統計的推測と幾何学的当てはめにおけるモデル選択

金谷 健一

群馬大学工学部情報工学科

平面上の点列への直線当てはめを例にとり、「統計的推測」と「幾何学的当てはめ」との本質的な違いを述べ、統計的推測が "繰り返しデータ" からの推論であるのに対して、幾何学的当てはめは "1 回の観測図形" からの推論であり、統計的推測のデータ数 ∞ の極限での「漸近的」性質が幾何学的当てはめでは誤差 0 の「無限小摂動」の極限で成立するという「双対性」を示す。そして、確率現象の発生機構を説明する「確率モデル」とデータ空間中の多様体としての「幾何学的モデル」を対比することにより「確率的モデル選択」と「幾何学的モデル選択」の相違を述べる。これに基いて「赤池の AIC」と「幾何学的 AIC」の導出や誤差の推定方式がどのように異なるかを示す。

キーワード: 統計的推測、幾何学的当てはめ、確率モデル、幾何学的モデル、幾何学的 AIC、モデル選択

Model Selection in Statistical Inference and Geometric Fitting

Kenichi Kanatani

Department of Computer Science, Gunma University, Kiryu, Gunma 376-8515 Japan

Taking line fitting to points in two dimensions as a typical example, we point out the fundamental difference between *statistical inference* and *geometric fitting*: statistical inference is to reason from *repeated data*, while geometric fitting is to reason from a *single observation*. We also point out their *duality* in the sense that the *asymptotic* properties of statistical inference in the limit of an infinite number of observations hold for geometric fitting in the limit of *infinitesimal perturbations*. It follows that noise estimation for the *stochastic model*, which describes the mechanism of random phenomena, is different from that for the *geometric model*, which describes the constraint manifold in the data space. We contrast *stochastic model selection* with *geometric model selection* and describe the difference between *Akaike's AIC* and the *geometric AIC* in their derivations.

Key words: statistical inference, geometric fitting, stochastic model, geometric model, geometric AIC, model selection

謝辞: 本研究の一部は文部省科学研究費基盤研究C (2) (No. 11680377) によった。本論文の内容に関して有益な議論を頂いたオーストラリアのモナシュ大学の David Suter 教授、英国レディング大学の Steve Maybank 教授、理化学研究所の甘利俊一博士、電気通信大学の韓太舜教授、群馬大学の関庸一助教授、(株) 朋栄の松永力氏に感謝します。

*376-8515 桐生市天神町 1–5–1 群馬大学工学部情報工学科
Tel: (0277)30-1841, Fax: (0277)30-1801, E-mail: kanatani@cs.gunma-u.ac.jp

1. 序論

筆者はコンピュータビジョンやロボット工学において誤差のある画像やセンサデータから空間の幾何学的構造を推論する問題の理論的研究を進め、最適推定の一般論を展開するとともに [11, 12, 13, 15, 19]、「幾何学的 AIC」によるモデル選択を提唱し [14, 20]、さまざまな応用例を示した [16, 17, 18, 22, 23, 26, 31, 39]。これは国内国外のコンピュータビジョン研究者の関心を集め、いろいろな形で実践的に応用されている [24, 36, 37, 38]。

しかし、統計学者の間ではこれがよく知られている (赤池の)「AIC」[1, 32] とどこが異なるのか、なぜ異なるのかが十分理解されていなかったため、これを疑問視する人達もいるようである [29]。その原因は統計学では農業、医学、政治、経済、等への応用の枠組み (「統計的推測」) が確立しているため、無意識にその枠組みに束縛されているためではないかと思われる。

そこで本論文ではまず、コンピュータビジョンやロボット工学で遭遇する多くの応用は従来の統計学で扱われた問題とは性格が異なることを指摘し、これらを「幾何学的当てはめ」として定式化する。このような問題が統計学者によってまったく扱われていないわけではないが、変則的な問題として「Neyman-Scott 問題」[30]、「測定誤差モデル」[8]、「撹乱母数」などの特異性を強調する名称が用いられるとともに、独自に理論的体系化がされている [3, 4]。

しかし、画像やセンサデータを用いるコンピュータビジョンやロボット工学では本論文でいう「幾何学的当てはめ」が最も普遍的な問題設定であり、統計学でいう「統計的推測」はどちらかというと例外的な設定である。この原因は「観測」の概念の相違に起因する。

本論文では幾何学的当てはめの最も単純な問題として平面上の点列への直線当てはめを例にとり、統計的推測との本質的な違いを述べる。そして、統計的推測が “繰り返しデータ” からの推論であるのに対して、幾何学的当てはめは “1 回の観測図形” からの推論であり、統計的推測のデータ数 ∞ の極限での漸近的性質が幾何学的当てはめでは誤差 0 の無限小摂動の極限で成立するという意味の双対性を述べる。

また確率現象の発生機構を説明する「確率モデル」とデータ空間中の多様体としての「幾何学的モデル」を対比することにより「確率的モデル選択」と「幾何学的モデル選択」の相違を述べる。これに基いて赤池の AIC と幾何学的 AIC の導出や誤差の推定方式がどのように異なるかを示す。

2. 直線の当てはめ

平面上の N 点 (x_α, y_α), $\alpha = 1, ..., N$ に直線 $Ax + By + C = 0$ を当てはめる問題を考える (図 1)。各点

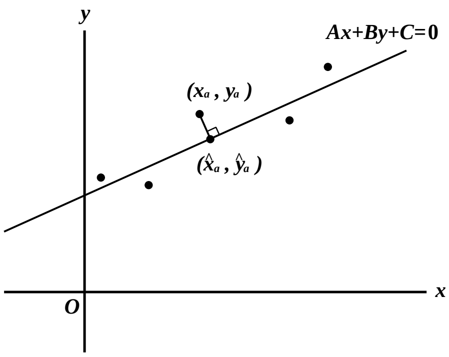

図 1: 直線の最適当てはめ。

はこの直線上の真の位置 $(\bar{x}_\alpha, \bar{y}_\alpha)$ から期待値 0、未知の分散 σ^2 の独立な正規分布に従う誤差だけずれているとする。この確率モデルのもとでの最尤推定解は次のように表される尤度を最大にするものである。

$$P = \prod_{\alpha=1}^{N} \frac{e^{-(x_\alpha - \bar{x}_\alpha)^2/2\sigma^2 - (y_\alpha - \bar{y}_\alpha)^2/2\sigma^2}}{2\pi\sigma^2} = \frac{e^{-J/2\sigma^2}}{(2\pi\sigma^2)^N} \tag{1}$$

