

ベイズ型コウホート・モデル

—標準コウホート表への適用—

統計数理研究所 中 村 隆

(1981年12月 受付)

はじめに

同じ項目について継続的に標本調査を実施し、結果を比較すれば、単独の調査結果からは得られない知見を得ることができる。しかし、調査結果を単に並列するのではなく、継続調査全体を有機的に結びつける視点が必要である。そうすれば、さらに多くの情報を引出すことが可能になる。

そのような視点を提供する分析法の1つがコウホート分析である。

コウホート (cohort) とは——語源的にはローマ時代の軍団の単位を表わすが——主として出生をほぼ同時期にする人口集団のことであり、はじめ人口統計学の分野で死亡率などの変動を分析するために導入された概念である。

この概念を用いたコウホート分析は、近年、世論調査・社会調査データの蓄積とともに、様々な意見、態度、行動現象の時代的变化を研究するために、広く行動科学の分野で使われている。とくに政治学の分野では、早くから投票行動や政党支持の変化についての研究でしばしば利用されてきた。コウホート分析の総説やそれを用いた研究の歴史的流れについては、Glenn [4]、吉田 [12] をみるとよい。

これから議論するコウホート分析は、簡単にいえば、ある集団全体としての意見や態度の分布の時間的变化を、成員全体が受ける時代の影響、各成員の加齢による影響、そして時代や加齢の影響を受けない世代固有の部分、に分離しようとする方法である。各時点での意見の分布はそれぞれ別の横断的調査の結果であっても、それらを一貫したコウホートという視点を導入することによって、継続調査全体を有機的に結びつけながら分析するのである。

ところが、コウホート分析に方法論上の問題があることを、多くの文献——たとえば、安田 ([11], pp. 267–276)、Mason, Winsborough と Poole [6]、Fienberg と Mason [3]——が指摘している。それは、時代・年齢・コウホートの3効果が混交していて、原理的に分離できないという問題である。モデル構築のことばでいう識別問題 (identification problem) の1つである。この問題を克服しないかぎり、コウホート分析は名ばかりで実質がないと言われても仕方がない。

そこで本稿では、この識別問題を克服する新しい方法として、パラメータの漸進的变化の仮定を取込んだベイズ型コウホート・モデルを提案する。

まず第1節では、標準コウホート表について説明する。つづいて第2節では、コウホート分析における識別問題について述べ、その解決策について触れる。第3節では、識別問題に対処するベイズ型コウホート・モデルを定式化し、パラメータの推定法と ABIC を用いたモデル選択について述べる。最後に第4節では、仮設データにモデルを適用することによってモデルの性能を調べ、ついで国民性調査データを解析した例を示す。

1. 標準コウホート表

本稿では、ひとまず分析対象を、取扱いの容易な標準コウホート表に限る。説明のために、6次にわたる国民性調査データを念頭に話をすすめよう（国民性調査については、4.3節を参照）。

Table 1. Standard cohort table showing hypothetical data (percentages) in which all variation is due to cohort effects.

表1. 標準コウホート表
——コウホート効果のみの人工データ(%)——

Age Group	Period (Survey)					
	KS1 1953	KS2 1958	KS3 1963	KS4 1968	KS5 1973	KS6 1978
20—24	27	25	26	21	17	16
25—29	25	27	25	26	21	17
30—34	23	25	27	25	26	21
35—39	21	23	25	27	25	26
40—44	19	21	23	25	27	25
45—49	16	19	21	23	25	27
50—54	13	16	19	21	23	25
55—59	11	13	16	19	21	23
60—64	9	11	13	16	19	21
65—69	8	9	11	13	16	19

標準コウホート表とは、Table 1 のような年齢区分×調査時点の形式の表である。各セルには、ある質問項目に対する正反応の観測比率が入る。

標準コウホート表の特徴は、調査時点の間隔と年齢区分の幅とが一致していることである。たとえば、国民性調査のような5年ごとの調査ならば、年齢区分も5歳刻みでなければならない。こうしておく、母集団の同じコウホートは次の調査時点では1つ上の年齢区分にそっくり移る。Table 1 の(1,1)要素は、1953年の第1次国民性調査の時、20～24歳の人々、すなわち、昭和3年～7年生まれのコウホートに対応するが、このコウホートは表中を南東方向にセルをたどり、1978年の第6次調査の時には、45～49歳の年齢区分にあるわけである。

