

## 曲線のあてはめとアルゴリズムアニメーションについて

北海道大学 工学部 水田正弘

### 1. はじめに

探索的データ解析における基本的な考え方に、グラフィックスの利用がある。データを様々な統計グラフによって表現することによって、単に数値を眺めているときには予想もできなかった特徴を見いだすこともある。Goto et al. (1991) では、1970年から現在まで発表された統計グラフを整理・分類している。

しかし、統計学一般、特に統計教育の分野におけるグラフィックスの利用を、統計グラフに限定すべきではない。情報処理を中心とした各分野では、scientific visualization, visual programming, visual languages, algorithm animationなどをキーワードとして様々なグラフィックスの利用が研究されている。

今回の報告では、統計教育における「グラフィックス」の利用の可能性について、データの可視化、システムの操作の可視化、プログラムの可視化、アルゴリズムの可視化に分けて検討する。また、「曲線のあてはめ」を例として、この種のシステムを考察した。

### 2. 様々な可視化について

データ解析に関連する分野において「可視化」することが望ましい対象がいろいろ存在する。以下では、それらについて紹介する。

#### 2.1 データの可視化

データの可視化 (visualization of data and information about data) とは、我々が通常使う統計グラフを含んだ概念である。さらに、統計データベース (穂鷹・佐藤(1988)) などにおけるデータ構造を可視化することも指している。

#### 2.2 システムの操作の可視化

New New Sをはじめ最近の統計解析システムでは操作性が向上している。マウスやアイコンを効果的に利用することにより、よりよいマンマシンインタフェースが提供されている。

システムを直接操作する感覚を作り出す、direct manipulation (直接操作) の考え方は参考になる。さらに最近では、virtual reality (人工現実感・仮想現実感) が注目されている (服部(1991), クルーガー (1991))。

#### 2.3 プログラムの可視化

FORTRAN, BASIC, Cをはじめとするプログラム言語の大部分はすべて文字列で記述する。従って、キーボードに慣れない人や、計算機それ自体に興味がないユーザにとってプログラムの学習は大変困難である。

かつて「プログラムを書く前に必ずフローチャートを作成するように」教育していた時期があった。1つの命令を1つのフローチャートで表す学生が出ること、トップダウンの考え方を阻害することなどから、近年では、フローチャートをはじめに書かせる教育はほとんどなされて

いない。しかし、フローチャートなどによりプログラムを視覚的に構成する方法は初心者にとって心理的負担が少ないと考えられる。そこで、マウスとキーボードを使って、各種チャートを構成すると自動的にプログラムなどが完成するシステムの研究がなされている（水田（1990）、菊地（1991））。

## 2.4 アルゴリズムの可視化

計算機教育において、プログラムが単に書けることが教育における最終的なゴールではない。例えば、並べ換え（sorting）を行うのに、同じ結果を得る複数のアルゴリズムが存在する。利用する場面に応じて適切なアルゴリズムを選択・作成する能力が重要になる。

一般に、文章で説明されたアルゴリズムを理解するのは易しいことではない。むしろ、ソースコードを読む方が楽なことさえある。そこで、アルゴリズムを可視化することが考えられている。前節で紹介した各種チャートでもアルゴリズムを表すことができるが、むしろアルゴリズムを動きのあるアニメーションで表示することが効果的である（木下（1991））。

統計学において、アルゴリズムの中心が線形代数である手法ではチャートによる表現が理解しやすいかもしれない。しかし、計算機指向的な手法では、アルゴリズムアニメーションが有効であると思われる。

## 3. 曲線あてはめに関するシステムの例

Hastie and Stuetzle (1989) により提案された principal curves をはじめとする曲線あてはめ法では、グラフィックスの利用が基本的である。すなわち、得られた曲線の妥当性を確認するにはデータ点と曲線を表示するのが有効である。

Principal curves のアルゴリズムは、projection step と expectation step と呼ばれる2つの操作の繰り返しにより構成されている。いくつかのパラメータの与え方によっては収束の速さや収束する曲線が変わる。このような収束のプロセスを動画像によって表現することにより、アルゴリズムの実行過程を示すことができる。