ただし次のように置いた。

$$J = \sum_{\alpha=1}^{N} \Big((x_\alpha - \bar{x}_\alpha)^2 + (y_\alpha - \bar{y}_\alpha)^2 \Big) \tag{2}$$

未知数は $(\bar{x}_\alpha, \bar{y}_\alpha)$, A, B, C, σ^2 であり、これらの間には次の制約がある。

$$A\bar{x}_\alpha + B\bar{y}_\alpha + C = 0 \tag{3}$$

式 (1) を最大にするにはその対数尤度の符号を換えた

$$-\log P = \frac{J}{2\sigma^2} + N \log 2\pi\sigma^2 \tag{4}$$

を制約条件 (3) のもとで最小化すればよい。σ^2 と残りの変数は別々に最小化できるので、まず σ^2 で微分して 0 と置くと次のようになる。

$$-\frac{J}{2\sigma^4} + \frac{N}{\sigma^2} = 0 \tag{5}$$

これから次の最尤推定解を得る。

$$\hat{\sigma}^2 = \frac{\hat{J}}{2N} \tag{6}$$

ただし J の最小値 (残差) を $\hat{J}$ と置いた。これを求めるために制約条件 (3) に対するラグランジュ乗数 λ_α を導入し、$J/2 + \sum_\alpha \lambda_\alpha (A\bar{x}_\alpha + B\bar{y}_\alpha + C)$ を $\bar{x}_\alpha$, $\bar{y}_\alpha$ でそれぞれ微分して 0 と置くと次式を得る。

$$x_\alpha - \bar{x}_\alpha = \lambda_\alpha A, \qquad y_\alpha - \bar{y}_\alpha = \lambda_\alpha B \tag{7}$$

これは 2 点 (x_α, y_α), $(\bar{x}_\alpha, \bar{y}_\alpha)$ を結ぶ線分が直線 $Ax + By + C = 0$ に直交することを意味する。$(\bar{x}_\alpha, \bar{y}_\alpha)$ がその直線上にあることから λ_α が次のように定まる。

$$\lambda_\alpha = \frac{Ax_\alpha + By_\alpha + C}{A^2 + B^2} \tag{8}$$

ゆえに式 (7) から $(\bar{x}_\alpha, \bar{y}_\alpha)$ の最尤推定解 $(\hat{x}_\alpha, \hat{y}_\alpha)$ は次式となる。

$$\hat{x}_\alpha = x_\alpha - \frac{(Ax_\alpha + By_\alpha + C)A}{A^2 + B^2}$$
$$\hat{y}_\alpha = y_\alpha - \frac{(Ax_\alpha + By_\alpha + C)B}{A^2 + B^2} \qquad (9)$$

これは点 (x_α, y_α) から直線 $Ax+By+C=0$ に下ろした垂線の足である (図 1)。これらを式 (2) の $(\bar{x}_\alpha, \bar{y}_\alpha)$ に代入すると、J は次のように書ける。

$$J = \sum_{\alpha=1}^{N} \frac{(Ax_\alpha + By_\alpha + C)^2}{A^2 + B^2} \qquad (10)$$

A, B, C を 0 でない任意の定数倍しても直線 $Ax + By+C = 0$ は変わらないことを反映して上式は A, B, C の 0 次同次式となっている。しかし、例えば $A^2 + B^2 = 1$ と正規化すれば解が符号を除いて一意的に定まる。

式 (10) を C について微分して 0 と置けば、直線は点集合の "重心" を通ることがわかる。重心を原点とする座標系に移せば J は A, B の 2 次形式となり、直線の方向は点集合の "(慣性) モーメント行列" の最大固有値に対する固有ベクトル方向であることがわかる (詳細省略 [10, 11])。

3. 最尤推定のパラドックス

前節の議論は疑いの余地がないように見えるが、統計学の常識からは奇妙である。$\hat{J}$ は当てはめた直線から各点までの垂直距離の 2 乗和であり、垂直距離は期待値 0、分散 σ^2 の正規分布に従うから $\hat{J}/\sigma^2$ は χ^2 分布に従う。N 点あるから、その自由度は真の直線から測ると N であるが、残差を最小にするように直線を当てはめているので直線がもつ 2 自由度だけ減り、$N-2$ となる。χ^2 変数の期待値はその自由度に等しいから $\hat{J}$ は次の期待値を持つ。

$$E[\hat{J}] = (N-2)\sigma^2 \qquad (11)$$

($E[\,\cdot\,]$ は期待値を表す)。ゆえに式 (6) の最尤推定量 $\hat{\sigma}^2$ の期待値は次のようになる。

$$E[\hat{\sigma}^2] = \frac{N-2}{2N}\sigma^2 \qquad (12)$$

これは点数 N が多い極限で $E[\hat{\sigma}^2] \to \sigma^2/2$ となる。しかし統計学の教科書 (例えば [35]) によれば、最尤推定量はごく一般的な条件のもとに (理想的には正規分布に従う誤差の場合に) データ数 ∞ の極限で真の値に収束する (一致性)。このパラドックスはどう解釈すればよいのであろうか。

4. 確率モデルと漸近解析

よく調べるとこの不一致の原因は式 (9) の最尤推定量 $(\hat{x}_\alpha, \hat{y}_\alpha)$ が $N \to \infty$ で真の位置 $(\bar{x}_\alpha, \bar{y}_\alpha)$ に収束しないことにある[1]。もし収束すれば $\hat{J}$ は真の位置からデータまでの距離の 2 乗和に近づく。その期待値は $2N\sigma^2$ であるから、式 (6) の $\hat{\sigma}^2$ は正しく真の値 σ^2 に収束する。