標準コウホート表に対して、調査間隔と年齢区分が一致していない場合を、一般コウホート表と呼ぶ。

2. コウホート分析の方法論上の問題

2.1 識別問題

説明のために、本節に限って次のようなモデルを考える。次節以下で述べるモデルもこのモデルの変形版である。

標準コウホート表の母集団における各セルの正反応確率を p_{ij} とし、これを

$$(2.1) \quad p_{ij} = \mu + \mu_i^A + \mu_j^P + \mu_k^C, \\ i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K,$$

と表わすモデルである。ここで、添字 i は年齢に、添字 j は時代に対応する（以下、年齢区分を単に年齢、調査時点を時代と呼ぶ）。 $\mu, \mu_i^A, \mu_j^P, \mu_k^C$ はそれぞれ総平均効果、年齢効果、時代効果、コウホート効果を表わし、

$$(2.2) \quad \sum_{i=1}^J \mu_i^A = \sum_{j=1}^J \mu_j^P = \sum_{k=1}^K \mu_k^C = 0,$$

を満足するものとする。 I, J, K はそれぞれ年齢，時代，コウホートの区分数である。標準コウホート表の性質から， $k=j-i+I, K=I+J-1$ であることを注意しておく。

さて，表記の簡便のために，添字 ij を $l=i+(j-1)I$ に改め， p_{ij} の代りに p_l と表わす。さらに， $L=IJ$ とおく。ベクトル $\mathbf{p}'=(p_1, \dots, p_L)$ を用いれば，(2.1) は (2.2) の制約条件を考慮して，次のように書くことができる：

$$(2.3) \quad \mathbf{p} = D\boldsymbol{\mu}.$$

ここで，

$$(2.4) \quad \boldsymbol{\mu}' = (\mu, \mu_1^A, \dots, \mu_{I-1}^A, \mu_1^P, \dots, \mu_{J-1}^P, \mu_1^C, \dots, \mu_{K-1}^C),$$

$$(2.5) \quad D = \begin{matrix} L \times M \\ \left(\begin{array}{cccccccc} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & -1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & 0 & 0 & 1 & 0 \end{array} \right) \end{matrix}$$

($I=J=3$ の場合)

$$(2.6) \quad M = 1 + (I-1) + (J-1) + (K-1) = 2(I+J) - 3,$$

である。

一般に(2.3)のようなモデルにおいて， $L \geq M$ で $\text{rank}(D) = M$ ならば， L 次元ベクトル \mathbf{p} は M 次元の部分空間に属し， $\boldsymbol{\mu} = (D'D)^{-1}D'\mathbf{p}$ で， $\boldsymbol{\mu}$ と \mathbf{p} は1対1に対応する(ここで， D' は D の転置行列を表わす)。しかし，コウホート・モデルでは， $D = (\mathbf{d}_1^A \dots \mathbf{d}_{I-1}^A \mathbf{d}_1^P \dots \mathbf{d}_{J-1}^P \mathbf{d}_1^C \dots \mathbf{d}_{K-1}^C)$ と列ベクトルに分けたとき，

$$(2.7) \quad \sum_{i=1}^{I-1} i \mathbf{d}_i^A - \sum_{j=1}^{J-1} j \mathbf{d}_j^P - \sum_{k=1}^{K-1} k \mathbf{d}_k^C = -J\mathbf{1},$$

という線形関係があるので， $\text{rank}(D) = M-1$ となっている(ここで， $\mathbf{1}' = (1, \dots, 1)$ である)。したがって， L 次元ベクトル \mathbf{p} は $(M-1)$ 次元の部分空間に属することになり，1つの \mathbf{p} に対して無数の $\boldsymbol{\mu}$ が対応する。データにこのモデルをあてはめ，もっとも適合のよい \mathbf{p} を決めることはできても，その \mathbf{p} から分析目的の $\boldsymbol{\mu}$ を特定することができない。