## 4. おわりに

統計教育に「グラフィックス」を利用する方法は数多く存在する。今回の報告では、ビデオを利用して関連するいくつかのシステムを紹介した。

## 参 考 文 献

- Goto, M., Matsubara, Y., Yoden, A., Tsuchiya, Y. and Wakimoto, K. (1991). Statistical graphics: a classified and selected bibliography, *J. Japan Statist. Soc.*, **21**, 97-121.
- Hastie, T. and Stuetzle, W. (1989). Principal curves, *J. Amer. Statist. Assoc.*, **84**, 502-516.
- 服部 桂 (1991). 『人工現実感の世界』, 工学調査会, 東京.
- 穂鷹良介, 佐藤英人 (1988). 『統計データベースの設計と開発』, オーム社, 東京.
- 菊地修一郎 (1991). 視覚的構造化機能を有するプログラム作成支援システムの開発, 平成2年度北海道大学工学部卒業論文.
- 木下 茂 (1991). 教育用アルゴリズムアニメーションシステムに関する研究, 平成2年度北海道大学工学部卒業論文.
- クルーガー, M.W. (1991). 『人工現実——インタラクティブ・メディアの展開——』(下野隆生 訳), トップラン, 東京.

水田正弘 (1990). 視覚的操作によるデータ解析システムについて, 計算機統計学, 3, 23-29.

## 統計教育におけるグラフの効用

統計数理研究所 馬場 康 維

### 1. グラフの効用

グラフ表現というと多くの人々はヒストグラム, 散布図あるいは実験の結果を折れ線グラフに表わしたものなど, 何らかの情報の視覚的な表現法, 記述的なものという認識を持つであろう. しかし, 観測値のグラフから見出された関数関係によって科学における様々な重要な発見が成されてきたことを考えてみれば, グラフは単なる記述の道具ではなく, 記述することによって思考・推論を助けるという重要な役割を担っていることが分る. また, グラフのトポロジカルな特徴の視察に置換えることによってある種の積分方程式を解くプロセスを簡略化するためのファイマンダイアグラムのように, 数学的なグラフも存在する.

コンピュータの普及もその一因となり, 統計学の分野に限ってみても, 今日では, 様々な実用的なグラフ表現法が提案されている. グラフの特徴は, 直観に訴えることによる分りやすさと情報の多さである. グラフの中でも, 理解しやすいグラフは, 統計的な概念や数学的な概念, あるいは統計モデルなどとの対応が明白で, それらをうまく表現しているものが多い. 元来, 分析のためのもの, あるいは結果の表示のためのものとして提示されたものであっても, この様なグラフは, 統計教育においても概念やモデルの説明に有効である.

ここでは, 統計教育の場でグラフが統計的な概念の説明に役に立つ例を説明する.

### 2. 統計教育におけるグラフの利用

グラフは統計教育の様々な場面で用い得るが, 大別して以下のようなものがある.

#### (1) 学習の動機づけ

統計的な概念の説明に, 実際のデータを見せその背景を説明したり, データの中身に興味を持たせ学習意欲を持たせるために用いる.

#### (2) 統計的および数学的な概念の説明

ヒストグラムから分布の概念を導入する, 散布図から相関の概念を導入する等, グラフによって各種の統計的な概念を説明するのに有効である. 特に, 重相関, 偏相関などの概念は特性ベクトルによる幾何学的な説明が役立つ.

#### (3) 統計的なモデルの説明

重回帰分析, 数量化I類等のモデルの性質を理解させる. これについては, 次章で述べる.

### 3. 数量化I類と星座グラフ

数量化I類による推測のモデルの説明に星座グラフが有効である. 図1は数量化の結果得られた予測式の構造を表わしたもので, 説明アイテムに与えた数量を半円の角度で示したものである. こういったグラフで以下のことが説明できる.

(1) 予測値は説明アイテムに対する反応パターンによって決る. したがって, 説明アイテム