最尤推定量 $(\hat{x}_\alpha, \hat{y}_\alpha)$ が真の位置 $(\bar{x}_\alpha, \bar{y}_\alpha)$ に収束しない理由は明らかである。いくら点数 N を増やしても、増えた点はそれまでの点を直線に沿って補正するための情報を持っていないからである。

それでは「最尤推定量は $n \to \infty$ の極限で真の値に収束する」という命題は誤りなのであろうか。子細に検討すると次のことがわかる。教科書でいう "データ数"n は正しくは「観測回数」である。

ランダム現象にひそむ構造を観測の繰り返しによって検証することが統計学の基本理念である。したがって統計学では「データを多数回観測すればその構造を記述するパラメータが推定できる」という立場をとる。このため、各観測データは同一のパラメータをもつ確率分布 (確率モデル) から発生されると仮定される。このとき観測回数 n を増やせば最尤推定量が真の値に収束することは直観的にも理解できる。

一方、直線当てはめにおける未知パラメータは $\{(\bar{x}_\alpha, \bar{y}_\alpha)\}$, A, B, C, σ^2 であるが、式 (3) の制約があり、かつ A, B, C には定数倍の不定性があるから実質的に $N+3$ 個である。この $N+3$ 個のパラメータをもつ現象から **1** サンプル $\{(x_\alpha, y_\alpha)\}_{\alpha=1}^{N}$ が観測される。

データ点を一つ増やせば新しいパラメータ $(\bar{x}_{N+1}, \bar{y}_{N+1})$ が追加され、$N + 4$ 個のパラメータをもつ新たな現象から 1 サンプル $\{(x_\alpha, y_\alpha)\}_{\alpha=1}^{N+1}$ が観測される。すなわち、どんなにデータ点を増やしても増やしただけパラメータも増え、観測されるのは常に新たな現象からの 1 サンプルである。

言い換えると、どんなに複雑な構造を持っていても観測するのは "1 個の図形" である。すなわち、統計的推測は複数の観測からの推論であるのに対して、幾何学的当てはめは一つの図形からの推論である。

観測回数は常に $n=1$ であるから、教科書にある最尤推定の漸近的性質は決して実現しない。したがって直線当てはめ、およびそれを一般化した幾何学的当てはめではデータ点 $N \to \infty$ の漸近的性質を論じることは意味がない。

5. 無限小摂動解析

この問題を従来の統計学の枠組みに合わせる便法の一つは、データ点と共に増える未知数 $\{(\bar{x}_\alpha, \bar{y}_\alpha)\}$ を "未知の確率分布から得られるサンプル" とみなすことである。そして $\{(\bar{x}_\alpha, \bar{y}_\alpha)\}$ そのものではなく、その分布を特徴づけるパラメータを「隠れ変数」として推定すればよい。これなら「データ数」N を「観測

[1] 統計的推測ではこのような変則的なパラメータを撹乱母数、局外母数などと呼ぶ [6]。

回数」n と同一視できるから $N \to \infty$ の漸近挙動が従来の枠組みで議論できる。統計的推測においてこの考えを一般化したものはセミパラメトリックモデルと呼ばれている [3, 4]。

これは数学的には便利かもしれないが実際的ではない。真の位置 $\{(\bar{x}_\alpha, \bar{y}_\alpha)\}$ は現在観測しているシーンや対象物体の性質から定まるものであり、事前情報を得ることは困難である。“類似の問題から傾向をつかんでその分布を定義する” ことは実際問題として不可能である。また、これはデータが多い極限の挙動を問題とするものである。しかし、コンピュータビジョンやロボット工学ではデータ数が少ないことが多く、「隠れ変数」やセミパラメトリックモデルを用いない理論体系が望まれる。これは可能である。

次のような仮想的状況を考えよう。直線を当てはめるべき N 点を n 回繰り返して観測したとする。実際には同一の装置 (センサや画像処理) を用いる限り、何度観測しても同じ結果しか得られないが、仮に毎回同じ特性を持つ別々の装置を用いるとする。第 α 点の a 回目の観測位置を $(x_\alpha^{(a)}, y_\alpha^{(a)})$ とし、n 回の観測値の平均を次のように置く。

$$x_\alpha = \frac{1}{n}\sum_{a=1}^{n} x_\alpha^{(a)}, \qquad y_\alpha = \frac{1}{n}\sum_{a=1}^{n} y_\alpha^{(a)} \tag{13}$$

観測回数 n を増すと (x_α, y_α) は真の位置 $(\bar{x}_\alpha, \bar{y}_\alpha)$ に収束するから、最尤推定の漸近的性質が実現するはずである。

このとき、各観測装置の誤差の分散が σ^2 なら、式 (13) による (x_α, y_α) の分散は σ^2/n になる。すなわち、仮想的に観測回数を増すことは実質的にデータの誤差を減少させること (高級な観測装置を用いること) に等価である。言い換えれば「観測回数が多い」極限で実現する漸近的性質は「誤差が小さい」極限でも成立する。これが漸近解析と摂動解析の双対性である [11, 15]。

統計的推測ではランダム変動を “繰り返し観測” で克服するので、観測回数を増やすと精度が著しく向上する (同じ精度を得るのに観測回数が少なくてよい) 推定方式が望ましい。それに対して幾何学的当てはめでは精密機器を用いた “1 回の観測” を行うので、装置の性能が上がると精度が著しく向上する (同じ精度を得るのに装置の性能が低くてよい) 推定方式が望ましい [11, 15]。これが統計的推測と幾何学的当てはめにおける「観測」という概念をめぐる基本的な相違である。