たとえば，Figure 1 のような年齢・時代・コウホート効果 $\boldsymbol{\mu}$ を与えて(2.3)により \mathbf{p} を生成したとしよう。この \mathbf{p} から， $\mu_2^A = \mu_3^A$ という付加条件で $\boldsymbol{\mu}$ を再現したのが Figure 2， $\mu_2^C = \mu_3^C$ という付加条件で $\boldsymbol{\mu}$ を再現したのが Figure 3 である。(もちろん $\mu_7^C = \mu_8^C$ という付加条件を与えれば，Figure 1 が再現できる。) Figure 1 から Figure 3 までを比べてわかることは，与える付加条件の違いによって，それぞれの効果に関して異なる解釈を下す可能性があることである。

モデルが識別不足の状況にあり，何らかの付加条件を与えなければモデルを特定できないわ

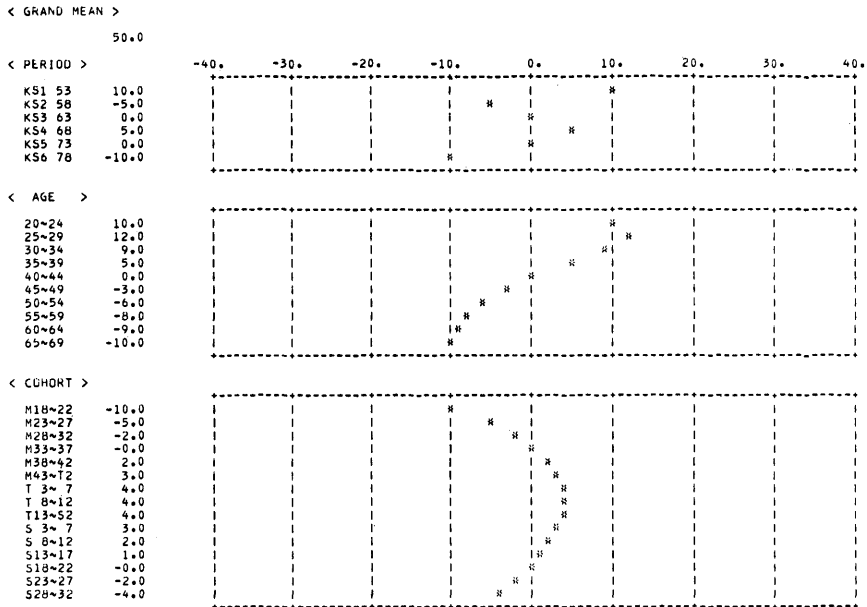


Figure 1. Hypothetical parameter μ from which the data used in Figure 2 and Figure 3 are generated.

図 1. 仮設パラメータ μ (これにより生成したデータを図2と図3で使う).

