\frac{\partial}{\partial a}F(b,a)=\int_{-\infty}^{\infty} f(t)\frac{\partial}{\partial a}\left\{\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)\right\}dt=\int_{,-\infty}^{\infty} f(t)\left\{-\frac{1}{2}a^{-\frac{3}{2}}\psi\left(\frac{t-b}{a}\right)+a^{-\frac{3}{2}}\psi'\left(\frac{t-b}{a}\right)\right\}dt \tag{7.31}$$

$$\begin{aligned}\therefore \frac{\partial}{\partial a}F(b,a)&=a^{-\frac{3}{2}}\int_{-\infty}^{\infty} f(t)\left\{-\frac{1}{2}\psi\left(\frac{t-b}{a}\right)+\psi'\left(\frac{t-b}{a}\right)\right\}dt \\ &=a^{-\frac{3}{2}}\left[\int_{-\infty}^{\infty}-\frac{1}{2}f(t)\psi\left(\frac{t-b}{a}\right)dt+\int_{-\infty}^{\infty} f(t)\psi'\left(\frac{t-b}{a}\right)dt\right]\end{aligned} \tag{7.32}$$

$$\therefore \frac{d}{da}g(a)=-\frac{\dfrac{\partial}{\partial a}F(b,a)}{\dfrac{\partial}{\partial u}F(b,a)}\quad on\quad U(a_0)$$

$$u_0=g(a_0),\quad F\big(g(a),a\big)=C \tag{7.33}$$

以上により，連続 Wavelet 変換曲面上の等高線描画 Algorithm の存在が示される。([45])

7.3 Parameter の決定

4 章，式(4.1) は 3 要素の場合だが，ここでは，2 要素の場合を考えてみる。

0 次の Gauss Wavelet 関数　$f(x)=e^{-\frac{x^2}{2}}$ が 0 になるところは，0，一般的には $x=\mu$。

Mexican hat 関数 $f''(x)=(1-x^2)e^{-\frac{x^2}{2}}$ が 0 になるところは，±1 したがって一般的には $x=\mu\pm\sigma$。従って，　$x_1=\mu_1-\sigma_1$，　$x_2=\mu_1+\sigma_1$ とおくと $\mu_1=\frac{x_1+x_2}{2}$，　$\sigma_1=\frac{x_2-x_1}{2}$，　$x_4=\mu_2+\sigma_2$，$x_3=\mu_2-\sigma_2$ とおくと $\mu_2=\frac{x_3+x_4}{2}$，　$\sigma_2=\frac{x_4-x_3}{2}$。

これから，$x_2<x_3$ となるためには $x_3-x_2>0$ よって $\mu_2-\mu_1>\sigma_1+\sigma_2$ となればよい。

一般論として，正規分布関数を考えてみる。図 7.1，図 7.2 に見られる様に，Wavelet 関数は Parameter a によって伸縮され，Parameter b によって平行移動される。正規分布の確率密度関数では平均値は最大値をあたえる点で有るため，そこにおける 1 階の導関数は 0 になる。

$$f'(x)=-\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}(x-\mu)}{\sqrt{2\pi\sigma^2}}=0 \qquad at \quad x=\mu \tag{7.34}$$

また,標準偏差の位置 $x=\mu\pm\sigma$ は 2 階の導関数が 0 となる変曲点の位置である。

$$f''(x)=\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}(x^2-2x\mu+\mu^2-\sigma^2)}{\sqrt{2\pi\sigma^4}}=0 \quad at \quad x=\mu\pm\sigma \tag{7.35}$$

正規分布は，平均をロケーション Parameter とし，標準偏差を ScaleParameter として持つロケーション Scale 密度関数であるという性質と，ScaleParameter a(伸縮 拡大)トランスレート b (平行移動)を基底関数(Mother Wavelet)に対応させて信号解析を行う Wavelet 解析を用いてＧMMの解析を試みる。

まず，1 次の Gauss 型 Wavelet 関数を用いての正規分布を Wavelet 変換してみると次式のようになる。

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\left(-\frac{\sqrt{2}\mathrm{e}^{-\frac{(\frac{x-b}{a})^2}{2}}(\frac{x-b}{a})}{\pi^{1/4}}\right)dx=\frac{\sqrt{2}a\mathrm{e}^{-\frac{(b-\mu)^2}{2(a^2+\sigma^2)}}(b-\mu)}{\pi^{1/4}\sqrt{\frac{1}{a^2}+\frac{1}{\sigma^2}}(a^2+\sigma^2)} \tag{7.36}$$

この式において $a=\sigma \quad and \quad b=\mu$ となれば(7.36)式は 0 となる。

また，2 次の Gauss 型 Wavelet 関数を用いて正規分布関数を Wavelet 変換してみると次式のようになる。

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\left(-\frac{2\mathrm{e}^{-\frac{(\frac{x-b}{a})^2}{2}}(-1+(\frac{x-b}{a})^2)}{\sqrt{3}\pi^{1/4}}\right)dx=\frac{2a^2\mathrm{e}^{-\frac{(b-\mu)^2}{2(a^2+\sigma^2)}}(a^2-(b-\mu)^2+\sigma^2)}{\sqrt{3}\pi^{1/4}\sqrt{\frac{1}{a^2}+\frac{1}{\sigma^2}}(a^2+\sigma^2)^2} \tag{7.38}$$

上式において，充分に小さな a と $b=\mu\pm\sigma$ においては上式の値は 0 となる。幾つかの例を用いて，層別可能な状況を見ていく。

図 7.5 では，単一の正規分布曲線を入力信号として，Wavelet 変換を行う。
図 7.5 a）は入力信号の正規分布曲線を示し，図 7.5 b)では，正規分布曲線の 1 次の Gauss 関数による Wavelet 変換を示した。図 7.5 c）2 次の Gauss 関数(メキシカンハット関数)による Wavelet 変換を示す。

図 7.5 b）の結果，Wavelet 変換は平均値を与える点において 0 値を示している。そしてその結果，図 7.5 c)は，Wavelet の値は平均値の点で極大値をとりる。$\mu-\sigma,\mu+\sigma$ で Wavelet の値がそれぞれ 0 になっていることが示される。

図 7.5 c）に見るように 0 の等高線がカーブしている。従って，Scale 値を何処にするかによって，横軸(Translate 値 b)の値は変わってくる。

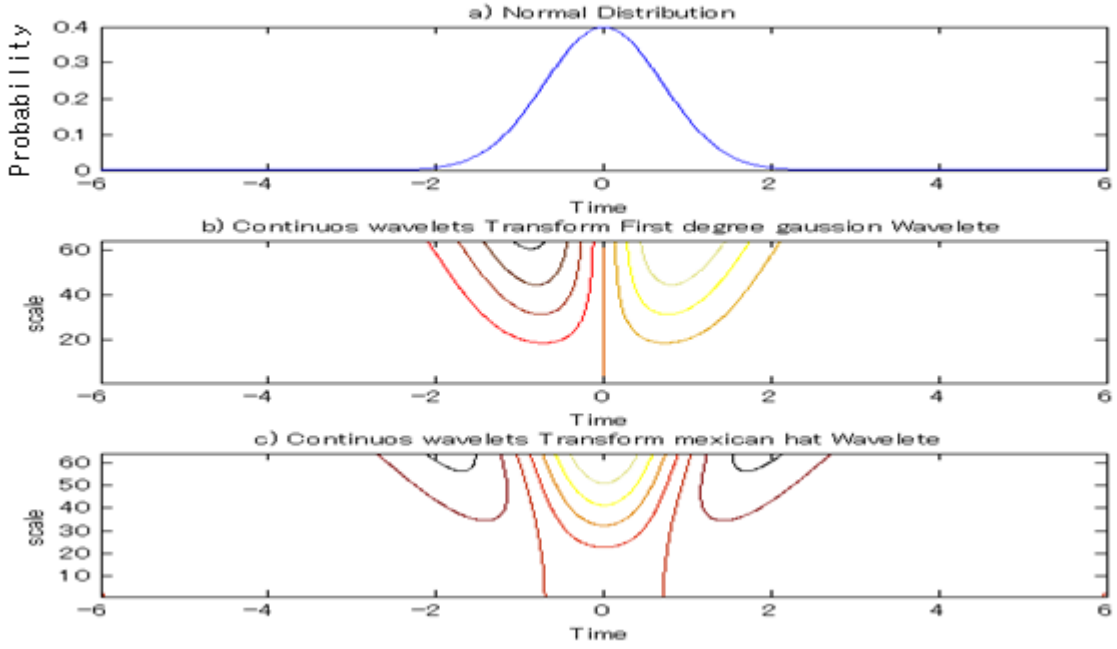


図 7.5 標準正規分布の Wavelet 変換

上段)正規分布曲線　中段）1 次の Gauss・Wavelet 変換　下段）2 次の Gauss(メキシカンハット)Wavelet 変換　を示す。

実験 7.1

$0.5\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-15)^2}{2}}+0.5\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-18)^2}{2}}$ を Wavelet 変換すると次のような結果を得る。

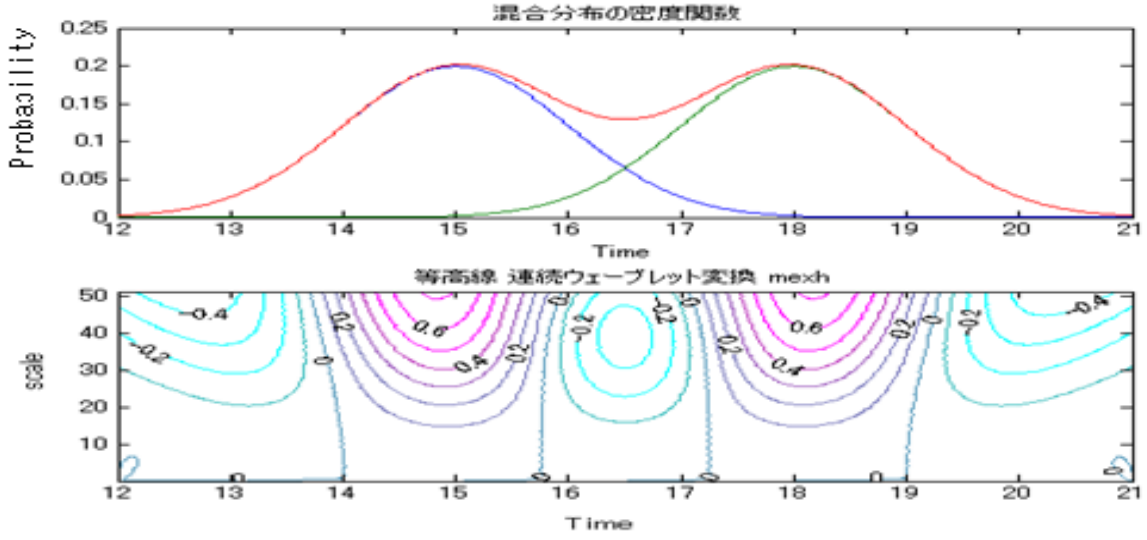


図 7.6 $N(15,1^2)$x0.5+ $N(18,1^2)$x0.　5 とその Wavelet 変換

この例は， $\mu_2-\mu_1>\sigma_1+\sigma_2$ を満たす。また，式(7.17)を満たしている。

x_2, x_3 はそれぞれ N(18, 1^2)， N(15, 1^2)の影響を受け x_2 は約 0.22 小さく， x_3 は約 0.22 大きくなっている。上段)正規分布曲線，下段) 2 次の Gauss(メキシカンハット)Wavelet 変換をしめしている。

実験 7.2

$0.5\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-15)^2}{2\times2^2}}+0.5\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-17)^2}{2\times2^2}}$ を Wavelet 変換すると次のような結果を得る。この実験では $\mu_2-\mu_1=\sigma_1+\sigma_2$ となり二つの分布の Parameter を計算するための，4 つの値を読み取ることが出来ない。式(7.17)を満たしていない。

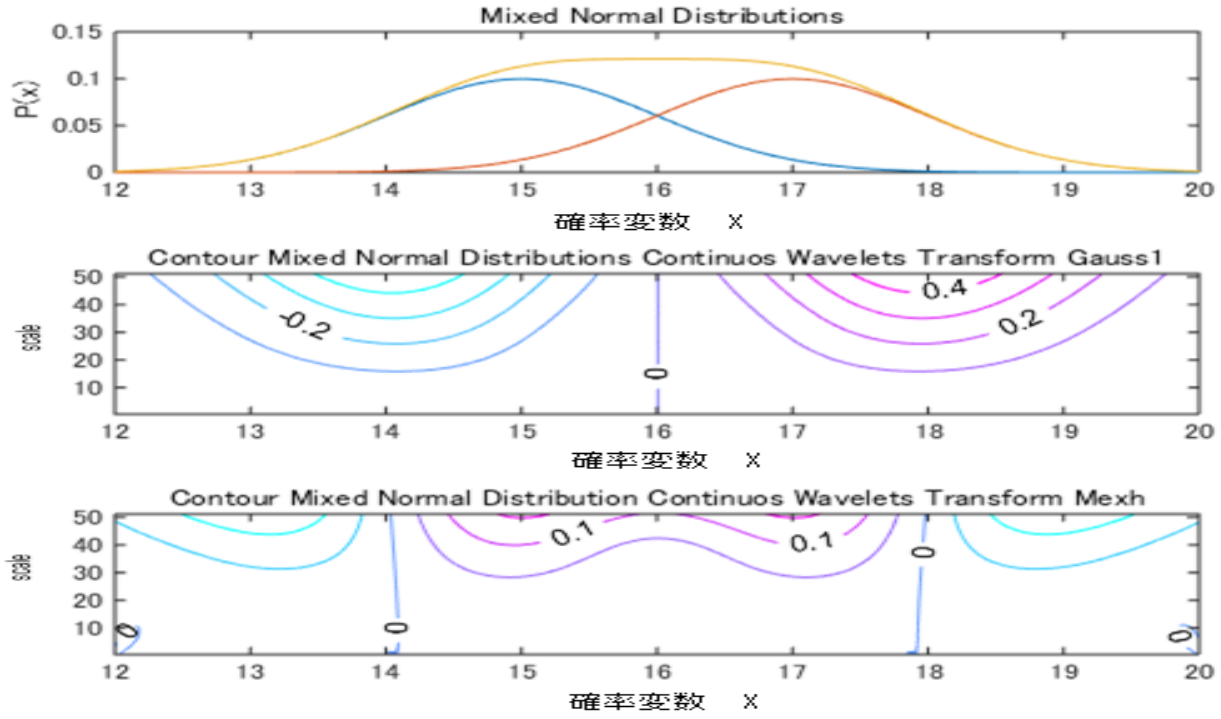


図 7.7 $0.5\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-15)^2}{2\times2^2}}+0.5\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-17)^2}{2\times2^2}}$ の Wavelet 変換

実験 7.3

標準正規分布を Mexican hat 関数を用いて連続 Wavelet 変換した結果はつぎの通りである。ここでは $f(x)$ が正規分布の確率密度関数の和の式で与えられるため Wavelet 関数

は Gauss 関数の関連した Mexican hat 関数 $\psi(x)=\left(\frac{2}{\sqrt{3}}\pi^{-1/4}\right)(1-x^2)e^{-x^2/2}$

