JAIST Repository
https://dspace.jaist.ac.jp/

Title	統計的モデルの考え方 ： 確率分布から多変量解析まで
Author(s)	日高，昇平；西田，豊
Citation	日本認知科学会・学習と対話研究分科会・第46回研究会，2012(1)：21-26
Issue Date	2012-03-22
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/10930
Rights	日高 昇平，西田 豊，日本認知科学会・学習と対話研究分科会・第46回研究会，2012(1)，2012，21-26.
Description	日本認知科学会・対話と学習研究会・第46回研究会，テーマ：認知科学における数理・計算モデル：体験的入門チュートリアルと研究の実際，日時　2012年 3月22日（木）9:30～18:30，場所　東京工業大学・大岡山キャンパス学術国際情報センター（情報棟）３階・第一実習室、大岡山西9号館研究会議室607

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

統計的モデルの考え方：
確率分布から多変量解析まで

Statistical modeling: probabilistic distribution to multivariate analysis

日高 昇平[1]・西田 豊[2]

Shohei Hidaka, Yutaka Nishida

[1]北陸先端科学技術大学院大学・[2]大阪大学

[1]Japan Advanced Institute of Science and Technology, [2]Osaka University

Abstract: In the present manuscript, we introduce a basic idea underlying statistical modeling as application for research in cognitive science. In order to introduce the minimum prerequisites, a hypothetical case was studied using binomial distribution as a binary decision making model. Walking through this case study, the idea of probability distribution, model selection, statistical test, and parameter estimation were briefly outlined. In the latter half of the paper, we studied a multivariate clustering model, structure discovery process from unlabeled data, as similarity judgment on conceptual space. The underlying assumption, parameter estimation, and interpretation of parameters in terms of cognitive process were discussed.

確率論の考え方は、実験から理論まで幅広く認知科学の各分野で用いられている(Nickerson, 2004)。たとえ、”同じ被験者”が”同じ刺激”を受けたとしても、多くの場合、まったく同じ反応を繰り返す事は少ない。これは、実験者が想定する”同じ被験者”が必ずしも同じではない(今日の私と明日と私がまったく同じ”気分”であるか分からない)、あるいは、実験で想定した範囲で”同じ刺激”であっても、思わぬ要因が結果に影響する可能性があり、実はわずかに異なる刺激であるかもしれない。従って、実験では全ての要因を統制する事ができないために、想定した範囲で同一の状況であっても、異なる結果を得る場合がある。従って、このような、実験で想定されない各種・雑多な要因をまとめて、「ノイズ」と呼び、想定される要因だけでなく、ノイズを含めて人の行動を確率過程とみなす。この意味で、認知過程のモデリングだけでなく、認知科学実験を設計・遂行・分析を行うためにも、確率の基本的な考え方を踏まえることは重要である。本稿では、確率の基礎知識から、多変量解析までを、技術的な内容よりも、”なぜそう考えるのか”に焦点を当てて紹介したい。

確率・確率分布

まずは、比較的単純な具体例について考えてみよう。N 人の被験者から、各 1 回ずつ、２つの選択肢の中から、１つを選ぶ、という実験を行ったとする。最も単純なモデルの一つは、各被験者が同じ確率的なメカニズムに基づいて、回答を行った、と仮定するものだろう。この実験では、二肢強制選択課題なので、その一つを選ぶ確率を p,もう一方を選ぶ確率を(1-p)とする。被験者 *i* が選択肢 1 を選ぶ事象を X_i=1、その確率を p, 選択肢 2 を選ぶ事象を X_i=0、その確率を(1-p)と表すとしよう。

$$P(X_i = 1) = p,\ P(X_i = 0) = 1 - p$$

この実験では、N 人の被験者の回答のうち、何人が選択肢 1 あるいは 2 を選んだのか、これが主な結果である。従って、選択肢 1 を選んだ被験者の人数を M とすると、その確率は以下のようになる。

$$P\left(\sum_{i=1}^{N} X_i = M\right) = \frac{N!}{(N-M)!M!} p^M (1-p)^{N-M}$$

M の取りうる範囲(M=0,1,2, .., N)の確率の集合

を確率分布と呼び、特にこの式は二項分布と呼ばれる。

モデル選択・検定

仮に実験者がこの実験前に、二つの仮説を持っていたとする。

仮説 1: 被験者は二つの選択肢 1・2 を等しい確率で選ぶ。すなわち p=1/2。

仮説 2: 被験者は選択肢 1 を選択肢 2 の 2 倍の確率で選ぶ。すなわち p=2/3。

このような仮説を立てて実験を行った結果、20 人の被験者について、選択肢 1 を選んだ被験者は 14 人であったとしよう(N=20, M=14)。さて、この実験結果に照らし合わせると、それぞれの仮説はどの程度尤もらしいのだろうか。