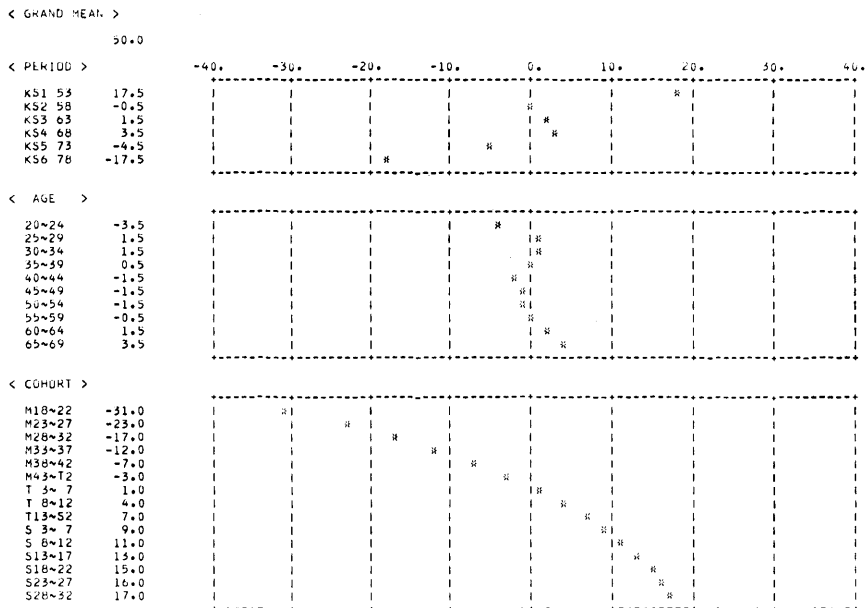


Figure 2. Reconstructed parameter μ for the data generated from that in Figure 1, under the constraint $\mu_2^A = \mu_3^A$.

図 2. $\mu_2^A = \mu_3^A$ という制約で再現したパラメータ μ (図1によるデータについて).

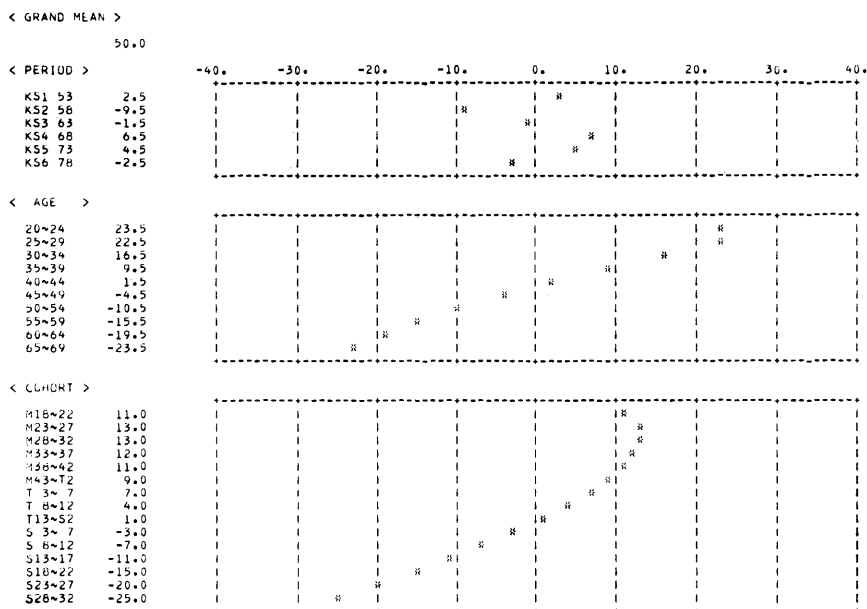


Figure 3. Reconstructed parameter μ for the data generated from that in Figure 1, under the constraint $\mu_2^C = \mu_3^C$.

図 3. $\mu_2^C = \mu_3^C$ という制約で再現したパラメータ μ (図 1 によるデータについて).

けである。では、その付加条件を客観的にどう与えればよいか？ それがコウホート分析における最大の難問である。

2.2 コウホート分析における識別問題の克服法

従来のコウホート分析において識別問題を克服するための一般的やり方は、いくつかの先験的な条件をパラメータに与えてモデルを過剰識別にし、そのような数個のモデルについて、モデルによる予測値と実際の観測値との適合度を考慮してよいモデルを選択する、というものである。

先験的条件としては、たとえば、時代・年齢・コウホート効果のいずれかが存在しないという条件、あるいは年齢効果などが直線的に変化するという条件、または第 1・第 2 コウホート効果が等しくかつ第 $K-1$ ・第 K コウホート効果が等しいという条件などである。

これらの先験的条件のほとんどは、いくつかの隣り合う効果の階差が小さい（またはゼロで等しい）という形式にまとめることができる。しかし、条件を取入れる効果の場所が特定化されているために、広い範囲の異なる条件について比較するのは非常に手間のかかる作業とならざるをえないし、先験的条件の恣意性を免かれぬ。