及び 4 次の Gauss Wavelet 関数 $\psi(x)=(x^4-6x+3)e^{-x^2/2}$ とする。この結果，平均のところで Wavelet の値がが最大値をとり，Wavelet の値が 0 のところでそれぞれ $\mu-\sigma,\mu+\sigma$ を示している。その結果，平均が 10，分散が 20^2 であることが読み取れる。

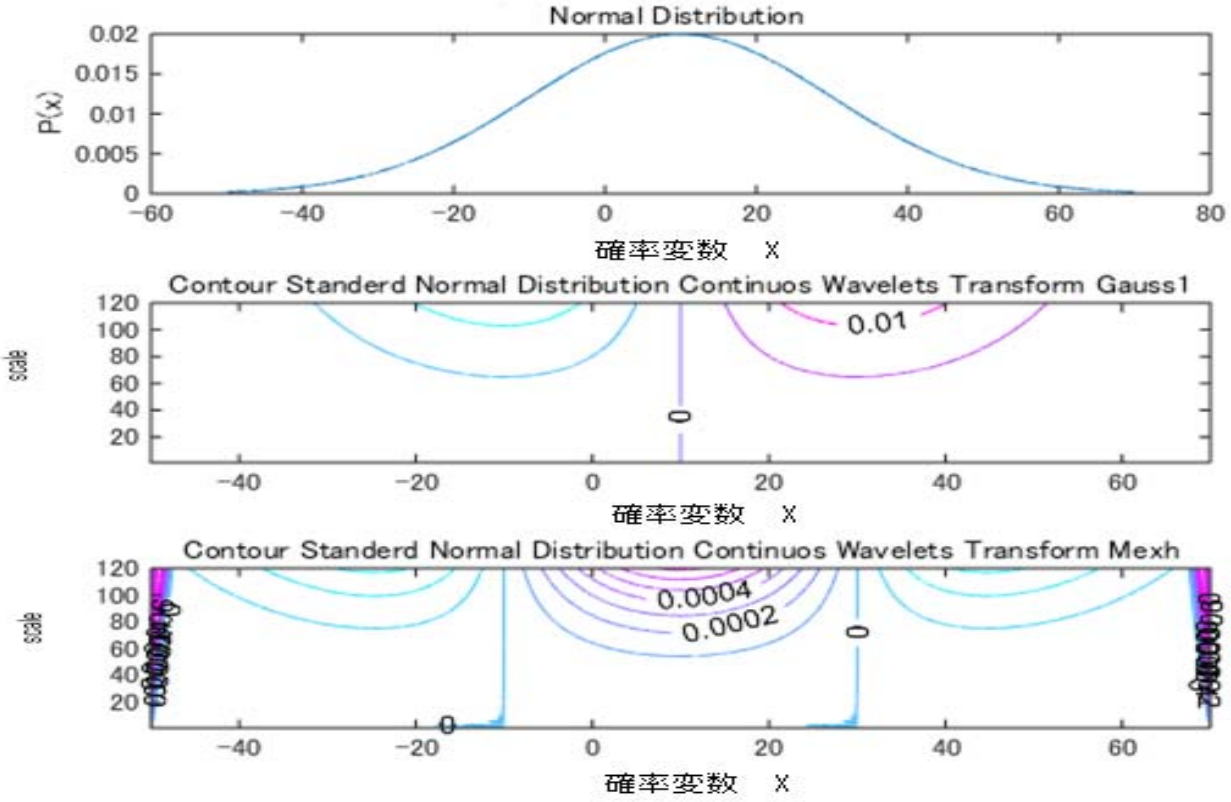


図 7.8 $N(10,20^2)$ とその Wavelet 変換

実験 7.4

$0.7\times\frac{1}{\sqrt{2\pi}}e^{\frac{-(x-15)^2}{2}}+0.3\times\times\frac{1}{\sqrt{2\pi}}e^{\frac{-(x-17)^2}{2}}$ のデータ分布に Mexican hat 関数を用いた場合は次の通りである。

図 7.9 の観測データは平均 15 分散 1，平均 19 分散 1 の正規分布を 7 対 3 の割合で加えたものである。 図 7.9 より，0 点を左から x_1,x_2,x_3,x_4 とすると $x_1=14,x_2=16,x_3=18,x_4=20$ が読み取れる。

$$\begin{cases} x_1 = \mu_1-\sigma_1 \\ x_2 = \mu_1+\sigma_1 \end{cases} \quad \begin{cases} x_3 = \mu_2-\sigma_2 \\ x_4 = \mu_2+\sigma_2 \end{cases}$$

したがって，上式より

$$\mu_1=\frac{1}{2}(x_1+x_2),\quad \sigma_1=\frac{1}{2}(x_2-x_1),\quad \mu_2=\frac{1}{2}(x_3+x_4),\quad \sigma_2=\frac{1}{2}(x_4-x_3).$$

$\mu_1=15,\ \sigma_1=1,\ \mu_2=19,\ \sigma_2=1$ となる。

ω_1,ω_2 も簡単な連立方程式により求めることが出来る。

特定の点 α における観測データの値を f_α とする。

次の連立方程式が成り立つ。

$$\begin{cases} \omega_1 + \omega_2 = 1, \\ \omega_1\dfrac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(\alpha-\mu_1)^2}{\sigma_1^2}} + \omega_2\dfrac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(\alpha-\mu_2)^2}{\sigma_2^2}} = f_\alpha. \end{cases}$$

いま，点 18 における混合分布の確率密度関数の観測値を用いて計算すると

$$f_{18}=0.075\quad ,\frac{1}{\sqrt{2\pi}}e^{-\frac{(18-15)^2}{2}}=0.004\quad ,\quad \frac{1}{\sqrt{2\pi}}e^{-\frac{(18-19)^2}{2}}=0.242$$

より $\omega_1=0.7,\quad \omega_2=0.3$ が求まる。

4 次の Gauss Wavelet 関数を用いた場合も類似の計算で求められる。

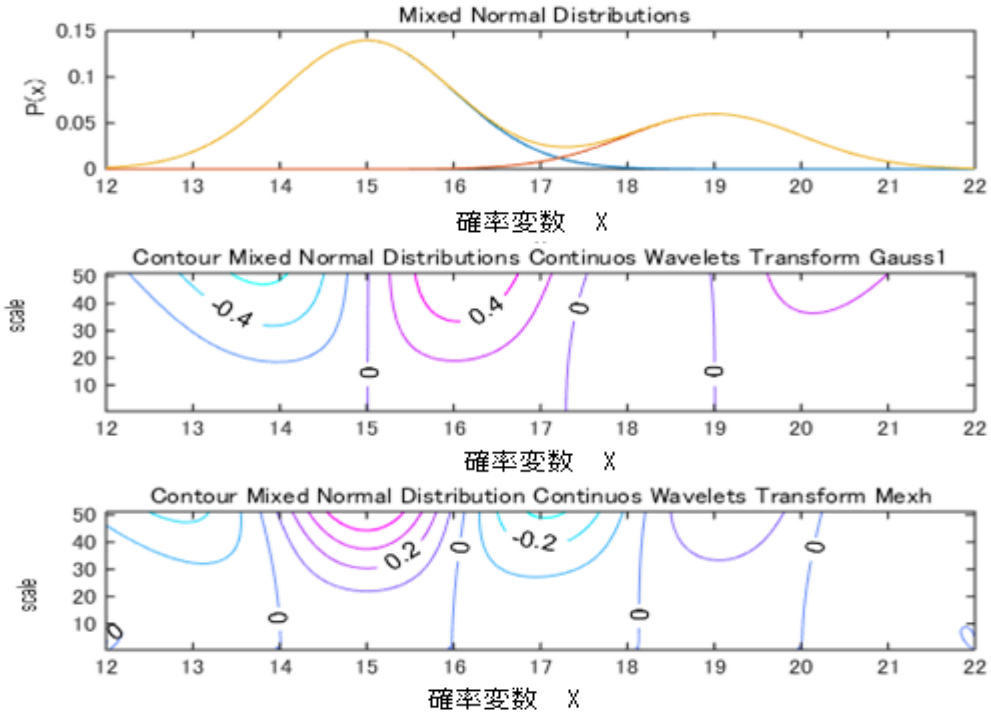


図 7.9　$0.7\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-15)^2}{2}}+0.3\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-19)^2}{2}}$ のサンプルデータの Wavelet 変換

実験 7.5

等平均，異なる分散の場合の解析

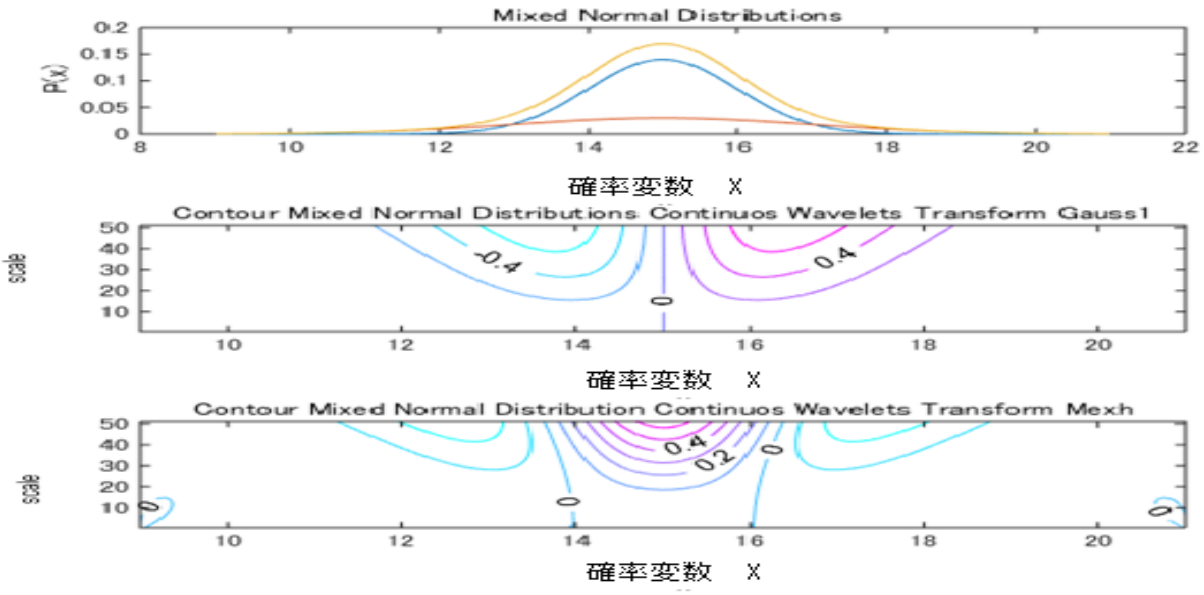


図 7.10 $0.7\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-15)^2}{2}}+0.3\times\frac{1}{\sqrt{2\pi}2}e^{-\frac{(x-15)^2}{2x4}}$ 等平均，異なる分散の Wavelet 変換

実験 7.6

等平均，等分散の場合の解析

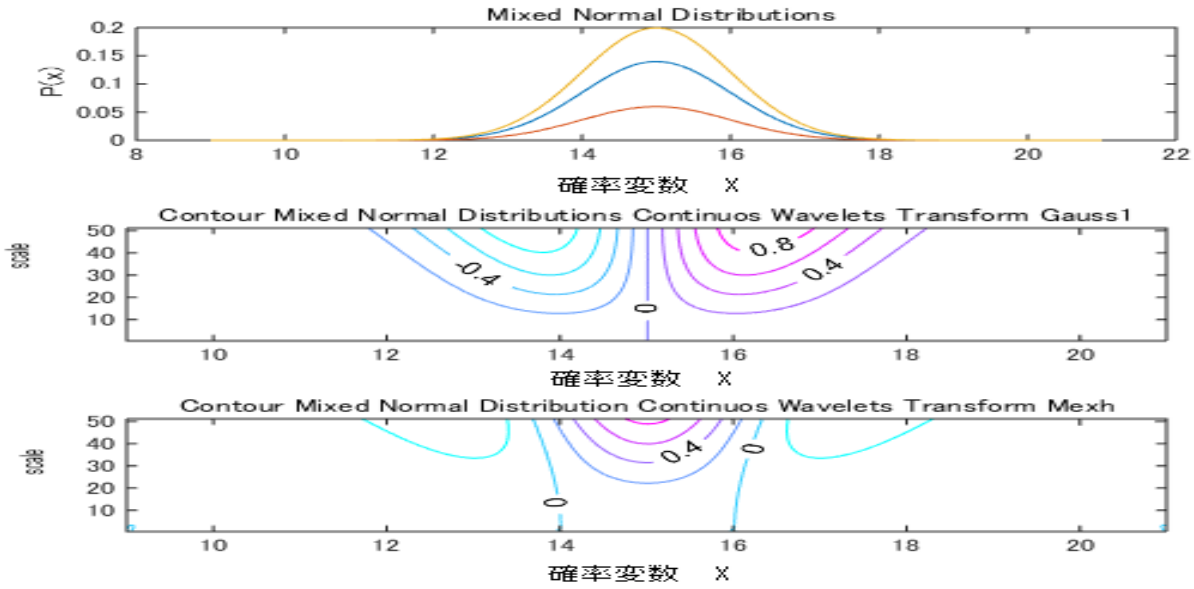


図 7.11 $0.7\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-15)^2}{2}}+0.3\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-15)^2}{2}}$ 等平均，等分散の Wavelet 変換

図 7.10,図 7.11 はどちらも式(7.17)を満たしていないので 2 つの分布を分離できない例である。　図でみたように等高線が 0 になる b は一定の値ではなく，a の値により変化している。そこで，どの a によって b を決めるかが問題になってくる。
何らかの意味で最適であることを示す指標を探さなくてはならない。その指標として，今回は，Wavelets Power Spector を用いる。([43]，[46]）

1 次,2 次(Mexican hat 関数)の Gaussian Wavelet の等高線が 0 になる場所は，それぞれ平均及び標準偏差に関係した値を与える位置としたため，その目的のために Scale・Parameter a の値を決定しなければならない。

Scale a を大きく取ることは,進化計算における大域的探索にあたり，Scale a を小さくとることは局所探索に相当する。そのため，探索レベルを決定するための尺度として Wavelets Power Spector を用いる。信号の与えるスペクトルが最大になる Scale・Parameter を選び，その位置での等高線が 0 の値が与えられる Translate 値 b により標準偏差を示す点とする。
信号 $x(t)$ に含まれていた総エネルギー大きさは,その 2 乗積分されたとして定義される。

$$E=\int_{-\infty}^{\infty}\left|x(t)\right|^2 dt=\left\|x(t)\right\|^2 \tag{7.39}$$

ある Scale 値 a と位置(Translate 値)b での信号エネルギーの相対的値は，二次元の Wavelet・energy 確率密度関数から与えられる。

$$E(a,b)=\left|CWT(a,b)\right|^2 \tag{7.40}$$

この，Spector が最大になるような Scale 値 a を選び，その近辺で Wavelet 変換の等高線が 0 になるような位置(Translate 値) b を Parameter の推定値とすることにより，信号エネルギーが大きくなる推定値ということができる。([46]）

実験 7.5，実験 7.6 の図 7.10,図 7.11 は提案する手法にとっては好ましくない状態であるが，あえて取り上げた。また，次の実験では辛うじて(7.17)式を満たす。精度が良くはないが要素分布の確認ができる例である。