6. 誤差の確率モデル

直線当てはめの例では各点は真の位置から正規分布に従う誤差だけずれていると仮定した。このような “確率モデル” はどうして正当化されるのであろうか。

統計的推測では “繰り返し” を前提とするので、データの確率分布とは同一の量を独立に何回も観測したときの値のばらつきと定義される。より具体的には、完全に同一の状況を再現するのは不可能なので、“重要と思われるパラメータ” (制御変数) を同一にして何度も観測したときの “その他の要因” の影響を「誤差」(モデル化誤差) と定義する。“その他の要因” が数多く、かつ独立であれば中心極限定理により正規分布に従うことが期待される。このような記述はそれを解析する統計学者の主観に依存するので、(確率) モデルと呼ぶ。そのようなモデル化が妥当かどうかをデータから検証することが統計的推測の目的である。

一方、画像データのように原理的に “1 個の図形” しか得られない場合はこのような定義ができない。画像に写ったある点の位置はそれを画像処理で検出すれば一意的に定まり、画像や処理が同一である限り何度繰り返しても同じ結果しか得られない。しかし、画像処理で検出した点の位置は画像の離散化やカメラの撮像の感度に左右され、画像処理も近似的な手続きであるので、得られた位置は真の位置から多少はずれている。このずれは厳密な測定が不可能であることに起因する「誤差」(測定誤差) である。

しかし、何度観測しても同じ値しか得られないから、そのずれの程度は経験的に評価するしかない[2]。そこで例えば経験的に 2 画素程度のずれであろうと思われるときには、誤差は期待値 0、標準偏差 2 画素の正規分布に従い、現在観測されているのはその 1 サンプルであるとみなす。この意味で分布の正規性の検証は原理的に不可能である[3]。

しかし、経験的とはいっても誤差の共分散行列を “測定” する手段は存在する。例えば白い領域中の白点のようにその点とその近傍とが区別しにくいときは共分散行列が大きいとみなせる。その大きさはテンプレートマッチングの残差の評価から得られる [7, 27, 33, 34]。これを用いると物体と背景の境界 (エッジ) 上の点ではその境界に沿った細長い分布が得られる。ただし “共分散行列” といっても、そもそも確率モデルが虚構であるから、厳密には “識別しやすさの程度” という意味しかない。しかし、画像処理やコンピュータビジョンではこのような形式的な “確率モデル” に基いた最適推定によって有効な結果が得られる。これは実質的にあいまいなデータを除去し、識別しやすい特徴のみを選択することに相当するからである。

7. モデルとは何か

[2] 観測不可能というのは “機器や計算処理では真の値が得られない” という意味であり、人間が見ると真の位置が判断できることが多い。この意味で “真の” 位置は人間の認識に関わる問題であり、理論化が難しい。

[3] このような検証不可能な分布を導入することは統計的推測でいうベイジアンと呼ばれる立場に相当する。

統計的推測でいう“モデル”とは前述のように現象のモデル化であり、誤差は完全記述の不可能性を意味する。この“確率モデル”の検証や選択が課題である。一方、画像においては誤差は完全測定の不可能性を意味する。そしてその“確率モデル”が定義できるが、これは例え虚構であっても画像の入出力装置や濃淡分布に依存し、画像処理の方法にもよる“物理現象”である。装置や画像が与えられれば固定され、検証や選択の対象ではない。

コンピュータビジョンが「画像認識」、「画像理解」とも呼ばれるのは、その目的が“画像の意味内容”を知ることにあるからである。そして、その可能性としていろいろな「モデル」が立てられる。直線当てはめでは“観測したデータ点の真の位置は同一直線上にある”というモデルを立てた。このような意味内容を記述するモデルを「幾何学的モデル」と呼ぶ。コンピュータビジョンやロボット工学での課題は、このような幾何学的モデルをデータから検証したり、それを用いて画像に写っている対象物の性質を計算することである。同じ“モデル”という用語を用いても確率モデルと幾何学的モデルとは異なる意味を持つ。

8. 幾何学的当てはめ

直線当てはめを一般化すると次のように定式化できる [12, 15]。N 個の m 次元データ $\boldsymbol{x}_1$, ..., $\boldsymbol{x}_N$ が与えられ、各 $\boldsymbol{x}_\alpha$ は真の位置 $\bar{\boldsymbol{x}}_\alpha$ から期待値 $\boldsymbol{0}$、共分散行列 $V[\boldsymbol{x}_\alpha]$ の独立な正規分布に従う誤差 $\Delta\boldsymbol{x}_\alpha$ だけずれているとする。そして、それらの真の位置 $\bar{\boldsymbol{x}}_\alpha$ は p 次元ベクトル $\boldsymbol{u}$ でパラメータ化された r 個の拘束条件

$$F^{(k)}(\bar{\boldsymbol{x}}_\alpha, \boldsymbol{u}) = 0, \qquad k = 1, ..., r \tag{14}$$

を満たすとする。データ $\{\boldsymbol{x}_\alpha\}$ の定義域をデータ空間、未知ベクトル $\boldsymbol{u}$ の定義域をパラメータ空間、r を拘束条件のランクと呼ぶ[4]。このとき r 個の方程式 $F^{(k)}(\boldsymbol{x}, \boldsymbol{u}) = 0$ はデータ空間内に次元 $d = m - r$(余次元[5]r) の多様体[6] $\mathcal{S}$ を定義する。拘束条件 (14) は真の値 $\{\bar{\boldsymbol{x}}_\alpha\}$ が $\mathcal{S}$ 上にあることを要求している。問題は観測データ $\{\boldsymbol{x}_\alpha\}$ からパラメータ $\boldsymbol{u}$ を推定することである。

データの共分散行列 $V[\boldsymbol{x}_\alpha]$ は定数倍を除いて既知とし、次のように書けるとする。

$$V[\boldsymbol{x}_\alpha] = \epsilon^2 V_0[\boldsymbol{x}_\alpha] \tag{15}$$

未知の定数 ϵ をノイズレベル、定性的な挙動を表す既知の $V_0[\boldsymbol{x}_\alpha]$ を正規化共分散行列と呼ぶ。

このように仮定するのが実際的である理由は、まず多くの場合にデータの誤差の絶対的量を推定するのは困難であるのに対して、その定性的挙動を観測することは比較的容易だからである。例えば 6 節に述べたテンプレートマッチングの残差から得られるのは定数倍を除いた共分散行列の評価である。