カテゴリの順序性を考えれば、年齢効果やコウホート効果の隣り合う変化が小さいというパラメータの漸進的变化の条件はそれほど厳しいものではない。これまでのように特定の効果に固定的な条件を入れるのではなく、条件を入れる場所を分散させて広い範囲の条件の場合を比較できればよいわけである。

また、節約原理はなるべく少ないパラメータ数のモデルの方がよいということであるが、これとは別に、なるべくパラメータの変動が小さいモデルがよいという節約原理も考えられる。パラメータの漸進的变化の条件はこのような意味でも識別問題を克服する方策として有力であ

る。

3. ベイズ型コウホート・モデル

3.1 ロジット・コウホート・モデル

2項型データの分析で多く用いられるロジット・モデルを基本に考える。前節(2.1)の p_{ij} を、そのロジット $\ln p_{ij}/(1-p_{ij})$ で置きかえたものが、ロジット・コウホート・モデルである：

$$(3.1) \quad \ln p_{ij}/(1-p_{ij}) = \mu + \mu_i^A + \mu_j^P + \mu_k^C.$$

(2.1)式でもそうであったが、コウホート・モデルでは高次の交互作用がない、ということに注意しよう。たとえば、どの年齢あるいはどのコウホートに対しても時代効果は同じ大きさであり、その事情は年齢、コウホート効果についても同様である。

あるいは見方を変えれば、コウホート効果は時代と年齢の交互作用項であるともいえる。ただし、同じコウホートで同じ値をとるという制約があるわけである。

さて、 $q_i = \ln p_i/(1-p_i)$ とおき、ベクトル $\mathbf{q}' = (q_1, \dots, q_L)$ を用いれば、ロジット・コウホート・モデルは、

$$(3.2) \quad \mathbf{q} = D\boldsymbol{\mu},$$

と書くことができる。 $\boldsymbol{\mu}$ と D は(2.4)、(2.5)でそれぞれ定義されたものと同じである。

このモデルの尤度 $f(\mathbf{n}|\mathbf{p}(\boldsymbol{\mu}))$ は、標準コウホート表の各セルの観測された正反応頻度 n_i が互いに独立に生起確率 p_i の2項分布に従い、 \mathbf{p} が(3.1)のように表わされると仮定すれば、

$$(3.3) \quad f(\mathbf{n}|\mathbf{p}(\boldsymbol{\mu})) = \prod_i \binom{N_i}{n_i} p_i^{n_i} (1-p_i)^{N_i-n_i},$$

である。ここで、 N_i は各セルの標本の大きさであり、 $\mathbf{n}' = (n_1, \dots, n_L)$ である。

通常は最尤法により $\boldsymbol{\mu}$ を求めればよい。しかし、コウホート・モデルでは第2節に述べた理由からそれが困難である。そこで、次に示すようにパラメータの漸進的変化の条件を組入れることにする。

3.2 パラメータの漸進的変化の条件

2.2節で示唆したパラメータの漸進的変化の条件を、次に示すようなパラメータの1次階差の重み付き2乗和が小さくなること、として具体化する。すなわち、適当な重み $1/\sigma_A^2$ 、 $1/\sigma_P^2$ 、 $1/\sigma_C^2$ をとって、

$$(3.4) \quad \frac{1}{\sigma_A^2} \sum_{i=1}^{I-1} (\mu_i^A - \mu_{i+1}^A)^2 + \frac{1}{\sigma_P^2} \sum_{j=1}^{J-1} (\mu_j^P - \mu_{j+1}^P)^2 \\ + \frac{1}{\sigma_C^2} \sum_{k=1}^{K-1} (\mu_k^C - \mu_{k+1}^C)^2,$$

を小さくするという付加条件である。ただし、それは(3.3)の f を大きくすることとのバランスで実現しなければならない。

この(3.4)は、

$$(3.5) \quad (\mathbf{z} - \boldsymbol{\alpha})' S^{-1} (\mathbf{z} - \boldsymbol{\alpha}),$$