実験 7.7

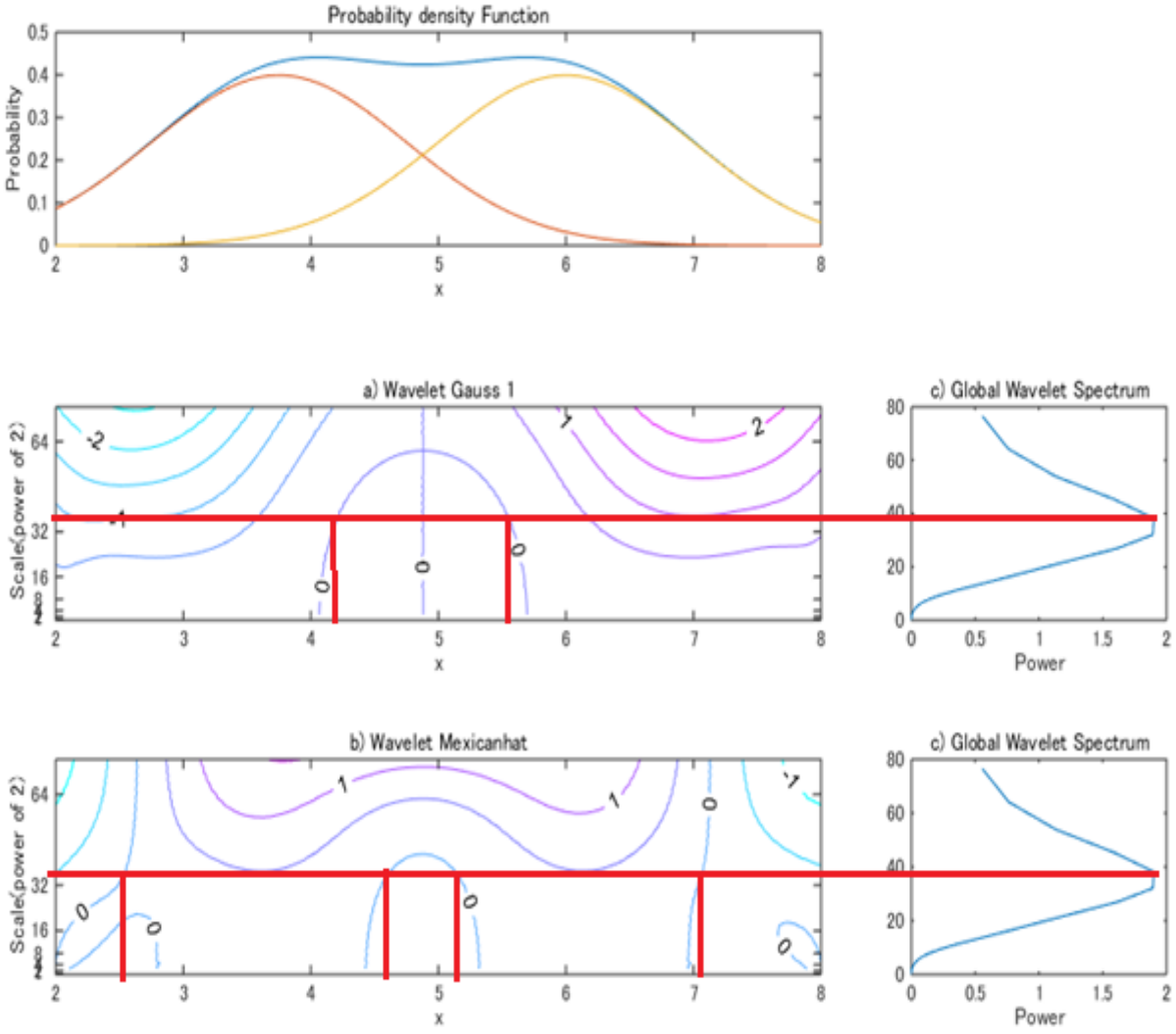


図 7.12 $\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-3.75)^2}{2}}+\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-6)^2}{2}}$ の Wavelet 変換

この実験では(7.17)式を十分満足していない。Parameter の精度も十分満足できるものではない。

実験 7.8

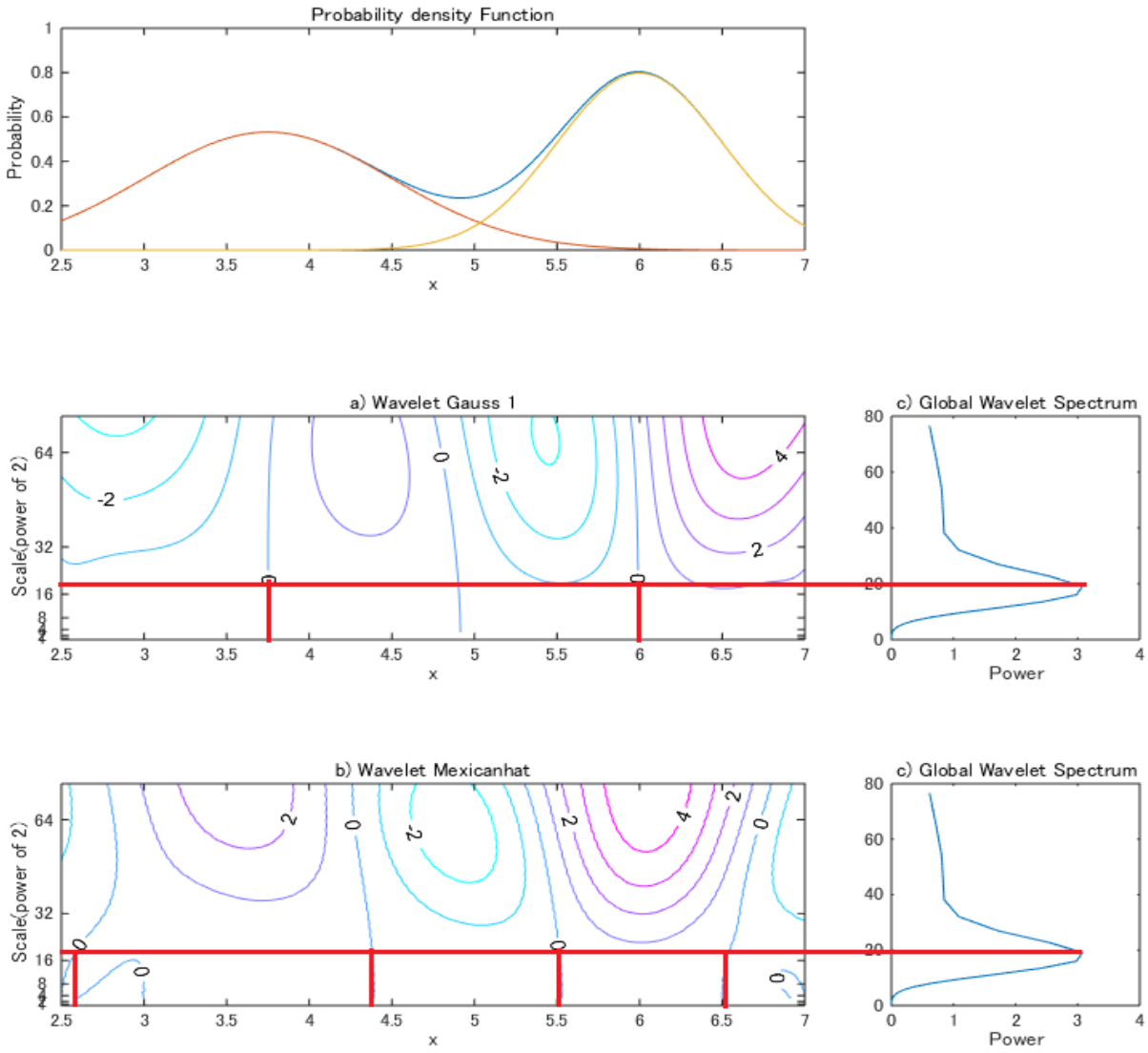


図 7.13 $\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-3.75)^2}{2\times 0.75^2}}+\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-6)^2}{2\times 0.5^2}}$ の Wavelet 変換

この実験では(7.17)式を十分満足している。Parameter の精度も十分満足できるものである。

7.4 花粉飛散データに関する例

杉花粉によって引き起こされた健康上障害は日本の早春における重要な問題である。私たちは，環境省のホームページからダウンロードされる 2004 年の笠間市の観測点での花粉分散データ及び，飯能市の観測点での花粉飛散データを使用する。

環境省はホームページ上で2月1日から5月31日まで毎時間花粉分散状況を発表している。杉花粉問題(花粉症)は,原因になって,日本の杉およびヒノキのような植物の花粉によって,くしゃみと鼻水のようなアレルギーを始める病気である。さらに，それは季節性アレルギー性鼻炎と呼ばれる。関東地方を例として取り上げると，杉花粉の散乱は 2 月はじめに始まり飛散 4 月末には減少する。

また，四月にヒノキ花粉の飛散乱が始まり，5 月末日ぐらいまでそれは継続する。くしゃみ，鼻水，鼻づまり，痒さ，目の異物のような感じは，花粉の散乱の量に比例して，ますます悪化する。

そこで，私たちは，関東地方における花粉の飛散状況を杉およびヒノキの花粉の分布へ分割を試みる。

データは 2 月 1 日から 5 月 31 日までの各時間ごとの花粉のカウント数が記載されているが，ここでは，各日の飛散カウント数を観測時間で割った，その日の一時間当たりの平均個数を用いた。また，年度も多少古いが，2004 年はデータの欠測時間がほとんどなくデータが揃っていたので，この年のデータとした。また，観測地点も所によっては殆どが杉花粉で面白みがないところもあるので，杉と檜が混ざるような地点を選んだ。

杉花粉を主に計測する花粉自動計測器を設置している。この装置は、内蔵された吸引ポンプで大気を吸引してレーザー光を照射すると花粉などの粒子で光が 散乱するので、その散乱光の量から粒子の大きさを判別し、散乱光の数から粒子の数を求めるものである。

（このホームページのデータは、全てこの自動計測器による測定値である)。

しかしながら,この方法での測定は必ずしも花粉だけをカウントし測定している訳ではない。

風の強い日は埃も計測される。特に春先になると，中国からの黄砂なども考えられる。

図 7.14 関東地方観測局名称図

(環境省花粉観測システム http://kafun.taiki.go.jp/)

(環境省花粉観測システム（はなこさん)) より

地元に近い観測所として，群馬県・栃木県のデータを用いて解析したいところだが，この地区は群馬県衛生環境研究所に見学に行ったときに面白みがないと聞いたので，埼玉県・茨城県の観測局のデータを用いた。

実験 7.9 笠間市のデータの解析

茨城県笠間市の花粉飛散データを用いて解析を行う。図からは杉花粉と檜花粉の飛散状況がはっきりと分離していて，典型的な花粉飛散の形である。

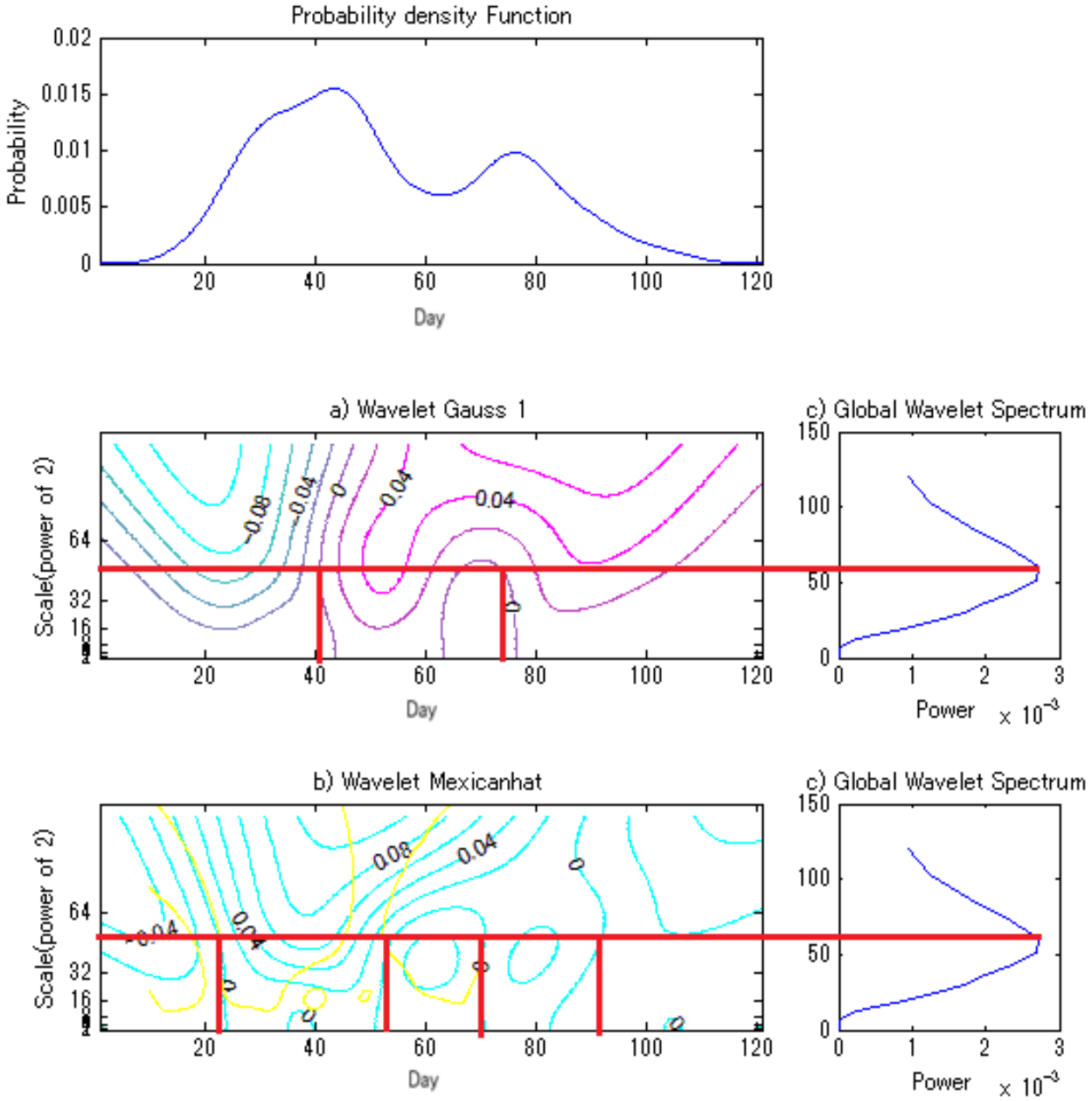


図 7.15　V.D.Spline 関数を入力信号とした Wavelet 解析

図 7.15 の上部の列は，V.D.Spline 関数表現の確率密度関数を示す。また，中間は，1 次の Gauss 型 Wavelet による解析を示す。また，下段はメキシカンハット Wavelet による解析を示す。

次の Parameter は次のようになる。

$$f(x)=0.605\times\frac{1}{\sqrt{2\pi}16}e^{-\frac{1}{2}\left(\frac{x-41}{16}\right)^2}+0.204\times\frac{1}{\sqrt{2\pi}10}e^{-\frac{1}{2}\left(\frac{x-75}{10}\right)^2}$$