もう一つの理由は最尤推定において共分散行列の定数倍は解に影響しないので、正規化共分散行列 $V_0[\boldsymbol{x}_\alpha]$ のみを知れば十分だからである。実際、$V_0[\boldsymbol{x}_\alpha]$ に関するマハラノビス距離の 2 乗和

$$J = \sum_{\alpha=1}^{N} (\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha)) \tag{16}$$

を制約条件 (14) のもとで最小化すればよい[7]。ただし、ベクトル $\boldsymbol{a}$, $\boldsymbol{b}$ の内積を $(\boldsymbol{a}, \boldsymbol{b})$ と記す。

誤差が小さいと仮定して式 (14) を線形化し[8]、ラグランジュ乗数を導入して制約条件を除去すると式 (16) は次のように書ける [15]。

$$J = \sum_{\alpha=1}^{N} \sum_{k,l=1}^{r} W_\alpha^{(kl)} F^{(k)}(\boldsymbol{x}_\alpha, \boldsymbol{u}) F^{(l)}(\boldsymbol{x}_\alpha, \boldsymbol{u}) \tag{17}$$

ただし $W_\alpha^{(kl)}$ は $(\nabla_{\mathbf{x}} F^{(k)}, V[\boldsymbol{x}_\alpha] \nabla_{\mathbf{x}} F^{(l)})$ を (kl) 要素とする $r \times r$ 行列の逆行列[9]の (kl) 要素である。

解 $\hat{\boldsymbol{u}}$ はノイズレベル ϵ によらないが、その共分散行列 $V[\hat{\boldsymbol{u}}]$ は ϵ に依存する。無限小摂動 $\epsilon \to 0$ の極限で $V[\hat{\boldsymbol{u}}] \to \boldsymbol{O}$ であることは当然であるが、$O(\epsilon^4)$ の項を除いて精度の理論限界を達成することが証明できる [12, 15, 19]。

これらは正に、統計的推測において観測数を n とするとき、最尤推定量の共分散行列が $n \to \infty$ の漸近的極限で $\boldsymbol{O}$ に収束する (一致性) のみならず、$O(1/n^2)$ の項を除いてクラメル・ラオの下界を達成すること (漸近有効性) の双対命題である。このように統計的推測の $n \to \infty$ の漸近的極限での性質が幾何学的当てはめでは双対的に $\epsilon \to 0$ の無限小摂動の極限で成立する。

9. モデルとモデル選択

統計学ではデータが生成されるランダム現象の解明を目的とする。データ $\boldsymbol{x}$ はパラメータを用いて確定的な式とランダム誤差によって表現され、そのパラメータを推定することが統計的推測である。抽象化すると、パラメータ $\boldsymbol{\theta}$ をもつ確率密度 $p(\boldsymbol{x}|\boldsymbol{\theta})$ から発生したデータ列から $\boldsymbol{\theta}$ を推定することである。パラメータ

[4] データが m 次元空間のある m' 次元多様体に拘束されていたり、パラメータ $\boldsymbol{u}$ が p 次元空間のある p' 次元多様体に拘束されていてもよい [12, 15]。また式 (14) の r 個の式には従属関係があってもよい。その場合のランクは独立な式の個数 r' である [15]。

[5] 一般次元空間の“曲面”に直交する方向の次元。

[6] 一般次元空間の“曲面”。

[7] データが m 次元空間のある m' 次元多様体に拘束されているときは $V_0[\boldsymbol{x}_\alpha]$ はランク m' の特異行列となるので、$V_0[\boldsymbol{x}_\alpha]^{-1}$ は (ムーア・ペンローズの) 一般逆行列 $V_0[\boldsymbol{x}_\alpha]^-$ に置きかえる [15]。

[8] 無限小摂動の極限を解析していることに注意。

[9] 拘束条件 (14) に従属関係があり、独立な式の個数 (ランク) が r' のときは、大きい r' 個以外の固有値を 0 に置き変えてから (ムーア・ペンローズの) 一般逆行列 (ランク r') を計算する [15]。

$\boldsymbol{\theta}$ は確定的な要因を記述する方程式の係数、および誤差の期待値や分散のようなランダム誤差の特性を記述するパラメータを含んでいる。この確率密度に複数の可能性 $p_1(\boldsymbol{x}|\boldsymbol{\theta}_1)$, $p_2(\boldsymbol{x}|\boldsymbol{\theta}_2)$, ... がある場合に、各々を (確率) モデルと呼ぶ。それぞれは "関数空間の 1 点" である [2, 28]。そして、どれが妥当かを判定するのが (確率) モデル選択である。

これに対して、データ $\boldsymbol{x}$ の満たすべき "陰関数" $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{u}) = 0$ のパラメータ $\boldsymbol{u}$ を推定するのが幾何学的当てはめであり、直線当てはめが典型的な例である。 $\boldsymbol{u}$ は幾何学的構造を記述する方程式の係数から成り、ランダム誤差の特性を記述するパラメータを含んでいない。そして、データ $\boldsymbol{x}$ をパラメータ $\boldsymbol{u}$ の式によって陽に表すことができない。この二つが統計的推測との根本的な相違である。その拘束条件に複数の可能性 $\boldsymbol{F}_1(\boldsymbol{x}, \boldsymbol{u}_1) = \boldsymbol{0}$, $\boldsymbol{F}_2(\boldsymbol{x}, \boldsymbol{u}_2) = \boldsymbol{0}$, ... がある場合に、各々を (幾何学的) モデルと呼ぶ。それぞれはデータ空間中の "多様体" である。そして、どれが妥当かを判定するのが (幾何学的) モデル選択である [14, 15, 20]。

同じ "モデル" といっても幾何学的モデル $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{0}$ は現象の確率的機構には無関係である。現象の確率的機構は既知の正規化共分散行列 $V_0[\boldsymbol{x}_\alpha]$ をもつ正規分布と仮定され、もはや選択の余地がない。選択すべきはデータの従属関係である。