仮説 1・2 はそれぞれ、異なるパラメータ p を持つ二項分布として表現できる。このように、仮説を確率的に表現したものが、確率(統計的)的モデルである。実際にモデルに基づき確率を計算をしてみると(あるいは図 1 の確率分布関数(青)の M に対応する数値を読み取る)、

$$\frac{(20)!}{6!14!}\left(\frac{1}{2}\right)^{N} = 0.0370$$

$$\frac{(20)!}{6!14!}\left(\frac{2}{3}\right)^{M}\left(\frac{1}{3}\right)^{N-M} = 0.1821$$

従って、仮説 1 に基づくと 0.0370, 仮説 2 に基づくと 0.1821 の確率で実験で得られた結果が発生することになる。従って、この 2 つの仮説を比べた場合、仮説 2 の尤度が高く、これを選ぶのが妥当と言えるだろう。このように仮説を数理的に表現したモデルを、ある基準(この場合尤度)に基づいて選択する事をモデル選択と呼ぶ。

モデル選択では、複数のモデルを立て、相対的にどのモデルが最も良くデータを説明するか検討した。一方、個々のモデルについて、それぞれがどの程度の確率でデータを説明できるか知りたい場合がある。

仮説 3: 被験者は二つの選択肢 1・2 を等しい確率で選ぶ。すなわち p=1/2。実験結果はこのような過程で得られる値以下になる。

仮説 3 は仮説 1 と類似しているが、ある値だけではなく幅をもって仮説を表現していることに注意(M≦13)。この仮説 3 は、この実験において、被験者が等確率で無作為に回答を行った場合に起こりうる範囲を表現している。実験者が、このような等確率仮説の範囲を超えて、高い M の値を得られたかどうか知りたい場合、仮説 3 は帰無仮説(null hypothesis)と呼ばれる。もし帰無仮説に従って結果が生起する確率が十分に低いならば、仮説 3 を棄却する。これを計算すると(あるいは図 1 の累積確率分布関数(緑)の対応する値を読み取る)、

$$\sum_{M=0}^{13}\frac{(N)!}{(N-M)!M!}p^{M}(1-p)^{N-M} = 0.9423$$

を得る。従って仮説 3 の下で M≧14 が発生する確率は 0.0577 であり、仮に 0.05 を検定の水準として採用していた場合、仮説 3 は棄却されない事になる。

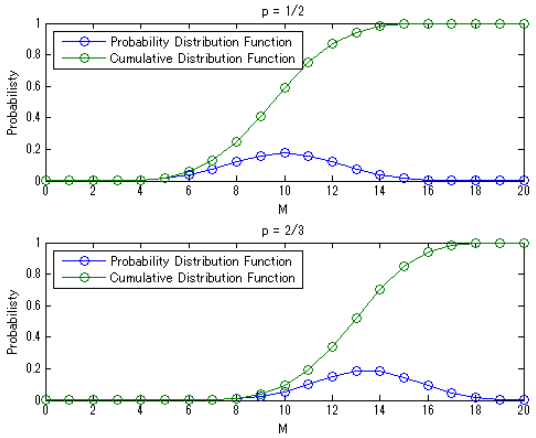

図 1: p=1/2(上), p=2/3(下)のパラメータを持つ二項分布の確率分布関数(青), 累積確率分布関数(緑)

最尤推定法：モデルパラメータ推定

次に、実験者が事前にある特定の仮説を持っておらず、逆に、実験結果から、妥当な仮説を生成したいと考えていたとしよう。単純化のため、選択モデルを二項分布に限定した場合、可能な仮説はあるパラメータ p を持つ二項分布となる。従って、実験結果から妥当な仮説を生成するには、実験結果を生成する確率を最大にするパラメータを持つモデルを選択するのが一つの方法である。これを最尤推定法と呼び、尤度、すなわちパラメータが与えられたときにデータが生成される確率を最大にするパラメータの推定を行う(Fisher, 1922)。最大尤度の計算に関しては、尤度ではなく、対数尤度の最大化とし