で表わすことができる。ただし、 $\mathbf{z}' = (0, \dots, 0)$ は形式だけのために用いてある。また、

$$(3.6) \quad \boldsymbol{\alpha} = D_s \boldsymbol{\mu},$$

$$(3.7) \quad D_s = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 \end{pmatrix}$$

($I=J=3$ の場合)

$$(3.8) \quad S = \begin{pmatrix} \sigma_A^2 I_{I-1} & 0 & 0 \\ 0 & \sigma_P^2 I_{J-1} & 0 \\ 0 & 0 & \sigma_C^2 I_{K-1} \end{pmatrix},$$

である。ここで、 I_N は $N \times N$ の単位行列を表わす。

(3.5)を小さくするということは、(3.5)を正規分布の密度関数の形に書き直した

$$(3.9) \quad \pi(\boldsymbol{\mu} | \sigma_A^2, \sigma_P^2, \sigma_C^2) = (2\pi)^{-(I+J-2)} |S|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{z}-\boldsymbol{\alpha})' S^{-1} (\mathbf{z}-\boldsymbol{\alpha}) \right\},$$

を大きくすることと同じである。こう考えると、コウホート・モデルにパラメータの漸進的変化を取入れる問題は、

$$(3.10) \quad f(n | \boldsymbol{\mu}) \cdot \pi(\boldsymbol{\mu} | \sigma_A^2, \sigma_P^2, \sigma_C^2),$$

の最大化を考えればよいことになる。

$\pi(\boldsymbol{\mu} | \sigma_A^2, \sigma_P^2, \sigma_C^2)$ はパラメータ $\boldsymbol{\mu}$ の密度関数とみなすことができ、ベイズ理論でいう事前分布である。したがって、(3.10)を最大にするように $\boldsymbol{\mu}$ を決めることは、ロジット・コウホート・モデル (3.2) のパラメータ $\boldsymbol{\mu}$ に事前分布 (3.9) を仮定した時の事後分布 (3.10) のモードによって推定値を求めること、と解釈できる。

3.3 パラメータの推定

$\sigma_A^2, \sigma_P^2, \sigma_C^2$ を与えれば、(3.10) の事後分布が決まり、その最大化問題は最尤法に帰着する。一般に、指数分布族の場合、最尤方程式は重み付き最小2乗法を利用して解くことができる (Bradley [2])。

本稿のコウホート・モデルの場合、パラメータ推定の具体的手順は付録 A1 に示すように行なえばよい。

こうして求めた推定パラメータを $\hat{\boldsymbol{\mu}}$ と書くことにする。また、

$$(3.11) \quad \bar{D} = \begin{pmatrix} D \\ D_s \end{pmatrix},$$

$$(3.12) \quad \bar{S} = \begin{pmatrix} W & 0 \\ 0 & S \end{pmatrix},$$

$$(3.13) \quad W = \begin{pmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_L \end{pmatrix}, \quad w_l = 1/N_l p_l (1-p_l),$$

$$(3.14) \quad \mathbf{m}' = (N_1 p_1, \dots, N_L p_L),$$

を定義しておく.

3.4 ABIC によるモデル選択

残された問題は, $\sigma_A^2, \sigma_P^2, \sigma_C^2$ をどう決定するかである. Akaike [1] は, こうした問題を超パラメータ (パラメータの事前分布のパラメータ) の決定問題として定式化し,

$$(3.15) \quad \text{ABIC (A Bayesian Information Criterion)} \\ = -2 \ln \int f \cdot \pi \, d\boldsymbol{\mu},$$

を最小にする超パラメータを採用する方式を提案している. ここで, $\int f \pi \, d\boldsymbol{\mu}$ は周辺尤度である.

ベイズ型コウホート・モデルの場合, (3.15) を最小にする $\sigma_A^2, \sigma_P^2, \sigma_C^2$ を直接解析的に求めるのは困難である. ただし, ある $(\sigma_A^2, \sigma_P^2, \sigma_C^2)$ を与えたときに求まる $\boldsymbol{\mu}$ を使って, その場合の ABIC の値を求めることは比較的容易である.