この結果を使用して，

Kolmogorov-Smirnov　検定の結果は下に示される。

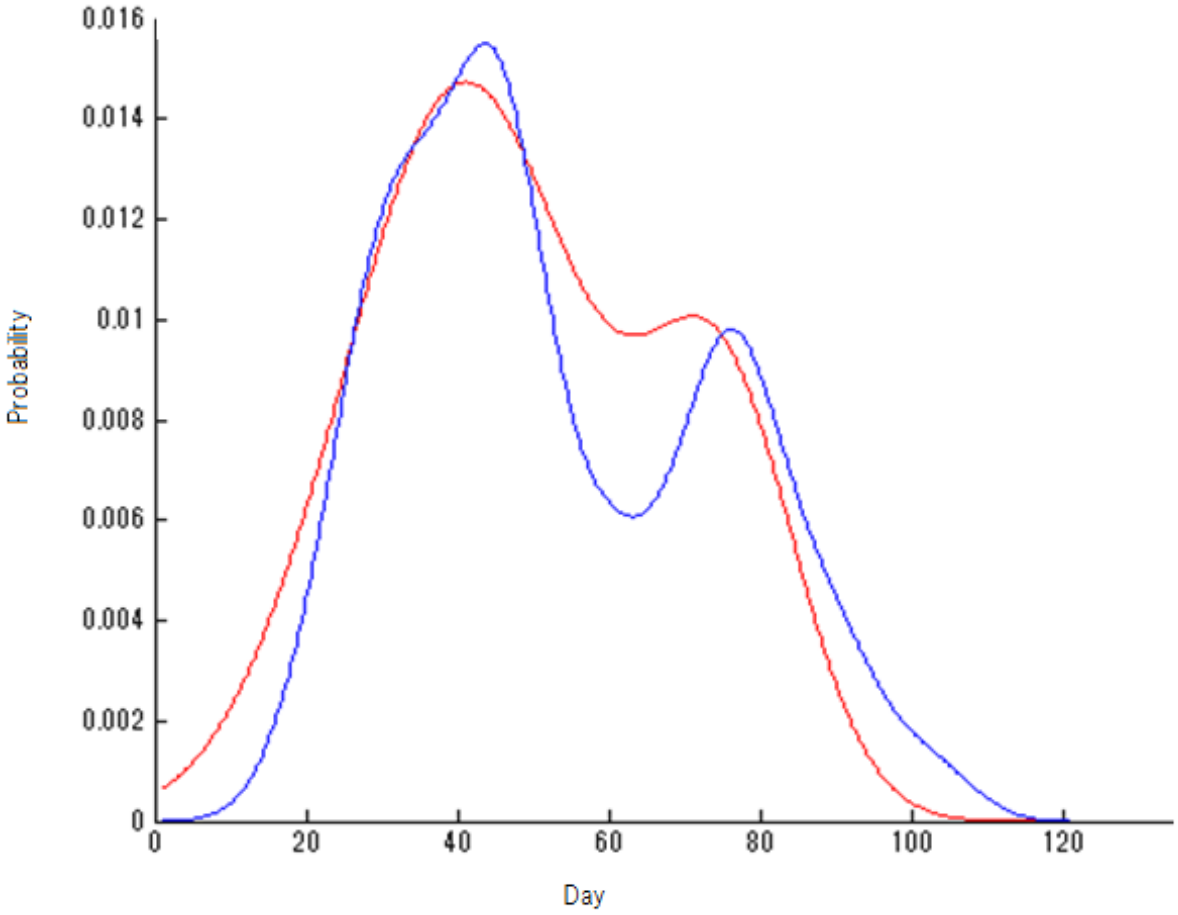


図 7.16 GMM Estimation & Kernel Estimation

図の状況では一見，Kolmogorov-Smirnov 検定で棄却されそうである。

検定は，対立仮説は，Kernel 密度および GMM 推定が異なる連続分布であるということである。結果 h は 1 ならば検定が 5% 有意水準で帰無仮説を棄却である。
0 なら棄却されない。

今，h= 0，テストの漸近の p-値は 0.0691 および検定統計量は 0.18 であるので棄却されない。３月の１０日ぐらいに杉の飛散状況はピークになり，檜花粉は４月２０日あたりが花

粉の飛散状況がピークになる。笠間の杉および檜の花粉飛散の分布状況は 2 つの花粉の飛散時期を分離可能である。

次の図は帯域幅=4 の Kernel 密度への分析である。

特に杉の林業が盛んであるならば 4 月初めからの花粉の飛散分布は小さくなり，観測されない。

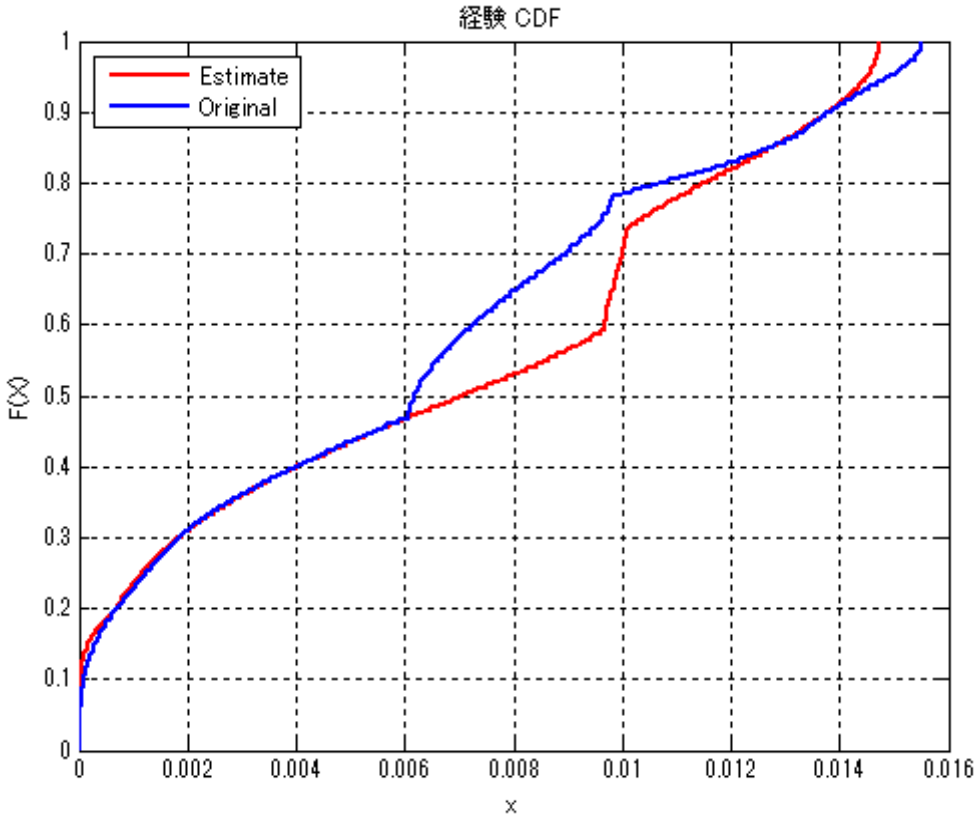


図 7.17　Kolmogorov-Smirnov 検定

実験 7.10 飯能市のデータの解析

埼玉県飯能市の花粉飛散データの解析を行う。ここでは入力信号として V.D.Spline 関数による推定を用いた。

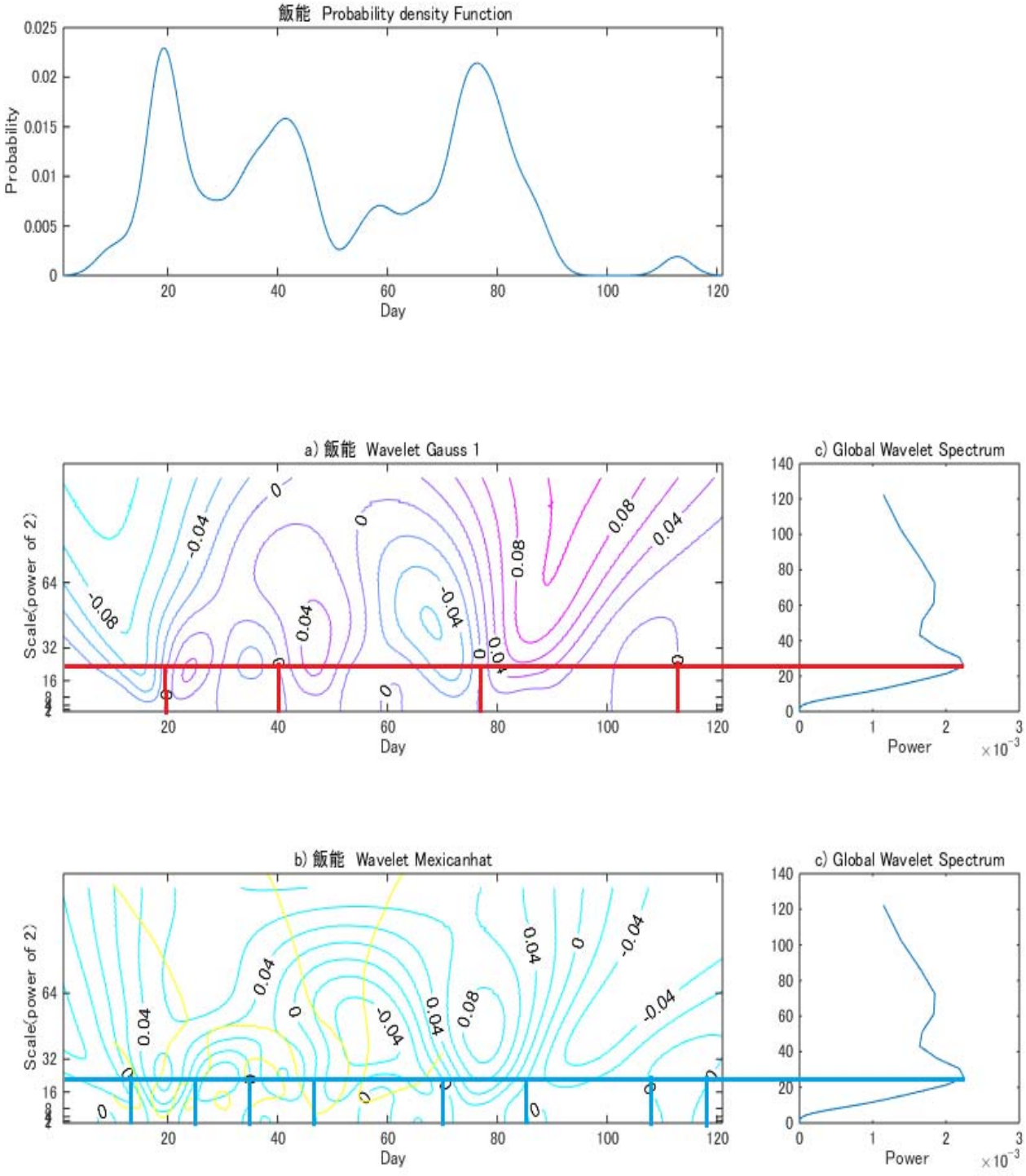


図 7.18 飯能市飛散データ

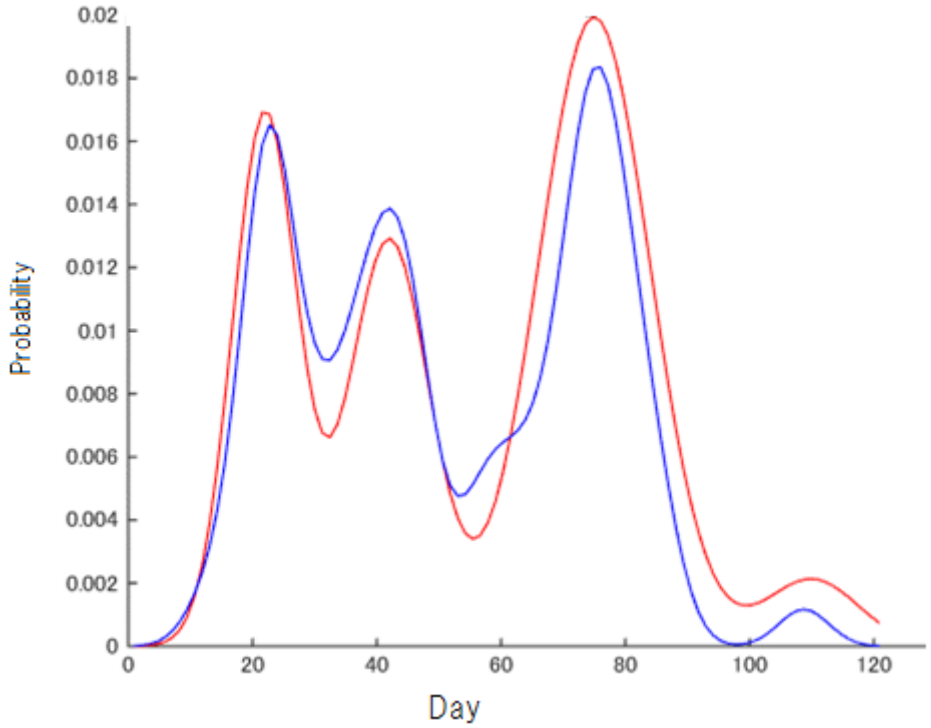


図 7.19 飯能市の当てはめ

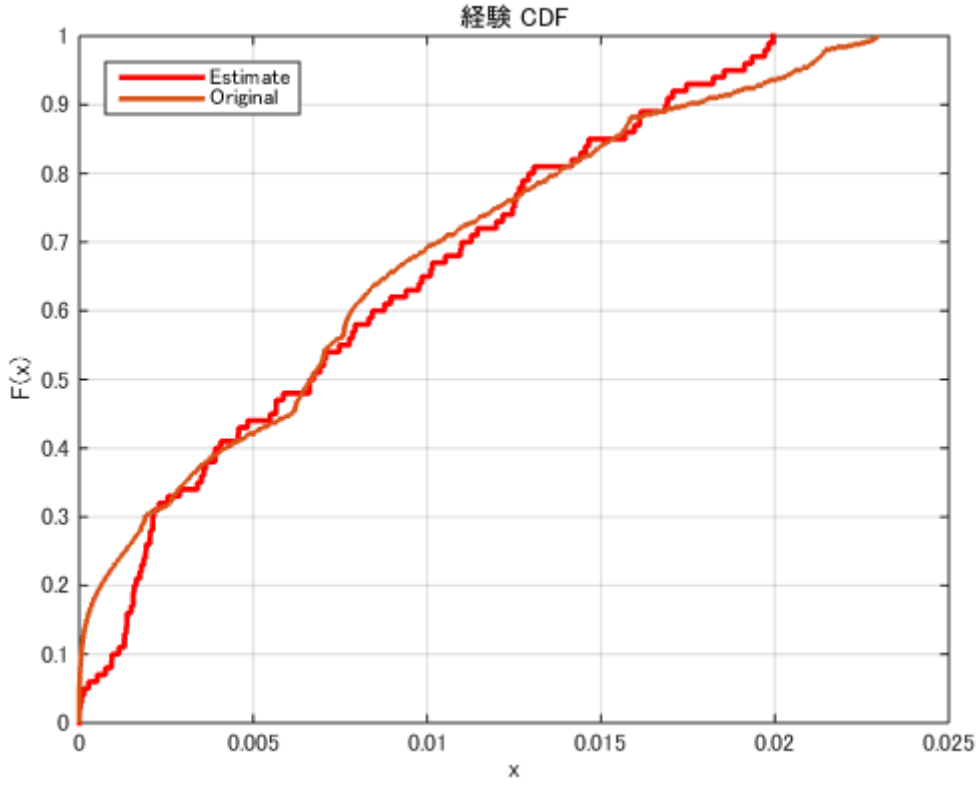


図 7.20 飯能市 Kolmogorov-Smirnov 検定

今，h= 0，テストの漸近の p-値は 0.0528 および検定統計量は 0.1455 である。

飯能では，気象状況によるためか，4 つの飛散状況に分けられる。

$$f(x)=0.22\times\frac{1}{\sqrt{2\pi}5.2}e^{-\frac{1}{2}\left(\frac{x-22}{5.2}\right)^2}+0.21\times\frac{1}{\sqrt{2\pi}6.5}e^{-\frac{1}{2}\left(\frac{x-42}{6.5}\right)^2}+0.45\times\frac{1}{\sqrt{2\pi}9}e^{-\frac{1}{2}\left(\frac{x-75}{9}\right)^2}+0.04\times\frac{1}{\sqrt{2\pi}7.5}e^{-\frac{1}{2}\left(\frac{x-110}{7.5}\right)^2}$$