ても一般性を失わないため(対数関数は単調増加関数であるため)、計算の利便性のため対数尤度を用いる。

$$L = \log P\left(\sum_{i=1}^{N} X_i = M\right)$$

$$= \log \frac{(N+M)!}{N!M!} + M \log p + (N-M)\log(1-p)$$

L を最大にする p の必要条件は、偏微分に寄って与えられる：$\frac{\partial L}{\partial p} = 0$。

従って、

$$\frac{\partial L}{\partial p} = M\frac{1}{p} - (N-M)\frac{1}{1-p} = 0$$

$$p = \frac{M}{N}$$

前述のように、N=20, M=14 という実験結果の場合、p=0.7 をパラメータとする二項分布が最も高い確率で実験結果を生成することになる。これは、われわれの日常的に行っている直感的な計算と一致している。つまり、20 回中 14 回起こる事象を観察した場合、われわれの多くは、おおよそ 70%の確率でこの現象は起こる、と直感的に推測するかもしれない。これを形式的に言い直すと、「われわれの 2 択現象に関する仮説生成は、二項分布をモデルとした最尤推定によって行われる」と言える。ただし、ここで挙げた最尤推定法は、モデルパラメータの一つの方法に過ぎ無いことに注意してほしい。他にも、尤度だけではなく事前確率分布を考慮するベイズ統計や、尤度の最大化による一つのモデルの選択ではなく、重み付けにより複数のモデル群として表現する方法、また以下に紹介する最小二乗法などがある。

より複雑なモデル

ここまでは、統計的モデルの考え方に焦点を当て、比較的単純な二項分布をモデルとして扱ってきた。当然ながら、認知過程が二項分布のように単純なモデルだけで説明できるわけではない。これからの解説では、より複雑なモデルとして多変量解析を取り上げる。ただし、モデル自体が複雑になろうとも、根底にある基本的な考え方が同じである事に注意してほしい。

データ解析を行うとき、その分析手法のモデルを意識しながら用いる場合はあまり多くないかもしれない。しかし、よく用いられる分散分析にも因子分析にもモデルがあり、分析者はそれらを利用するときにはモデルを当てはめていることになる（モデルが無く、アルゴリズムのみの分析手法もある）。

多変量解析におけるモデルとはデータをパラメータの関数 f によって近似する数式

$$\text{データ} \cong f(\text{パラメータ})$$

を指す。データとはもちろん観察（実験、調査）によって得られた数値である。パラメータとはデータを特徴付ける数値であり解析前には不明である。データ解析における目的はデータを最もよく近似するパラメータを推定することにある。

ここからは、データ解析において用いられる多変量解析手法が、データを近似するモデルとして表現されることを意識しつつ、分析手法としてではなく心的現象の表現モデルとして解釈していく。

多変量解析と認知機能

認知機能の役割には、入力されるデータに何らかの処理を行い、パターンの発見、結果の予測、情報の縮約などの目的を達成することなどがある。このような心的活動と同様の目的を持った多変量解析手法が多く存在する。例をいくつか挙げると、概念識別は判別分析に、概念形成はクラスタリングに相当し、類似性判断は多次元尺度法と関連が深い。

概念識別、判別分析はまずクラスラベルを伴った入力データから、クラス分類法を学習する。そしてクラスラベルがつけられていない未知のデータに対して適用し、そのクラスを予測することが目的である。概念形成、クラスタリングはクラスラベルがない入力データを少数の群に分類し、データパターンを発見することが目的である。多次元尺度法は（非）類似性データから対象を低次元の心的空間に布置することが目的であり、類似性判断における基準を見いだすことができる。

このように人間の心的活動は多変量解析の枠組みでモデル化できる部分が少なからずあ

る。人間は受けた刺激（入力）に統計的情報処理を行い、反応（出力）を返す、「直感的統計学者」(Peterson & Beach, 1967) であるといえるかもしれない。もちろん、人間が統計的分析をそのまま実行しているわけではないと考えられるが、1 つのモデルとしてとらえ、人間がどのような認知的活動を行っているかを考えてく手助けにはなる。