10. 赤池の AIC

統計的推測における (確率) モデル選択において素朴な考え方は、各々のモデルを最適に当てはめてその残差が最小になるものを選ぶことである。しかし、モデルはデータに最も適合するように推定しているので、自由度が最も多いモデルの残差が常に最小となる。そこでデータと同じ統計的な性質を持ち、かつモデルに適合させたデータとは独立な仮想的なデータに対する残差の期待値を評価しようというのが赤池の AIC[1] の着想である[10]。

これは期待値を評価するので、本質的には漸近評価は不要であるが、漸近評価を用いると AIC が簡単な形で表される。赤池は次の二つの「漸近近似」を用いた。

- 観測数が増加すると最尤推定量は真の値に収束する (大数の法則)。
- 観測数が増加すると最尤推定量は漸近的に正規分布に従う (中心極限定理)。

これらを用いて対数尤度の期待値を近似的に評価し、正規分布に従う変数の 2 次形式は χ^2 分布に従うこと、および χ^2 分布の期待値はその自由度に等しいという性質から赤池の AIC が導かれる [32]。これは統計学で広く用いられているだけでなく、コンピュータビジョンやロボット工学においても統計的推測の形の問題にいろいろ利用されている [5]。

[10] これは相互検証法と呼ばれる実験的手法やブートストラップ [9] と呼ばれるシミュレーションと類似の考え方である

11. 幾何学的 AIC

赤池の AIC は本質的に期待値の評価であって漸近解析に依存しないから、漸近解析が意味を持たない幾何学的当てはめにも適用できる[11]。金谷は幾何学的当てはめに「データと同じ統計的な性質を持ち、モデルに適合させたデータとは独立なデータに対する残差の期待値を評価する」という赤池の考えを適用した。データ $\{\boldsymbol{x}_\alpha\}$ から推定した $\{\bar{x}_\alpha\}$ の最尤推定量を $\{\hat{x}_\alpha\}$ とすると、その残差は次のように書ける。

$$\hat{J} = \sum_{\alpha=1}^{N} (\boldsymbol{x}_\alpha - \hat{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{x}_\alpha - \hat{\boldsymbol{x}}_\alpha)) \tag{18}$$

データ $\{\boldsymbol{x}_\alpha\}$ と独立で同じ統計的性質を持つ仮想的なデータを $\{\boldsymbol{x}^*_\alpha\}$ とすると、モデルのよさは次の残差で測られる。

$$\hat{J}^* = \sum_{\alpha=1}^{N} (\boldsymbol{x}^*_\alpha - \hat{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{x}^*_\alpha - \hat{\boldsymbol{x}}_\alpha)) \tag{19}$$

$\{\boldsymbol{x}_\alpha\}$ および $\{\boldsymbol{x}^*_\alpha\}$ に関する期待値をそれぞれ $E[\,\cdot\,]$、$E^*[\,\cdot\,]$ とすると次のようになる [15, 20]。

$$E^*[E[\hat{J}^*]] = E[\hat{J}] + 2(Nd + p)\epsilon^2 + O(\epsilon^4) \tag{20}$$

ここに $d = m - r$ は拘束条件 (14) がデータ空間に定義するモデル多様体 $\mathcal{S}$ の次元である。上式は残差 $\hat{J}$ はデータに適合させたため独立なデータに関する残差 $\hat{J}^*$ にくらべて平均的にほぼ $2(Nd + p)\epsilon^2$ だけ減少することを意味している。そこで幾何学的 **AIC** を次のように定義する[12]。

$$\text{G-AIC} = \hat{J} + 2(Nd + p)\epsilon^2 \tag{21}$$

式 (20) の評価はノイズレベルが小さい摂動解析に基づいており、漸近評価は行っていない。それでも漸近評価に基いて導いた赤池の AIC とほとんど同じ形が得られるのは次の二つの「摂動近似」を用いたからである。

- ノイズレベルが減少すると最尤推定量は真の値に収束する。
- 誤差は正規分布と仮定しているから、線形な拘束条件のもとでは最尤推定量は正規分布に従う。拘束条件が非線形でもノイズレベルが小さい極限では解の近傍での線形近似が正当化される。

[11] AIC はモデルとしての確率密度と真の確率密度との **Kullback-Leibler** 距離 (情報量)[25] の漸近的最小化という解釈もなされる。この解釈は幾何学的モデルには当てはまらない。

[12] 本論文では赤池の AIC と対比させて G-AIC と表記するが、幾何学的当てはめに定義されるのは式 (21) のみであり、AIC と G-AIC の 2 種類があるわけではない。このため文献 [14, 15, 20] では単に AIC と表記している。

そして、正規分布に従う変数の 2 次形式は χ^2 分布に従うこと、および χ^2 分布の期待値はその自由度に等しいという事実から式 (20) が得られる (ただし計算はかなり複雑になる [15, 20])。

このように、式 (21) は式 (20) から得られるのであって、赤池の **AIC** のモデル自由度の部分に未知数の数を代入して得られるのではない。統計的推測と幾何学的当てはめはその基盤が互いに双対であり、漸近評価による赤池の AIC の導出と摂動解析による幾何学的 AIC の導出は基本的に異なる。それにもかかわらず同じような形が得られるのは、 AIC の思想が本質的に期待値評価であり、漸近評価や摂動解析は単なる近似の技術に過ぎないからである。

12. ノイズレベルの推定

赤池の AIC と幾何学的 AIC の本質的な相違が現れるのは式 (21) 中の ϵ^2 の推定である。これが既知であれば問題ないが、未知の場合にはどう推定すればよいのであろうか。

統計的推測における確率モデルは何が確定的構造で何がランダムな要因 (誤差) かを記述するものであるから誤差もモデルの性質である。したがってその分散もモデルパラメータであり、最尤推定で推定される。それに対して前述のように幾何学的当てはめでは誤差の分散はモデルパラメータではない。