２月２０日前後にピークが１つあり，３月１０日あたりに小さな山がある。また，４月２０日あたりの山は檜花粉のピークだと思われる。５月２０日前後の小さな山は地名に名栗地区という地区名があることから栗の開花時期かもしれない。

7.5　Wavelet 解析品質管理問題への応用

6.4 節の品質管理問題への応用で用いたデータに Wavelet 解析の手法を用いてみる。

まず，V.D.Spline 関数による入力信号を示すと次のようになる。6.4 節の１次の Spline 関数(折れ線関数)と違い変動を削除して，滑らかに表示されている。

図 3.11 の Kernel 関数法の推定を参照。

実験 7.11

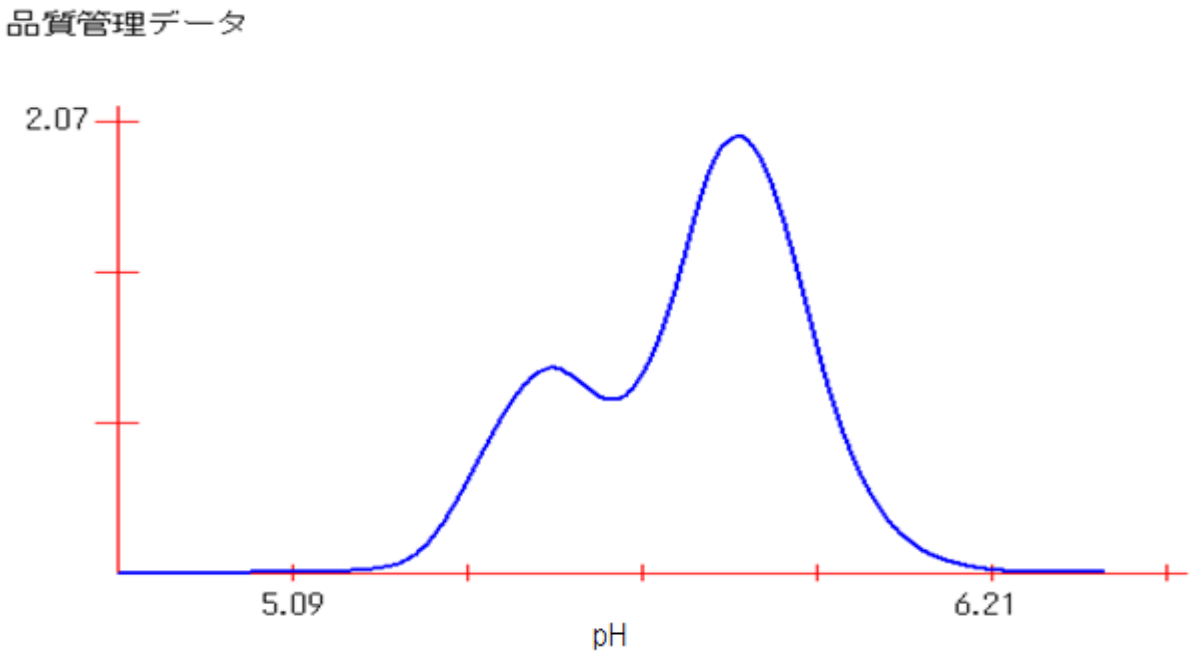


図 7.21 品質管理問題への応用で用いたデータの解析
V.D.Spline 関数による入力信号

この確率密度関数を，Wavelet 解析のへの入力信号として用いて混合分布の解析をおこなう。

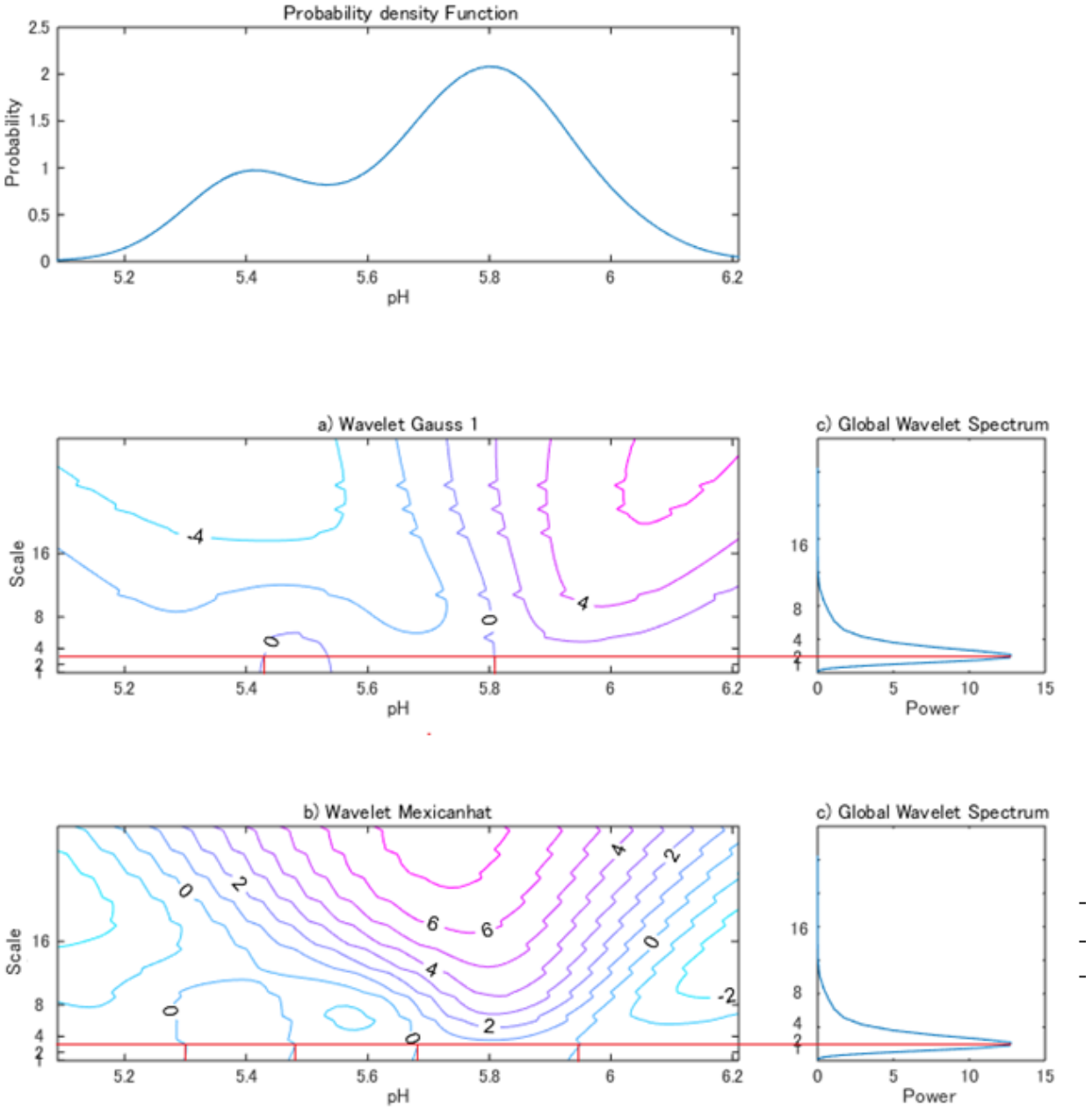


図 7.22 品質管理問題への応用で用いたデータの解析

上段が，V.D.Spline 関数による入力信号で，中段が 1 次の Gaussian Wavelets 解析による処理と Spectrum，下段が 2 次の Gaussian Wavelets 解析による処理と Spectrum，を表している。これから，表 7.1 が導き出される。

表 7.1 品質管理問題への応用で用いたデータの解析結果

	Wavelet 解析の手法		非線形最適化手法	
	第二分布	第一分布	第二分布	第一分布
平均	5.42	5.82	5.39	5.79
標準偏差	0.1	0.15	0.081	0.12
混合率	0.25	0.75	0.25	0.75

表 7.1　から解るように,1 対 3 の混合率は変わらず,他の Parameter も多少の差はあるが大きな違いはない。Wavelet 解析の手法，非線形最適化手法の両手法による結果は同様の結果を示している。

この結果，original のデータ数は 1041，だから Kernel 関数による確率密度関数の再表現には 1041 のデータが必要になり，V.D.Spline 関数表現で，knots と nodes で 35 のデータが必要で,正規混合分布では 5 っのデータが必要となる。

しかも，品質評価は 2.71 倍にあげることができる。

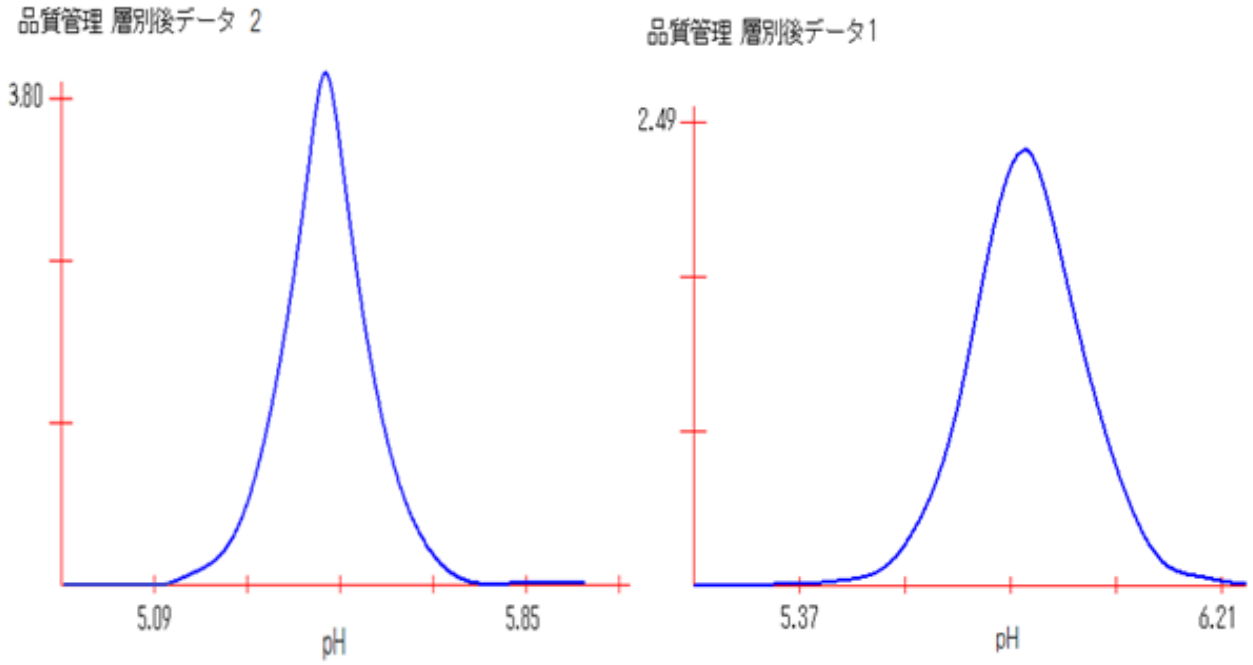


図 7.23　層別後データによる V.D.Spline 関数表現の分布

層別後データ 2 は 265 個のデータを用い，層別後データ 1 は 776 個のデータを用いた。

しかも，品質評価は層別後データ 1 では 2.623 倍，層別後データ 2 では 5.015 倍にあげることができる。　しかし，今回は平均値が 6 に近い層別後データ 1 の仕入先を用いることにした。　平均値を 6 に近づける操作ができれば，層別後データ 2 の仕入先を用いた方が品質向上に有効である。

品質を考えるうえで重要なことは，製造工程中に品質のバラツキが大きくなったり，平均値が規格の中心からズレたり，不良品を出す原因はたくさんある。そこで，この原因は機械にあるのか，作業者にあるのか，作業方法にあるのか，原材料にあるのか，原因別に分けて考えて解析する必要がある。このように，データを幾つかに分けることを層別という。

混合分布の解析は，品質管理の 7 つ道具の一つの，層別ための手法として十分に活用できる。

8. 結論

確率密度関数の推定法としての正規混合分布について考察をしてきた。この中で Histogram をはじめ Nonparametric な推定法は数多く発表されている。

これらの方法は，Band 幅が一定であり，データの変動に基ずき，所により Band 幅を可変にすることによってより柔軟にデータへの適応が可能になるものと思われる。これに答えるべき方法として５章で述べたVariation Diminishing Spline 関数表現による密度関数の推定を提案した。

ここでは，knots と nodes を用いた B-Spline 関数による確率密度関数の推定を行った。これは，Kernel 関数を B-Spline 関数とした確率密度関数の推定を行っていることである。それに伴って B-Spline 関数に必要に knots と nodes をどう決めるべきかを述べた。(5.27), (5.28), (5.29)により knots を決めることは(変動が少なければ knots の間隔が長くなる。) Kernel 関数による確率密度関数の推定においてデータの変動の大きさにより Band 幅が可変になることと同一である。

従って，より表現の自由度が増していることを示し，更にその多重度を決めるということは図 5.2 の V.D.Spline 関数のバリエーションに見られる様にデータに対する適応が敏感になり，データの変動への適応度を高めている。

それにより nodes を決め確率密度関数の推定することによりデータの変動の激しいところには敏感に,激しくないところには穏やかに適応する推定方法ができた。

信号処理でいえば，低い周波数では窓幅が広く,高い周波数では窓幅が狭くして解析することと同様である。

しかも，各関数の数が通常の Kernel 関数を用いた確率密度関数の推定よりも少なく再計算の際の効率化が図れている。

通常の Kernel 関数を用いた確率密度関数の推定では確率密度関数を表現するときには全てのデータが必要だけれど，B-Spline 関数による表現は knots と nodes だけで表現できるから，保存データ数も少なくて済む。

通常，確率密度関数の推定では確率密度関数を表現するときには，数十の単位のデータで

行うことはく，数百，またはそれ以上のデータを用いる。したがって，Kernel 関数を用いた確率密度関数の推定による確率密度関数の再表現ではそれだけのデータを用いなければならない。V.D.Spline 関数による確率密度関数の推定では，多くて数十の単位で表現できる。さらに，Semiparametric な混合分布モデルでは，更に少なく十前後のデータ数で表現が可能になる。