以下では、例として *K*-means クラスタリング (MacQueen, 1967) を取り上げ、モデル、パラメータ推定、アルゴリズムについて説明を行う。

分類モデルの考え方

今、イヌ、サル、キジ、…といった動物たちを A、B、C の 3 つのカテゴリに分類することを考えよう。動物たちは「吠える」、「体毛」、「狩猟を好むか」、…といった観点から数値的に評価されているとする。イヌがどれくらい吠えるかの評価値を *x*(イヌ, 吠える)で表すとする。動物評価値がデータとして手元に有り、このデータを近似するモデルを考える。モデルとはパラメータの関数として表現されるものであったが、どのようなパラメータが必要だろうか。

分類課題の目的はどの動物がどのカテゴリに所属するかという出力を得ることである。すなわち所属を表現するパラメータが必要になる。所属するか否かは 2 値であるので 0 と 1 で表現し、メンバシップと呼ぶことにする。たとえば、イヌが A に所属するということを少し形式的に書くと、「*u*(イヌ, A) = 1」となる。このときイヌは B には所属しないので「*u*(イヌ, B) = 0」、同様に「*u*(イヌ, C) = 0」である。

メンバシップパラメータだけでは動物評定データを近似することは難しいので、他のパラメータを追加する。動物をカテゴリ A、B、C に分類するのだからそれぞれの代表点を用いれば情報が要約されて都合がよいかもしれない。カテゴリ A の「吠える」についての代表点を形式的に書けば「*c*(A, 吠える)」となる。代表点の値をセントロイドと呼ぶことにする。

これらのパラメータの組み合わせによって個体の評定値を近似するモデルとして

$$x(\text{イヌ, 吠える}) \cong u(\text{イヌ, A}) \times c(\text{A, 吠える}) + u(\text{イヌ, B}) \times c(\text{B, 吠える}) + u(\text{イヌ, C}) \times c(\text{C, 吠える})$$

が考えられる。これは評定値がメンバシップとセントロイドの積の合計として近似されるモデルである。

たとえばイヌがカテゴリ A に所属するならば *u*(イヌ, A) = 1、 *u*(イヌ, B) = 0、 *u*(イヌ, C) = 0 で表現されるので、カテゴリ B と C の値はキャンセルされカテゴリ A の値だけが残る。すなわち、カテゴリ A に所属するイヌの評定値 *x*(イヌ, 吠える) はカテゴリ A の代表値 *c*(A, 吠える)によって近似されるモデルと考えられる。

モデルとパラメータ

以上の例を一般化して記述する。分類課題で想定する入力データの形式は *n* 個体×*p* 変数の行列 $\mathbf{X} = [\mathbf{x}_1', \ldots \mathbf{x}_i', \ldots \mathbf{x}_n']'$ である。ここで、データ $\mathbf{x}_i$ をメンバシップ **U**（*n* 個体×*K* カテゴリ）とセントロイド **C**（*K* カテゴリ×*p* 変数）の積によって近似するモデル

$$\mathbf{x}_i \cong \sum_k u_{ik}\mathbf{c}_k \ldots (1)$$

を考える。メンバシップ $\mathbf{U} = \{u_{ik}\}$は、ある個体がどのカテゴリに所属するかを表現する値で、個体 *i* カテゴリ *k* に所属するとき $u_{ik} = 1$ となり、所属しない場合 $u_{ik} = 0$ となる。メンバシップには制約があり、個体はいずれか 1 つだけのカテゴリに所属しなければならない。セントロイド $\mathbf{C} = [\mathbf{c}_1', \ldots \mathbf{c}_k', \ldots \mathbf{c}_K']'$は、カテゴリの中心を表現する座標ベクトルの集合である。メンバシップとセントロイドは両方とも推定対象のパラメータである。カテゴリ数 *K* はハイパーパラメータもしくはチューニングパラメータなどと呼ばれ、あらかじめ決めておかなければならない。

最小二乗法：モデルパラメータ推定

代表的なパラメータ推定法として先に紹介した最尤推定法のほかに最小二乗法がある。最小二乗法とは、誤差の 2 乗和を最小にするパラメータを求める方法である。(1)に誤差 **e** を書き加えると

$$\mathbf{x}_i = \sum_k u_{ik}\mathbf{c}_k + \mathbf{e}_i \ldots (2)$$

となり近似ではなくなり、両辺が等号で結ばれる。最小化の対象となる誤差の 2 乗和をパラメータの関数として表現したものを目的関数と呼び、その値が最小となるパラメータが最適解