幾何学的当てはめでは、直線当てはめを例にとって誤差の分散 σ^2 の最尤推定量 (6) が誤った推定値を与えること、およびその理由は既に述べた。正しい推定は次の通りである。

$$\hat{\sigma}^2 = \frac{\hat{J}}{N-2} \tag{22}$$

これは残差 $\hat{J}$ の期待値が式 (11) で与えられることから得られ、 σ^2 の不偏推定量である。一般化すると、式 (16) の残差 $\hat{J}$ を用いて次のようにすればよい。

$$\hat{\epsilon}^2 = \frac{\hat{J}}{rN-p} \tag{23}$$

幾何学的には $\hat{J}$ はデータ点から式 (14) の定義するモデル多様体 $\mathcal{S}$ への (マハラノビス) 距離の 2 乗和である。この多様体に直交する方向の次元 (余次元) が r であるから、残差 $\hat{J}$ の期待値は $rN\epsilon^2$ となるはずであるが、曲面をデータに最適に当てはめているので、その自由度だけ減って $(rN-p)\epsilon^2$ となる。

しかし、これはモデルが正しい場合にのみ正しい推定値を与える。モデルが正しいかどうか不明な場合はどうすればよいのであろうか。金谷 [14, 20] はモデル選択を退化の検出に適用することを想定してこの困難を解決した。

画像やセンサデータの解析においてはシーンに関して成立する知識 (物体は剛体運動をする、等) を利用する。これが拘束条件 (14) に対応する。しかし、それが例外的に退化した場合 (例えば運動が 0 である、等) では一般的な条件を前提にした計算が破綻する。厳密な退化でなくても退化に近いと計算が不安定になる。このような場合に、モデル選択により退化を自動的に検出し、退化を記述するモデルに切り換えれば計算が安定化される [16, 17, 18, 22, 23, 26, 31, 39]。

退化が生じるということは拘束条件 (14) に新たな式 (ある量が 0 である、等) が加わることを意味する。しかし拘束条件 (14) そのものは成立している。そのような「一般モデル」は「退化モデル」が成立しようがしまいが成立するから、 ϵ^2 は一般モデルの残差 $\hat{J}$ から式 (23) によって推定すればよい。

これは統計的推測との際立った相違である。統計的推測では誤差は記述の不完全性に起因し、その分散は (確率) モデルのパラメータ $\boldsymbol{\theta}$ に含まれている。これは一般にモデルごとに異なるので、候補モデルごとに独立に推定される。これに対して幾何学的当てはめでは、誤差は測定の不完全性に起因するので、データの意味内容を表す (幾何学的) モデルのパラメータ $\boldsymbol{u}$ には含まれていない。したがってどの候補モデルにも共通でなければならないから、常に真と仮定される一般モデルから推定するのが妥当である。

“データ数の多い” 極限の統計的推測と “ノイズレベルの小さい” 極限の幾何学的当てはめの双対関係から、統計的推測の “データ数” n と幾何学的当てはめの “ノイズレベル” ϵ とが互いに双対であるとも解釈できる。実際、金谷の統計的最適化理論 [15] の多くの部分は ϵ を $1/\sqrt{n}$ に置きかえると通常の統計的推測理論に帰着する。したがって、幾何学的当てはめにおけるノイズレベルの推定は統計的推測における “データ数の推定” と同じ役割を果している。

統計的推定において (実際にはあり得ないが仮に) データ数が未知とする。残差の期待値はデータ数と誤差の分散に依存するから、誤差の分散が既知なら残差から推定できる。当然、これをモデルごとに最尤推定することはできない。なぜならデータ数は観測過程で定まるものであり、 (確率) モデルごとに異なってはならないからである。これは真と仮定されるモデルから共通に推定しなければならない。

同様に幾何学的当てはめではノイズレベルは観測過程で定まるものであり、 (幾何学的) モデルごとに異なってはならず、真と仮定されるモデルから共通に推定しなければならない[13]。

13. まとめ

本論文では平面上の点列への直線当てはめを例にとり、「幾何学的当てはめ」と「統計的推測」との違い

[13] 赤池の AIC にデータ数 n の式を掛けたり足したりしても、 n は各モデルに共通であるから相対的な比較に影響しない。同様に、幾何学的 AIC にノイズレベル ϵ の式を掛けたり足したりしても、 ϵ は各モデルに共通であるから相対的な比較に影響しない。

を述べた。統計学は農業、医学、政治、経済、等への応用を指向して発展したため、それに適した「統計的推測」の枠組みが確立したが、画像やセンサデータを用いるコンピュータビジョンやロボット工学では本論文でいう「幾何学的当てはめ」が最も普遍的な問題設定である。

本論文では、統計的推測が“繰り返しデータ”からの推論であるのに対して、幾何学的当てはめは“1 回の観測図形”からの推論であり、統計的推測の数学的基盤となるデータ数 ∞ の極限の「漸近的解析」が幾何学的当てはめでは誤差 0 の「無限小摂動」の極限で成立するという「双対性」を述べた。

統計的推測における「(確率) モデル」は確率現象の発生機構を説明する“確率密度”であり、数学的には“関数空間の 1 点”である。これに対して幾何学的当てはめにおける「(幾何学的) モデル」はデータの依存関係を陰に指定する“拘束条件”であり、数学的にはデータ空間中の“多様体”である。このため、誤差の分散は確率モデルのパラメータであるが、幾何学的モデルのパラメータではなく、その推定の方法が統計的推測と幾何学的当てはめでは根本的に異なる。

最後に「確率的モデル選択」と「幾何学的モデル選択」を対比し、「赤池の AIC」と「幾何学的 AIC」の導出の相違を述べた。最近、長尾ら [29] も「幾何学的 AIC」や「幾何学的 MDL」と呼ぶ基準を定義しているが、これについては次報 [21] で述べる。

参考文献

[1] H. Akaike, A new look at the statistical model identification, *IEEE Trans. Automatic Control*, **16**-6 (1974), 716–723.

[2] S. Amari, *Differential-Geometric Methods in Statistics*, Springer, Berlin, 1985.

[3] 甘利俊一, 川鍋元明, 線形関係の推定 — 最小 2 乗法は最良であるのか?, 応用数理, **6**-2 (1996), 96–109.

[4] P.J. Bickel, C. A. J. Klassen, Y. Ritov and J. A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopking University Press, Baltimore, MD, 1994.