表 8.1　確率密度関数の再現に必要なデータ数

推定方法	データ数
Kernel 関数による確率密度関数の推定方法	採取した標本すべて(通常)数百以上
提案する V.D.Spline 関数による確率密度関数の推定方法	数十の単位
混合分布としての確率密度関数の推定方法	要素分布の数Ｘ3-1

R.A.Fisher のいう統計学の問題における，有用な情報を比較的少数の数値で表すという　Ⅲ データの簡約方法に関する研究　への貢献が可能になった。

有限の混合分布モデルに関する最初の主な解析はWelden(1892 と 1893)によって提供したデータに，2 つの正規分布の確率密度関数の混合分布の適合を Pearson(1894)によって試みられた論文である。

Pearson によって分析されたデータ集合は，ナポリ湾からサンプリングされた n = 1000 のカニの体長に対する額の比率上の測定から成った。

これらの測定は,それらおよびその 2 つの要素に適合された正規混合分布の密度に表示される。Welden は，これらのデータの Histogram 中の不調和がこの母集団が 2 つの新しい亜種の方へ発展させていた信号かもしれないと推測した。

Pearson は，優れた適合を得るために彼が開発した Moment 法を使用し，カニの 2 つの種があったという証拠として 2 つの要素の存在を解釈した。しかしながら，Moment 法は要素数が増えると高次のモーメントを使用しなければならなくなり,計算的にはあまり勧められるものではない。

混合分布の解析において，特によく用いられている方法として E-M Algorithm がある。他に，6 章で提案した，非線形最小二乗法を用いる方法がある。これらのいずれの方法も，

Parameter を計算するための方程式を立てるために要素数を予め設定していた。

本論文で提案する確率密度関数の V.D.Spline 関数による推定方法は入力信号として信号解析で用いられる。また，信号解析としては Wavelts 解析を用いた。この方法は，構成要素数の情報なしに解析を行うことができる。Wavelet 関数は Parameter a によって伸縮され，Parameter b によって平行移動される。そこで，正規分布の確率密度関数では平均値は最大値をあたえる点であり，標準偏差は変曲点の位置と関係するので，Gaussian Wavelets の１階・２階導関数である Gaussian １次・Gaussian 2 次(Mexican hat) 関数の０点を探索することにした。

しかし，Parameter a によって 0 の等高線も曲がるので Parameter b の位置も変化する。そこで．最適な Parameter b を見つけるために。そのため，a の探索レベルを決定するための尺度として Wavelets Power Spector を用いる。信号の与えるスペクトルが最大になる Scale・Parameter を選び，その位置での等高線が 0 の値が与えられる Translate 値 b を平均(Gaussian １次の時)・標準偏差(Gaussian 2 次の時)を示す点とする。

しかし．Wavelets Power Spector が最大点であることが妥当かというと，必ずしもそうではなさそうである。ほかにも何か基準となるものを探す必要がある。

統計では，混合分布モデルは，測定値セットが個々の観測データが属する部分母集団を識別することを要求せずに，母集団内の部分母集団の存在を表わすための確率モデルである。

形式的に，混合分布モデルは，母集団の中で観測データの確率分布を表わすことに相当する。

しかしながら，「混合分布」に関連した問題は部分母集団の中から母集団の特性を引き出すことに関係があるが，「混合分布モデル」は同時に母集団についての観測データだけを与えられた部分母集団の特性に関する統計的推測を作るために部分母集団の同一性情報なしで使用される。

混合分布モデルを観測データに適合させるいくつかの方法は，仮定された部分母集団の同一性が個々の観測データ(あるいはそのような部分母集団への軽重)に起因すると考えるステップを含んでいる。

これらのことを踏まえてさらに，適切な手法とその基準を求めてこれからの研究を続けてゆきたい。

謝辞

本研究をまとめるにあたり，一方ならぬ励ましとご指導をいただいた井田憲一教授に深く感謝を申し上げます。

足利工業大学　山城光雄教授には貴重なアドバイスをいただき感謝しております。

情報処理における論理と数値処理の倫理を教えていただいた故山内二郎先生，データ解析の見識を伝えていただいた故田口玄一先生，数学的素養を身に着けていただいた馬渡鎮夫先生に厚くお礼を申し上げます。

また，主査の鍾先生をはじめ，岡野先生，松本先生には細部にわたり丁寧なご指摘と論文の完成にご指導いただき，誠にありがとうございました。

また，論文執筆までの筋道を整えていただいた，足利工業大学 故横田孝雄教授に論文のまとめの報告が出来なかったのが非常に残念でなりません。心からお礼を申し上げます。

平成２７年　３月

参考文献

[1] R.A.Fisher:“Statistical Methods for Research Workers”, Oliver & Boyd Ltd., Publishers, (1963)

[2] J.W.Tukey :“Exploratory Data Analysis”, Addison-Wesley Publishing Company, (1977)

[3] D.Freedman and P.Diaconis :“On this histogram as a density estimator: L2 theory”,Zeit. Wahr. ver. Geb.,57,pp.453-476,(1981)

[4] D.W.Scott :“On Optimal and Data-Based Histograms”, Biometrika Vol. 66, No.3 Dec., pp.605-610,(1979)

[5] H. A.Sturges :“The Choice of a Class Interval”,Journal of the American Statistical Association, Vol. 21, No. 153, pp.65-66,(1926)

[6] M.Rosenblatt :“Remarks on some nonparametric estimates of a density function”,Ann.Math.Statist. 27, pp.832-837,(1956)

[7] M.Rosenblatt :“Curve estimates”, Ann.Math.Statist.42,pp.1815-1842,(1971)

[8] E.Parzen :“On estimation of a probability density function and mode”, Ann.Math.Statist. 33 pp.1065-1076, (1962)

[9] K.Pearson:“Contributions to the Mathematical Theory of Evolution”, Phil. Trans. R. Soc. Lond. A 185, pp.71-110, (1894)

[10] S.Newcomb:“A Generalized Theory of the Combination of Observations so as to Obtain the Best Result”, American Journal of Mathematics Vol.8, No.4 Aug. , pp.343-366, (1886)

[11] L.Devroye :“A Course in Density Estimation”,Birkhauser Boston,(1987)

[12] L.Devroye and L.Gyorfi :“Nonparametric Density Estimation: The L1 View”, John Wiley, (1985)

[13] R.A.Tapia and J.R.Thompson :“Nonparametric Probability Density Estimation”, Johns Hopkins Univ.Press, (1978)

[14] D.P.Doane : “Aesthetic frequency classification. American Statistician” , Vol.30, pp.181-183,(1976)

[15] B.W.Silverman : “Density Estimation for Statistics and Data Analysis” , Chapman and Hall,(1986)

[16] W.F.R.Welden : “On Certain Correlated Variations in Carcinus moenas” , Proceedings of the Royal Society, Vol.54 pp.318-329, (1893)

[17] B.S.Everitt and D.J.Hand : “Finite Mixture Distribution” , Chapman and Hall,(1981)

[18] C.G.Bhattacharya : “A SIMPLE METHOD OF RESOLUTION OF A DISTRIBUTION INTO GAUSSIAN COMPONENTS”,Biometrics, pp.115-135,(1967)

[19] G.D.Murray and D. M.Tttterington : “Estimation Problems with Data from a Mixture”, Appl.Statist, Vol.27, No.3, pp.325-334,(1978)

[20] E.A.C.Thomas : “Distribution free tests for mixed probability distributions”, Biometrika Vol.56,No.3,pp.475-484,(1969)

[21] D.M.Titterington, A.F.M.Smith and U.E.Markov: “Statistical Analysis of Finite Mixture Distributions”, John Wiley and Sons,(1985)

[22] A.P.Dempster, N.M.Laird and D.B.Rubin : “Maximum Likelihood from Incomplete Data via the EM Algorithm”, Journal of the Royal Statistical Society. Series B (Methodological) 39 (1),pp 1-38,(1977)

[23] T.N.E.Greville ed.: “Theory and Application of Spline Functions” , Acacemic Press, (1969)

[24] J.H.Ahlberg, E.N.Nilson and J.L.Walsh: “The theory of Splines and their Applications”, Academic Press,(1967)

[25] I.J.Schoenberg: “Contribution to the problem to approximation equidistatant data by analitic functions”, Quart. Appl. Math. Vol.4, (1946)

[26] I.J. Schoenberg : “On Variation Diminishing approximation methods” , On Numerical Approximation Methods,(1959)

[27] I.J.Schoenberg : “On Variation Diminishing Spline approximation methods” , Mathematica Vol.8,(31), (1966)

[28] T.L.Boullion and P.L.Odell:“Generalized Inverse Matrices”,
Willy-Inter Science, (1971)

[29] L.I.Boneva, D.Kendall and I.Stefanov :“Spline Transformation Tree New Diagonostic Aids for Statistical Data Analysis”,
J. Roy. Statisist. Soc. Ser, B 33 . (1971)

[30] I.J. Schoenberg :“Spline and Histogram”,ISNM Vol.21,pp.227-327,(1973)

[31] M. Marsden and I.J.Schoenberg:“On Variation Diminishing Spline Approximation Method”, Mathematics(cluj), Vol. 8,pp.61-82,(1966)

[32] M.Marsden:“An Identity for Spline Function with Application to Variation Diminishing Spline Approximation”, J.Approximation Theory
Vol.3, pp.7-49, (1970)

[33]塚越 清:“Variation Diminishing Spline 関数の knot の決定法について”
, 情報処理学会誌, Vol.18 No.6,(1977)

[34]塚越 清:“Variation Diminishing Spline 関数の knot の配置とその多重度の決定について″,情報処理学会誌,Vol.19 No.3,(1978)

[35]塚越 清:“Variation Diminishing Spline 関数による確率密度関数の効率的表現”,日本経営工学会誌,Vol.29,No.3,(1979)

[36]塚越 清:“確率密度関数の Variation Diminishing Spline 関数表現とその特性関数”,日本品質管理学会誌, Vol.10,No.2,(1980)

[37] H.B.Curry and I.J.Schoenberg:“On polya Frequency Function Ⅳ. Fundamental Spline Function and Their limit” ,
J.d'Anal.Math.Vol.17, pp.71-107,(1966)

[38]塚越 清, 武田 龍平:“混合分布の要素分布の推定(その I L2-norm)” ,
足利工業大学研究集録, Vol.40,pp.83-88,(2006)

[39]向島 達　他:“臨床検査細菌検査結果のコンピュータによる解析” ,
メディヤサークル,Vol.24, No.8,(1979)

[40]農林水産省ホームページ
http://www.maff.go.jp/j/keiei/hoken/saigai_hosyo/s_yoko/pdf/2-3_3.pdf

[41]田口玄一:″開発・設計段階の品質工学″,品質工学講座 1,日本規格協会,(1989)

[42]R. T. Ogden :“Essential Wavelets for Statistical Applications and Data Analysis”, Birkhauset,(1997)

[43]P.S.Addison:“The Illustrated Wavelet Transform Handbook”, Taylor&Francis,(2002)

[44]K.Tsukagoshi and S.Mawatari:“Extraction of Element Distribution of Gauss Mixture Distributions with Unknown Number of Elements”, Intelligent Engineering Systems Through Artificial Neural Networks, Vol.19, (2009)

[45]K.Tsukagoshi and K.Ida:“Extraction of Element Distribution of Gauss Mixture Distributions by Wavelet Power Spectrum”, Intelligent Engineering Systems Through Artificial Neural Networks, vol.20, (2010)

[46]C.Torrence and P. Compo :“A practical guide to Wavelet Analysis”, Bulletin of the American Meteorological Society, (1998)

[47] 塚越清,井田憲一:“ガウス系ウェーブレットによるGMMの解析”, 日本設備管理学会誌,Vol.23,No.3,(2011)

[48]K.Tsukakoshi and K.Ida:“Analysis of GMM by a Gaussian Wavelet transform” Procedia Computer Sciences Conference on System Engineering Research, Vol.8 (2012)

[49]K.Tsukagoshi and K.Ida:“The GMM Problem as one of The Estimation Methods of a Probability Density Function”,LNCS08210 Active Media Technology, (2013)

[50]K.Tsukagoshi and K.Ida:“The solution According Finite Mixture Distribution by GMM Problem as One of the Modes of Epression of Probability Density Functiom”, Procedia Computer Sciences Complex Adaptive Systems, Vol.20,(2013)

発表文献一覧

本論文関係筆頭論文

[1]塚越　清 ：“Variation Diminishing Spline 関数の knot の決定法について” ,
情報処理学会誌 Vol.18 No.6,pp.550-557,(1978 3)

[2] 塚越　清 ：“Variation Diminishing Spline 関数による確率密度関数の効率的表現”
日本経営工学会誌 Vol.29,No.3,pp.315-321,(1979　3)

[3] 塚越　清 ：“Variation Diminishing Spline 関数の knot の配置とその
多重度の決定について”,情報処理学会誌 Vol.19 No.3,pp.256-262,(1979 3)

[4] 塚越　清 ；“確率密度関数の Variation Diminishing Spline 関数表現と
その特性関数″,日本品質管理学会誌, Vol.10,No.2,pp.42-50,(1980　4)

[5] K.Tsukagoshi, S.Mawatari:“Extraction of Element Distribution of Gauss
Mixture Distributions with Unknown Number of Elements”,
Intelligent Engineering Systems Through Artificial
Neural Networks, vol.19 pp.587-594 ,(2009　11)

[6] K.Tsukagoshi, K.Ida:”Extraction of Element Distribution of Gauss Mixture
Distributions　by Wavelet Power Spectrum”,Intelligent Engineering
Systems Through Artificial Neural Networks, vol.20,　pp.461-468,(2010 11)

[7] 塚越　清　井田憲一:“ガウス系ウェーブレットによる GMM の解析”
日本設備管理学会誌 Vol.23 No.3,pp.139-144,(2011 11)

[8] K.Tsukagoshi, K.Ida: “Analysis of GMM by a Gaussian Wavelet transform”
Procedia Computer Sciences Conference on　System Engineering Research
Vol.8,pp.467-472,(2012　3)

[9] K.Tsukagoshi, K.Ida and T.Yokota:“The GMM Problem as one of The Estimation
Methods of a Probability Density Function”,
LNCS08210 Active Media Technology pp.358-368,(2013 10)

[10] K.Tsukagoshi,K.Ida:“The solution According Finite Mixture Distribution by GMM Problem as One of the Modes of Expression of Probability Density Functiom”, Procedia Computer Sciences Complex Adaptive Systems Vol.20, pp.421-426,(2013 11)

[11] K.Tsukagoshi,K.Ida:“Research on the Gaussian Mixture distribution analysis as estimation of Probability Density Function and it's the periphery”, SCSE 2015 (in press)

２０１５年　３月　現在

その他発表論文(査読付き)

[1] K. Tsukagoshi :“Variation Diminishing Spline Function repesentation of density function and it's characteristic function”, American statistical Assocoation Proceedings of the Statistical Computing,(1987 8)

[2]塚越　清　:“多群判別のＭＴＳ法における分散共分散行列の扱いにについて”, 品質工学会誌 Vol.11,No.6,pp.58-63,(2003 12)

[3] K. Tsukagoshi :“To Estimate of Inverse Matrix of Variance-Covariance Matrix in the MTS Method of Multi-Group Discriminant problem”, The 33rd International Conference on Computer and Industrial Engineering Proceedings, (2004 3)