と呼ばれる。(2)に対して最小二乗法の目的関数は

$$F(\mathbf{U},\mathbf{C}) = \sum_i \sum_k u_{ik} \|\mathbf{x}_i - \mathbf{c}_k\|^2 \ldots (3)$$

となる。先に記したパラメータ **U** に関する制約の数式表現：

$$\sum_i u_{ik} = 1 \text{ and } u_{ik} \in \{0,1\} \ldots (4)$$

を満たしながら(3)の値を最小にするパラメータ **U** と **C** を求めることが解析の目的となる。以上をまとめると、群内の散らばりが最も小さくなるような群の中心とデータ割り当て方を求めている。

アルゴリズム

アルゴリズムとは目的関数の値を最小にするパラメータを求める手順のことである。(3)には 2 つのパラメータが含まれているため、反復を繰り返しながら最適解を求める、交互最小二乗法と呼ばれる方法を用いる。

具体的にはパラメータ **U** を推定する際にはパラメータ **C** は与えられている値として扱い(3)を最小化する。パラメータ **C** を推定する際にはパラメータ **U** は与えられている値として扱い(3)を最小化する。この 2 つのステップを (3)の値が収束するまで繰り返す。

各ステップの最適解は次のように求まる。メンバシップパラメータ **U** は、

$$\hat{\mathbf{U}} = \{u_{ik}\} = \begin{cases} 1, k = \operatorname{argmin} D(\mathbf{x}_i, \mathbf{c}_k) \\ 0, \text{otherwise.} \end{cases} \ldots (5)$$

ここで $D(\mathbf{x}_i, \mathbf{c}_j) = \|\mathbf{x}_i - \mathbf{c}_k\|^2$ であり、個体 *i* とセントロイド *k* との距離を表す。個体 *i* は最も距離が近いセントロイド *k* に割り振られる。セントロイドパラメータ **C** は

$$\hat{\mathbf{c}}_k = \frac{\Sigma_i u_{ik} \mathbf{x}_i}{\Sigma_i u_{ik}} \ldots (6)$$

となり、カテゴリ *k* に所属する個体の平均を計算することにより求まる。

アルゴリズム全体は次のようになる。

Step1: **U**、**C** に初期値を設定する。
Step2: (5)を用いて **U** を求める。
Step3: (6)を用いて **C** を求める。
Step4: (3)の値の変化量があらかじめ決めた値よりも少なければ終了。そうでなければ Step2 へ戻る。

パラメータの解釈とモデルの修正

K-means クラスタリングを分類課題の認知モデルとして見るならば、推定されたパラメータを心理学的に解釈することが必要になる。今回のモデルに含まれるパラメータはメンバシップ **U** とセントロイド **C** であった。メンバシップはある個体がどのカテゴリに所属するかを表現する数値であったから、人間が行った分類結果に対応する。すなわち個体 a はカテゴリ I に所属し、個体 b はカテゴリ II に所属するといった具合である。セントロイドはカテゴリの中心を表現する座標ベクトルであったから、各カテゴリが心理的空間にプロトタイプとして表象されているという仮定に対応する。

多変量解析のモデルそのままではデータをうまく近似できない場合や、実験などで得られた知見をモデルに組み込みたいという場合はモデルを拡張、修正する必要がある。 たとえば変数への注意をパラメータとして導入するといったことが考えられる。このような認知モデルの改良には多変量解析モデルが基礎となっていても、現象についての知識が必要となる。

おわりに

本稿では、初等的な例題として二項分布を取り上げ、基本的な統計的モデルの考え方を紹介し、さらに、より複雑なモデルにおける、パラメータ推定、アルゴリズムという考え方を *K*-means クラスタリングを例として解説した。このほかにも認知機能と対応する統計手法は多くあると考えられる (Ashby, 1992)。

参考文献

Ashby, F. G. (Ed.) (1992). *Multidimensional Models of Perception and Cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics, *Philosophical Transactions of the Royal Society*, A, 222, 309-368.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley*

Symposium, **1**, 281–297.

Nickerson, R. S. (2004). *Cognition and Chance: The Psychology of Probabilistic Reasoning*. Psychology Press.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, **68**, 29-46.