[5] K. K. Boyer, J. Mirza and G. Ganguly, The robust sequential estimator: A general approach and its application to surface organization in range data, *IEEE Trans. Patt. Anal. Mach. Intell.*, **16**-10 (1994), 987–1001.

[6] A. J. Dobson, *An Introduction to Generalized Linear Models*, Chapman-Hall, 1990; 田中豊, 森川敏彦, 栗原考次 (訳), 「統計モデル入門 — 回帰モデルから一般線形モデルまで —」, 共立出版, 1993.

[7] W. Förstner, Reliability analysis of parameter estimation in linear models with applications to mensuration problems in computer vision, *Comput. Vision Graphics Image Process.*, **40** (1987), 273–310.

[8] W. A. Fuller, *Measurement Error Models*, Wiley, New York, 1987.

[9] B. Efron and R. J. Tibshirani, *An Introduction to Bootstrap*, Chapman-Hall, New York, 1993.

[10] K. Kanatani, *Geometric Computation for Machine Vision*, Oxford University Press, Oxford, U.K., 1993.

[11] 金谷健一, 「空間データの数理 — 3 次元コンピューティングに向けて —」, 朝倉書店, 1995.

[12] 金谷健一, 当てはめ問題の最適推定と精度の理論限界, 情報処理学会論文誌, **36**-8 (1995), 1865–1873.

[13] 金谷健一, 幾何学的補正問題の最適計算と精度の理論限界, 情報処理学会論文誌, **37**-3 (1996), 363–370.

[14] 金谷健一, 情報量基準による幾何学的モデル選択, 情報処理学会論文誌, **37**-6 (1996), 1073–1080.

[15] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier Science, Amsterdam, The Netherlands, 1996.

[16] 金谷健一, 自己評価を伴うアクティブビジョン, 日本ロボット学会誌, **15**-2 (1997), 268–274.

[17] K. Kanatani, Comments on “Symmetry as a Continuous Feature”, *IEEE Trans. Patt. Anal. Mach. Intell.*, **19**-3 (1997), 246–247.

[18] K. Kanatani, Comments on “Nonparametric Segmentation of Curves into Various Representations”, *IEEE Trans. Patt. Anal. Mach. Intell.*, **19**-12 (1997), 1391–1392.

[19] K. Kanatani, Cramer-Rao lower bounds for curve fitting, *Graphical Models Image Processing*, **60**-2 (1998), 93–99.

[20] K. Kanatani, Geometric information criterion for model selection, *Int. J. Comput. Vision*, **26**-3 (1998), 171–189.

[21] 金谷健一, 松永 力, 幾何学的ＭＤＬとそのメディア応用, 情報処理学会コンピュータビジョンとイメージメディア研究会, 2000-CVIM-122 (2000-5).

[22] Y. Kanazawa and K. Kanatani, Infinity and planarity test for stereo vision, *IEICE Trans. Inf. & Syst.*, **E80-D**-8 (1997), 774–779.

[23] 金澤靖, 金谷健一, 幾何学的 AIC による画像モザイク生成の安定化, 電子情報通信学会論文誌 A, **J83-A**-6 (2000), to appear.

[24] 木下敬介, M. Lindenbaum, 幾何学的 AIC によるカメラモデル選択, 情報処理学会コンピュータビジョンとイメージメディア研究会, 99-CVIM-114-8 (1999-1), 57–64.

[25] S. Kullback, *Information Theory and Statistics*, Wiley & Sons, New York, 1951.

[26] 松永力, 金谷健一, 平面パタンを用いる移動カメラの校正：最適計算，信頼性評価，および幾何学的 AIC による安定化, 電子情報通信学会論文誌 A, **J83-A**-6 (2000), to appear.

[27] D. D. Morris and T. Kanade, A unified factorization algorithm for points, line segments and planes with uncertainty models, *Proc. 6th Int. Conf. Comput. Vision*, January 1998, Bombay, India, pp. 696–702,

[28] M. K. Murray and J. W. Rice, *Differential Geometry and Statistics*, Chapman-Hall, London, 1993.

[29] 長尾淳平, 韓太舜, かく乱母数を含む場合の MDL 基準の構築と空間図形モデル推定問題への応用, 電子情報通信学会論文誌 A, **J83-A**-1 (2000), 83–95.

[30] J. Neyman and E. L. Scott, Consistent estimates based on partially consistent observations, *Econometrica*, **16**-1 (1948), 1–32.

[31] N. Ohta and K. Kanatani, Moving object detection from optical flow without empirical thresholds, *IEICE Trans. Inf. & Syst.*, **E81-D**-2 (1998), 243–245.

[32] 坂本慶行, 石黒真木夫, 北川源四郎, 「情報量統計学」, 共立出版, 1983.

[33] J. Shi and C. Tomasi, Good features to track, *Proc. IEEE Conf. Comput. Vision Patt. Recogn.*, June 1994, Seattle, WA, pp. 593–600.

[34] A. Singh, An estimation-theoretic framework for imageflow computation, *Proc. 3rd Int. Conf. Comput. Vision*, December 1990, Osaka, Japan, pp.168–177.

[35] 竹内啓, 「数理統計学」, 東洋経済新報社, 1963.

[36] P. H. S. Torr, An assessment of information criteria for motion model selection, *Proc. IEEE Conf. Comput. Vision Patt. Recogn.*, June 1997, Puerto Rico, pp. 47–53.

[37] P. H. S. Torr, Geometric motion segmentation and model selection, *Phil. Trans. Roy. Soc.*, A-**356**-1740 (1998), 1321–1340.

[38] P. H. S. Torr, A. W. Fitzgibbon and A. Zisserman, The problem of degeneracy in structure and motion recovery from uncalibrated image sequences, *Int. J. Comput. Vision*, **32**-1 (1999), 27–44.

[39] Iman Triono, N. Ohta and K. Kanatani, Automatic recognition of regular figures by geometric AIC, *IEICE Trans. Inf. & Syst.*, **E81-D**-2 (1998), 246–248.