[4]T. Yokota, K.Tsukagoshi, S.Wada. etc :“GA-Base Method for Optimal Weight Design Problem of 23 Stories Frame Structure”, LNCS08210 Active Media Technology, pp.421-426, (2013 10)

その他発表論文(査読なし)

[1]塚越　清: “Fuzzy 空間についての考察”,青山経営論集, Vol.8,No.3,(1974 1)

[2]塚越　清: “探索的データ解析システムの開発”
足利工業大学研究集録, Vol.15,(1983),

[3]塚越　清: “図形データのマッチング問題について”,
足利工業大学研究集録, Vol.17,(1985)

[4]池田・塚越・佐藤:“教育のための CAI システム”,
足利工業大学研究集録, Vol.19, (1986)

[5]塚越・佐藤 : “データ解析パッケージ言語の構築についての一考察”,
足利工業大学研究集録, Vol.19,(1986)

[6]塚越　清;“マハラノビスの距離による手書き文字のパターン認識”,
標準化と品質管理, Vol.45, No.7, (1992 7)

[7]塚越　清 :″群馬県各市町村の民力度と最高価格地の関係について”,
足利工業大学研究集録, Vol.28, (1995 3)

[8]塚越　清:“北関東三県の民力度からみた地価解析（３．群馬県編）”,
足利工業大学研究集録, Vol.29 , (1996 3)

[9]塚越　清:“北関東三県の民力度からみた地価解析（２．栃木県編）”,
足利工業大学研究集録, Vol.29, (1996 3)

[10]塚越　清:“北関東三県の民力度からみた地価解析（１．茨城県編）”,
足利工業大学研究集録, vol.29, (1996 3)

[11]塚越　清 :“マハラノビスの距離による手書き文字のパターン認識
（微分情報と積分情報による）”,足利工業大学研究集録, Vol.30 (1997 3)

[12]塚越　清: “民力度を用いた群馬県各市町村の最高価格地の多変量解析
による考察”,固定資産評価研究大会報告書 Vol.2,pp.107-118, (2001 3)

[13]塚越　清: “データの精度と誤差の評価”,教育研究会理化学部会会誌 ,
Vol.38,群馬県高等学校 (2002 11)

[14]塚越　清: “マハラノビスの距離の計算における多重共線性の問題について”,

足利工業大学研究集録, Vol.35, (2002 3)

[15] 塚越・武田:“混合分布の要素分布の推定(そのⅠ 12-norm)”,

足利工業大学研究集録, Vol.40, (2006 3)

[16] 高橋・塚越:“車の安全性要因に関する統計手法による解析”,

足利工業大学研究集録, Vol.45, (2012 3)

[17] 高橋・塚越・中野: “太田市の小中学校のインフルエンザ罹患データの解析”,

足利工業大学研究集録, Vol.46,(2013 3)

付録

データ

本論文で用いたデータの一部を掲載する。

図 3.1 ,図 3.4 ,実験 5.3 に用いたデータ

500

58.11 48.95 51.35 50.31 40.84 57.93 46.58 46.80 48.58 46.92 41.82 38.28 46.31 50.90
42.05 44.77 54.05 44.88 52.29 56.25 46.78 38.87 52.52 47.74 44.51 57.85 57.76 54.22
57.25 51.84 52.99 55.71 49.98 45.82 38.22 57.19 52.72 44.81 53.46 53.67 60.45 58.79
48.69 55.16 53.19 47.78 53.93 36.64 55.92 56.76 54.16 58.13 53.65 50.74 49.40 54.61
51.39 54.73 49.63 51.10 59.12 43.71 49.87 47.58 56.86 52.70 45.10 59.06 49.59 56.68
45.33 50.55 57.33 45.67 45.57 47.03 45.06 59.65 50.80 48.52 42.79 48.63 41.04 55.00
55.53 52.62 51.27 51.49 48.26 51.60 46.51 42.97 51.00 50.59 51.74 54.46 58.73 54.57
51.97 45.94 46.47 38.56 47.21 47.42 39.20 47.54 57.44 48.91 41.94 46.53 47.68 45.39
54.67 50.51 47.91 46.88 47.40 44.49 43.15 48.36 55.14 48.48 48.38 49.85 42.87 52.46
58.62 51.33 45.61 51.45 58.85 42.81 48.34 50.43 44.08 44.30 46.08 39.42 49.32 50.78
48.81 53.40 54.55 47.27 56.55 52.38 39.79 53.75 44.28 56.37 55.02 60.24 52.01 50.35
55.26 41.72 49.75 54.34 45.49 53.21 47.48 53.32 50.72 49.69 60.22 52.31 55.96 61.17
42.95 51.29 46.19 52.66 45.69 45.28 51.43 54.14 53.42 54.26 56.66 50.63 46.15 53.24
51.90 52.11 43.89 52.23 52.13 48.60 56.62 51.21 52.37 55.08 54.36 50.20 42.60 61.56
52.09 49.18 52.83 43.05 49.83 33.17 53.07 44.53 52.56 52.15 53.30 51.02 55.30 56.14
53.54 57.50 43.03 50.12 48.77 48.99 45.76 44.10 54.01 45.47 53.50 58.09 59.24 46.96
51.23 52.07 49.47 53.44 48.97 46.06 54.71 44.92 51.70 45.04 44.94 46.41 44.44 49.03
50.18 52.89 57.17 43.01 50.41 54.38 49.90 51.99 50.65 50.86 57.64 50.98 45.88 52.35
55.37 44.96 41.12 48.83 48.11 43.95 46.35 45.31 50.84 42.93 51.58 51.80 48.58 51.92

51.82 48.28 51.31 50.90 57.05 49.77 44.04 44.88 52.29 51.25 51.78 53.87 42.52 57.74
49.51 52.85 52.76 49.22 42.25 51.84 57.99 50.71 54.98 40.82 58.22 52.19 62.72 39.81
53.46 43.67 50.45 48.79 48.69 45.16 58.19 42.78 48.93 46.64 50.92 46.76 44.16 43.13
53.65 50.74 49.40 54.61 36.39 54.73 54.63 46.10 59.12 58.71 44.87 42.58 51.86 47.70
50.10 44.06 49.59 56.68 45.33 45.55 52.33 50.67 50.57 37.03 55.06 49.65 50.80 43.52
47.79 48.63 51.04 50.00 50.53 42.62 51.27 51.49 48.26 46.60 46.51 52.97 56.00 45.59
51.74 59.46 48.73 54.57 51.97 55.94 56.47 53.56 47.21 47.42 49.20 42.54 52.44 48.91
56.94 56.53 37.68 50.39 44.67 55.51 47.91 51.88 47.40 44.49 43.15 48.36 45.14 43.48
53.38 39.85 42.87 47.46 43.62 46.33 50.61 46.45 53.85 47.81 58.34 55.43 44.08 49.30
41.08 49.42 44.32 55.78 43.81 53.40 49.55 47.27 51.55 42.38 54.79 53.75 49.28 51.37
45.02 55.24 47.01 55.35 45.26 51.72 49.75 49.34 45.49 48.21 62.48 53.32 45.72 44.69
50.22 52.31 50.96 46.17 57.95 51.29 41.19 47.66 50.69 60.28 56.43 59.14 48.42 44.26
41.66 50.63 46.15 48.24 46.90 57.11 53.89 52.23 62.13 43.60 51.62 56.21 52.37 45.08
44.36 50.20 47.60 46.56 52.09 39.18 42.83 48.05 49.83 63.17 48.07 44.53 57.56 52.15
48.30 46.02 50.29 51.13 48.54 42.50 43.03 50.12 48.77 38.99 55.76 59.10 54.01 50.47
53.50 48.09 49.24 51.96 51.23 57.07 49.47 48.44 48.97 46.06 49.71 49.92 51.70 50.04
49.94 51.41 44.44 49.03 50.18 52.89 52.17 58.01 55.41 54.38

Silvermandata1

1 25 40 83 123 256 1 27 49 84 126 257 1 27 49 84 129 311 5 30 54 84 134 314
7 30 56 90 144 322 8 31 56 91 147 369 8 31 62 92 153 415 13 32 63 93 163 573
14 34 65 93 167 609 14 35 65 103 175 640 17 36 67 103 228 737 18 37 75 111 231 21 38
76 112 235 21 39 79 119 242 22 39 82 122 256

Silvermandata2

4.37 3.87 4.00 4.03 3.50 4.08 2.25 4.70 1.73 4.93 1.73 4.62 3.43 4.25
1.68 3.92 3.68 3.10 4.03 1.77 4.08 1.75 3.20 1.85 4.62 1.97 4.50 3.92
4.35 2.33 3.83 1.88 4.60 1.80 4.73 1.77 4.57 1.85 3.52 4.00 3.70 3.72
4.25 3.58 3.80 3.77 3.75 2.50 4.50 4.10 3.70 3.80 3.43 4.00 2.27 4.40

4.05 4.25 3.33 2.00 4.33 2.93 4.58 1.90 3.58 3.73 3.73 1.82 4.63 3.50
4.00 3.67 1.67 4.60 1.67 4.00 1.80 4.42 1.90 4.63 2.93 3.50 1.97 4.28
1.83 4.13 1.83 4.65 4.20 3.93 4.33 1.83 4.53 2.03 4.18 4.43 4.07 4.13
3.95 4.10 2.72 4.58 1.90 4.50 1.95 4.83 4.12

CBPC (図 6.2)

8 8 23 8 8 8 8 8 8 8 15 8 8 8 25 8
25 8 8 8 8 8 8 26 8 8 8 8 8 8 8 8
8 8 23 8 30 8 23 8 8 8 8 8 24 8 8 8
8 26 8 8 25 8 8 8 26 8 8 8 8 24 8 8
8 8 30 8 8 16 19 19 8 8 8 8 8 8 8 8
8 25 22 24 21 23 27 8 8 8 24 25 25 21 8 27
8 8 11 8 8 8 25 8 25 24 24 25 20 25 24 8
8 23 25 8 8 22 25 8 8 8 8 18 8 8 21 20
20 24 23 8 8 25 8 26 22 08 22 24 08 08 08 24
24 08 08 08 08 24 08 08 08 08 08 22 24 25 23 08
08 23 21 08 26 25 08 08 27 27 26 08 08 08 28 22
08 08 08 23 08 08 08 08 26 08 08 08 08 08 08 26
08 08 23 27 20 27 08 08 08 08 08 08 19 24 24 23
21 18 08 08 08 28 24 27 25 08 23 23 26 08 08 08
16 18 27 13 29 24 08 08 08 08 08 27 24 08 27 25
08 26 08 08 22 25 30 08 19 21 23 20 22 13 08 08
08 08 08 08 25 25 08 08 25 22 26 08 27 29 08 08
08 24 28 24 23 08 08 08 08 25 17 26 20 08 24 08
08 08 08 08 22 31 08 08 08 08 08 29 26 08 26 08
08 24 08 23 08 11 28 24 25 27 27 22 26 10 24 08
08 22 22 08 27 08 08 08 08 26 08 08 08 08 23 24
26 11 16 08 08 08 08 08 23 24 29 26 08 08 08 29
08 08 30 23 08 08 08 27 27 30 08 25 27 08 22 29

08 08 08 08 08 27 08 24 24 08 08 08 08 29 28 32
27 08 08 30 08 29 08 27 08 08 23 20 23 28 08 21
21 19 29 20 22 26 18 08 08 08 08 22 22 23 08 08
23 20 23 20 23 24 08 08 08 08 08 26 08 08 34 08
08 25 21 20 24 20 22 08 08 08 22 23 08 08 08 23
08 08 23 08 23 08 26 19 31 21 24 21 23 22 08 08
08 24 08 08 08 22 24 22 24 22 22 22 25 27 08 25
08 08 08 08 08 11 08 25 08 08 31 08 08 08 22 08
08 08 25 08 27 08 25 08 22 25 21 25 08 08 08 08
08 34 08 16 22 08 08 30 08 27 26 26 27 08 27 08
17 25 22 24 24 08 08 08 08 08 08 08 08 08 26 08
21 24 21 22 20 08 08 08

SBPC （図 6.3）

08 08 24 08 08 08 08 08 08 08 16 08 08 08 25 08
22 08 08 08 08 08 08 26 08 08 08 08 08 08 08 08
08 27 22 08 32 08 24 08 08 08 08 08 24 08 08 08
08 26 08 08 24 08 08 08 29 08 08 08 08 24 08 08
08 08 31 08 08 16 17 18 08 08 08 08 08 08 08 08
08 23 24 23 22 26 27 08 08 08 24 25 24 22 08 26
08 08 08 08 08 08 24 08 24 24 25 25 21 25 22 08
08 24 23 08 08 22 20 08 08 08 08 18 08 08 21 20
20 23 23 08 08 26 08 26 23 08 23 21 08 08 08 23
22 08 08 08 08 23 15 08 08 08 08 23 24 25 23 08
08 23 22 08 27 26 08 08 27 25 25 08 08 08 29 24
08 08 08 27 08 08 08 08 08 08 08 08 08 08 08 27
08 08 25 29 26 29 08 08 08 08 08 08 21 24 18 25
22 20 08 08 08 29 23 27 23 08 26 27 28 08 08 08
15 22 33 15 31 25 08 08 08 08 08 26 25 08 28 24
08 27 08 08 23 25 31 08 20 24 24 22 22 08 08 08

08 08 08 08 26 26 08 08 27 22 25 08 29 29 08 08
08 24 29 25 24 08 08 08 08 23 18 29 22 08 27 08
08 08 08 08 25 32 08 08 08 08 08 28 28 08 28 08
08 25 08 25 08 08 26 24 24 27 31 23 30 08 24 08
08 22 23 08 28 08 08 08 08 28 08 08 08 08 23 24
25 13 20 08 08 08 08 08 25 25 31 28 08 08 08 30
08 08 08 25 08 08 08 30 27 32 15 29 26 08 24 32
08 08 08 08 08 27 08 23 23 08 08 08 08 30 29 31
28 08 08 30 08 29 08 24 08 08 25 22 24 30 08 26
25 22 30 21 26 24 16 08 08 08 08 24 21 22 16 08
23 24 25 23 24 26 08 08 08 08 08 28 08 08 26 08
08 28 22 23 24 23 26 08 08 08 24 21 08 08 08 26
08 08 24 08 23 08 21 12 28 25 22 25 25 24 08 08
08 20 08 08 08 21 21 20 22 22 18 22 23 30 08 25
08 08 08 08 08 08 08 23 08 08 32 08 08 08 24 08
08 08 19 08 26 08 23 08 20 23 21 22 08 08 08 08
08 32 08 10 17 08 08 25 08 22 23 28 25 08 27 08
23 23 22 22 23 08 08 08 08 08 08 08 08 08 30 08
22 24 20 23 20 08 08 08

SM（図 3.6, 図 6.4）

26 16 22 08 11 08 08 08 12 08 22 08 08 12 24 08
22 08 16 08 08 08 08 11 08 08 08 28 08 08 08 08
08 25 08 08 23 12 13 23 14 13 14 12 22 08 08 08
23 23 16 13 22 08 11 08 23 08 08 16 14 22 08 08
08 14 25 29 08 18 22 21 11 13 08 08 08 24 08 18
08 13 21 22 21 11 11 08 08 08 20 22 22 20 11 13
08 08 27 08 08 08 25 23 25 11 15 13 14 13 08 08
08 08 13 08 08 16 13 08 08 08 08 23 08 08 12 13
08 22 20 11 24 24 11 08 24 08 08 08 08 08 12 14

29 11 08 08 08 20 08 08 24 15 15 15 22 24 23 17
08 15 20 08 22 22 17 08 24 24 22 08 24 23 24 23
14 08 08 27 08 08 08 08 13 08 08 08 08 08 08 22
23 16 22 23 23 22 15 08 13 08 13 08 22 22 25 22
21 18 08 08 11 24 18 20 21 13 22 22 25 21 08 08
24 23 30 35 11 23 08 18 08 23 23 22 27 08 24 23
22 11 08 08 13 11 25 23 17 21 24 21 22 08 08 08
23 25 08 08 11 23 08 08 11 21 24 08 11 08 08 15
14 24 20 26 25 13 08 08 08 21 08 08 08 28 12 24
23 11 08 08 23 08 08 13 08 27 26 26 23 11 13 25
13 23 25 21 08 23 27 24 24 08 08 08 25 23 22 08
08 24 26 08 24 08 13 08 08 25 11 25 08 08 08 22
23 19 19 08 08 11 08 08 08 22 25 23 10 10 08 26
13 13 26 25 08 13 08 23 08 20 15 25 20 08 21 08
08 08 11 08 08 24 13 23 23 15 08 11 08 08 08 22
24 13 12 26 08 23 12 21 08 08 21 24 24 22 08 19
20 08 20 21 20 08 23 20 08 08 08 21 26 10 21 08
08 21 18 21 21 08 08 08 08 08 08 19 08 13 23 12
08 08 21 08 13 18 11 08 14 15 21 16 08 08 08 08
12 12 08 08 21 11 10 27 24 20 20 08 23 19 08 08
08 19 15 08 08 22 08 11 20 24 19 08 19 20 08 08
17 08 08 08 08 08 08 23 18 08 25 08 08 13 21 08
12 13 22 13 21 08 21 08 20 20 21 26 11 08 08 08
20 08 08 24 08 08 08 25 08 25 12 22 18 08 23 24
23 24 24 08 22 17 08 08 21 22 08 08 08 24 23 08
23 24 08 22 21 10 08 15

TC(図 6.4)

24	15	27	12	13	16	13	13	15	21	30	08	20	26	31	12
27	28	15	33	14	08	14	32	34	20	17	34	11	12	14	20
33	31	12	15	35	12	20	30	18	18	20	31	27	26	12	15
17	14	20	13	28	12	11	14	34	12	19	21	19	29	08	13
08	19	20	32	17	27	29	29	15	19	13	11	12	13	15	15
15	15	29	27	26	15	15	15	15	15	23	25	22	28	14	12
11	31	24	17	12	10	29	31	28	18	18	15	16	20	15	16
16	17	16	12	12	16	16	12	14	16	08	25	08	11	15	16
27	24	23	17	29	31	13	34	29	17	11	08	14	26	08	08
33	12	08	17	17	25	17	15	16	29	26	10	31	30	30	33
16	14	30	14	30	30	36	11	27	28	29	13	29	13	29	30
12	16	17	34	12	14	16	20	30	10	12	13	13	12	19	29
35	26	31	28	27	31	24	12	23	12	28	12	30	29	30	30
28	20	27	15	12	31	30	32	31	19	32	35	30	13	12	12
29	30	37	32	31	28	13	31	11	11	11	31	27	13	20	30
12	31	14	12	12	13	31	11	17	31	30	28	12	28	11	08
30	11	29	18	16	13	13	08	14	12	30	12	20	32	13	13
12	31	28	28	27	14	27	12	11	27	08	36	11	29	27	27
32	15	13	13	35	37	13	28	17	26	30	30	31	28	12	35
13	29	11	28	29	31	28	25	14	26	28	16	31	27	28	20
08	27	29	14	30	12	35	14	33	30	12	11	18	14	30	28
28	34	36	12	08	13	11	11	32	32	32	30	14	12	12	36
12	08	08	31	11	31	17	30	33	34	28	35	34	11	29	38
17	18	20	15	19	29	11	27	26	20	17	11	17	29	30	31
31	16	17	31	19	31	18	18	14	14	29	28	28	28	17	27
30	17	30	30	30	12	29	15	18	17	18	29	29	19	28	16
11	15	30	34	35	28	31	31	28	13	12	29	12	18	32	14
14	32	29	30	29	29	23	27	24	32	27	24	13	18	15	30

18	19	30	27	29	20	31	30	30	31	27	15	29	28	12	13
13	14	28	26	11	32	12	16	17	15	15	14	13	30	12	32
27	11	14	15	15	21	20	14	28	08	34	15	15	16	27	08
11	14	31	13	32	30	28	14	29	30	31	28	11	34	13	15
20	32	11	27	27	16	13	30	15	30	25	31	31	11	30	21
30	28	31	30	32	15	08	31	11	12	13	16	08	27	26	30
27	30	15	28	30	10	10	28								

品質管理データ(図 3.2, 図 5.8, 図 6.5, 図 7.21)

6.01	5.4	5.92	5.92	5.93	5.74	5.84	5.39	5.86	5.67
5.9	6.01	5.84	5.71	5.96	5.74	5.81	5.39	5.95	5.61
5.85	6.05	5.88	5.74	5.9	5.82	6.07	5.85	5.8	5.67
5.93	5.91	5.7	5.76	5.85	5.82	5.92	5.83	5.81	5.65
5.93	5.91	5.75	5.68	5.88	5.82	5.84	5.82	5.87	5.58
5.87	5.87	5.78	5.74	5.62	5.84	6.1	5.81	5.84	5.59
5.85	5.89	5.84	5.89	5.66	5.81	6.11	5.73	6.08	5.58
5.86	5.87	5.88	5.8	5.72	5.86	6.04	5.74	6.03	5.54
5.84	5.73	5.93	5.68	5.75	5.78	6.07	5.68	5.93	5.55
5.88	5.67	5.91	5.97	5.76	5.87	6.07	5.77	5.91	5.61
5.97	5.68	5.9	5.83	5.78	5.89	6.05	5.69	5.75	5.57
5.74	5.7	5.81	5.8	5.8	5.91	5.99	5.72	5.81	5.56
5.87	5.64	5.8	5.85	5.74	5.87	6.02	5.68	5.94	5.5
5.85	5.66	5.82	5.71	5.82	5.52	5.87	5.63	5.9	5.51
5.89	5.76	5.95	5.8	5.71	5.71	5.87	5.72	6.13	5.5
5.81	5.81	5.97	5.71	5.63	5.43	5.84	5.73	5.98	5.59
5.87	5.78	5.59	5.7	5.66	5.33	5.88	5.68	5.97	5.6
5.81	5.8	5.6	5.82	5.7	5.39	5.87	5.73	6.04	5.61
5.88	5.8	5.6	5.65	5.88	5.39	5.84	5.71	5.97	5.63
6.21	5.73	5.56	5.73	5.81	5.42	5.84	5.78	6.07	5.71

5.82	5.76	5.71	5.75	5.92	5.83	5.75	5.86	5.93	5.69
5.84	5.76	5.62	5.79	5.82	5.79	5.65	5.69	5.9	5.83
5.86	5.71	5.55	5.81	5.84	5.82	5.74	5.78	5.89	5.75
5.85	5.7	5.64	5.8	5.86	5.43	5.8	5.72	5.88	5.74
5.69	5.84	5.6	5.7	5.88	5.32	5.83	5.61	5.89	5.7
5.82	5.8	5.63	5.79	5.92	5.81	5.78	5.74	5.87	5.72
5.83	5.88	5.63	5.76	6.03	5.78	5.75	5.72	5.87	5.77
5.74	5.86	5.6	5.62	5.92	5.71	5.69	5.62	5.91	5.78
6	5.72	5.44	5.7	5.85	5.45	5.76	5.6	5.81	5.67
5.69	5.7	5.37	5.68	5.95	5.49	5.75	5.32	5.92	5.84
5.69	5.91	5.48	5.9	5.89	5.32	5.65	5.72	5.7	5.68
5.68	5.9	5.55	5.71	6.03	5.88	5.65	5.69	5.82	5.77
5.83	5.9	5.51	5.81	6.07	5.71	5.69	5.79	5.92	5.73
5.8	5.72	5.49	5.7	5.93	5.76	5.75	5.77	5.93	5.66
6.2	5.76	5.88	5.73	5.91	5.77	5.75	5.7	5.84	5.64
5.74	5.69	5.9	5.46	5.94	5.79	5.53	5.77	5.87	5.83
5.82	5.8	5.94	5.4	5.93	5.81	5.75	5.66	5.86	5.71
5.8	5.84	5.84	5.65	5.86	5.71	5.6	5.67	5.9	5.69
5.73	5.79	5.83	5.64	5.82	5.78	5.8	5.82	5.86	5.69
5.75	5.76	5.83	5.71	5.74	5.89	5.7	5.63	5.79	5.82
5.79	5.77	5.87	5.79	5.85	5.81	5.63	5.73	5.9	5.75
5.94	5.79	5.88	5.77	5.85	5.91	5.69	5.54	5.85	5.75
5.7	5.79	5.86	5.46	5.87	5.74	5.7	5.65	5.76	5.85
5.79	5.78	5.93	5.56	5.8	5.77	5.68	5.58	5.75	5.9
5.77	5.86	5.91	5.63	5.85	5.72	5.82	5.56	5.85	5.74
5.74	5.87	5.83	5.81	5.72	5.7	5.89	5.56	5.84	5.78
5.83	5.88	5.82	5.74	5.77	5.96	5.84	5.76	5.73	5.78
5.8	5.82	5.71	5.46	5.78	5.82	5.85	5.77	5.88	5.99
5.82	5.83	5.8	5.45	5.83	5.8	5.82	5.85	5.84	5.76
5.75	5.79	5.74	5.42	5.85	5.9	5.85	5.78	5.72	5.38

5.9	5.8	5.78	5.34	5.47	5.39	5.87	5.88	5.73	5.43
5.88	5.77	5.88	5.45	5.4	5.38	5.48	5.9	5.8	5.26
5.86	5.74	5.83	5.44	5.4	5.53	5.36	5.85	5.81	5.31
5.84	5.09	5.86	5.49	5.45	5.5	5.41	5.87	5.86	5.35
5.79	5.4	5.79	5.44	5.48	5.42	5.44	5.82	5.74	5.48
5.73	5.67	5.78	5.48	5.39	5.34	5.55	5.83	5.72	5.51
5.64	5.68	5.97	5.34	5.34	5.44	5.53	5.88	5.73	5.32
5.68	5.75	5.95	5.43	5.45	5.38	5.54	5.69	5.74	5.3
5.79	5.81	5.77	5.32	5.49	5.38	5.53	5.67	5.79	5.35
5.71	5.86	5.81	5.4	5.42	5.38	5.53	5.68	5.69	5.49
5.68	5.86	5.99	5.41	5.37	5.4	5.53	5.75	5.7	5.37
5.78	5.76	5.95	5.42	5.41	5.38	5.57	5.73	5.81	5.35
5.7	5.8	5.94	5.38	5.42	5.34	5.59	5.79	5.82	5.44
5.72	5.8	5.99	5.32	5.52	5.4	5.54	5.78	5.8	5.55
5.85	5.76	5.94	5.33	5.41	5.33	5.48	5.83	5.78	5.52
5.62	5.81	5.91	5.3	5.45	5.35	5.51	5.85	5.7	5.57
5.89	5.63	6.1	5.3	5.59	5.4	5.53	5.75	5.7	5.41
5.82	5.61	6.04	5.43	5.39	5.44	5.46	5.81	5.69	5.66
5.83	5.7	6.01	5.37	5.44	5.26	5.4	5.86	5.74	5.52
5.64	5.68	5.84	5.47	5.48	5.37	5.53	5.9	5.72	5.6
5.68	5.69	6.02	5.39	5.41	5.23	5.51	5.82	5.7	5.57
5.79	5.67	6.04	5.4	5.4	5.26	5.47	5.79	5.78	5.64
5.66	5.64	6	5.33	5.41	5.31	5.48	5.76	5.8	5.57
5.71	5.63	6.06	5.26	5.39	5.27	5.45	5.8	5.81	5.58
5.71	5.63	5.91	5.41	5.52	5.18	5.43	5.81	5.87	5.6
5.8	5.71	5.92	5.42	5.43	5.33	5.38	5.8	5.72	5.64
5.71	5.7	5.91	5.5	5.44	5.3	5.5	5.75	5.94	5.6
5.67	5.65	5.85	5.47	5.41	5.35	5.46	5.79	5.92	5.59
5.71	5.71	6	5.36	5.37	5.3	5.85	5.82	5.84	5.57
5.78	5.68	5.93	5.42	5.32	5.28	5.37	5.74	6.03	5.45

5.8	5.72	5.98	5.37	5.38	5.28	5.59	5.72	5.52	5.42
5.88	5.59	6.01	5.37	5.38	5.21	5.47	5.75	5.76	5.36
5.91	5.62	5.98	5.45	5.41	5.28	5.32	5.79	5.96	5.44
5.86	5.55	6.02	5.53	5.27	5.41	5.41	5.74	5.97	5.51
6.08	5.61	6.04	5.49	5.28	5.32	5.43	5.7	5.97	5.4
5.85	5.6	5.96	5.38	5.3	5.45	5.44	5.72	6	5.31
5.84	5.56	5.99	5.42	5.37	5.26	5.54	5.8	5.89	5.25
5.93	5.51	5.95	5.39	5.37	5.3	5.43	5.82	5.88	5.45
6.03	5.52	5.97	5.3	5.32	5.42	5.41	5.74	5.79	5.39
5.91	5.55	6.07	5.35	5.34	5.32	5.47	5.72	5.8	5.44
5.85	5.6	6.04	5.25	5.23	5.35	5.48	5.82	5.43	5.6
5.96	5.68	5.94	5.31	5.32	5.39	5.36	5.84	5.51	5.44
5.95	5.72	5.98	5.36	5.8	5.45	5.45	5.83	5.88	5.49
6.02	5.63	6.17	5.25	5.95	5.4	5.54	5.89	5.44	5.51
5.76	5.67	6.03	5.32	5.93	5.35	5.38	5.88	5.39	5.35
5.93	5.61	5.99	5.39	5.97	5.35	5.45	5.96	5.9	5.42
5.83	5.59	5.88	5.34	5.89	5.36	5.4	5.98	6.05	5.37
5.94	5.64	6	5.33	5.97	5.38	5.4	5.83	5.94	5.38
5.84	5.62	6.02	5.35	5.87	5.38	5.38	5.43	5.94	5.98
5.84	5.69	6.1	5.38	5.86	5.97	5.35	5.52	5.92	5.93
6.04	5.92	5.75	6.1	5.62	5.87	5.76	5.93	6	5.95
5.93	5.91	5.76	5.89	5.69	5.94	6.08	5.93	6.18	5.98
5.73	5.94	5.71	5.87	5.7	6	6.06	5.6	6.04	6.03
6	5.96	5.72	6.01	5.71	5.71	5.92	5.75	5.92	6.04
5.9									

花粉データに関しては下記を参照

環境省花粉観測システム

http://kafun.taiki.go.jp/

環境省花粉観測システム（はなこさん）