そのために, まず, $\ln n_i / (N_i - n_i)$ が漸近的に平均 $\ln p_i / (1 - p_i)$, 分散 $\{N_i p_i (1 - p_i)\}^{-1}$ の正規分布に従うことを利用して, (3.15) 中の積分を求める. 定数項を適当に操作すれば,

$$(3.16) \quad \text{ABIC} \\ = 2 \sum_i \{n_i \ln n_i / m_i + (N_i - n_i) \ln (N_i - n_i) / (N_i - m_i)\} \\ + \boldsymbol{\mu}' D_s' S^{-1} D_s \boldsymbol{\mu} + \ln |S| + \ln |D' W^{-1} D + D_s' S^{-1} D_s|,$$

となる (ABIC の導出の詳細については, 付録 B を参照).

実際の場合では, $(\sigma_A^2, \sigma_P^2, \sigma_C^2)$ のいくつかの組合せを考え, それぞれについて, まず 3.3 節の手順で $\boldsymbol{\mu}$ を求め, つぎにその $\boldsymbol{\mu}$ を使って (3.16) の ABIC を評価する. それらの ABIC の中の一番小さい値に対応する $(\sigma_A^2, \sigma_P^2, \sigma_C^2)$ が与える $\boldsymbol{\mu}$ を, よいモデルのパラメータ値として採用することになる.

4. 適用例

4.1 仮設データ

仮設パラメータ $\boldsymbol{\mu}_0' = (\mu_0 \mu_0^A \mu_0^P \mu_0^C)$ を与え, このパラメータから (3.2) 式によって仮設データを生成する. ただし, ここでは標本誤差は加えていない. このデータについて, 提案したベイズ型コウホート・モデルを適用し, 元のパラメータがどのように再現されるかを検討することにしよう.

具体的には, $\boldsymbol{\mu}_0$ として Table 2 の最右列の数値を与えた場合について例示する. 再現のケースとして, 年齢, 時代, コウホートの各効果が単独の場合 (3 ケース), いずれかの効果が欠けている場合 (3 ケース), そして全効果が存在する場合の計 7 ケースを設定した. それぞれのケースについて, いずれも $N_i = N = 100$ ($i = 1, \dots, L$) とおいて \mathbf{n} ($= N\mathbf{p}$) を生成し, 超パラメータ $\sigma_A^2, \sigma_P^2, \sigma_C^2 \in \{\sigma^2/16, \sigma^2/4, \sigma^2, 4\sigma^2, 16\sigma^2\}$ とする 125 通りの組合せについてモデル選択を行なった. ここで, σ^2 は, $\ln n_i / (N_i - n_i)$ の分散 $\{n_i (N_i - n_i) / N_i\}^{-1}$ の相乗平均とした.

結果を Table 2 に示す. また, ケース 5, 7 について図示したのが Figure 4, Figure 5 である.

ケース 1 から 6 では, 推定パラメータはほぼ仮設パラメータを再現している. しかし, 正確

Table 2. An application of Bayesian cohort model to hypothetical data.
 表 2. ベイズ型コウホート・モデルの仮設データへの適用

Case No.		1	2	3	4	5	6	7	μ_0
Hypothetical Parameters	μ_0	μ_0	μ_0	μ_0	μ_0	μ_0	μ_0	μ_0	
		μ_0^A	0	0	μ_0^A	μ_0^A	0	μ_0^A	
		0	μ_0^P	0	μ_0^P	0	μ_0^P	μ_0^P	↓
		0	0	μ_0^C	0	μ_0^C	μ_0^C	μ_0^C	
Selected Hyper-Parameters	σ_A^2/σ^2	4	1/16	1/16	1	4	1/16	1	
	σ_P^2/σ^2	1/16	16	1/16	16	1/16	16	16	
	σ_C^2/σ^2	1/16	1/16	1	1/16	1	1	1	
Estimated Parameters	$\bar{\mu}$	-0.00	-0.00	-0.03	-0.00	0.03	0.02	0.02	μ_0
	μ^A	1.00	0.00	-0.00	0.97	0.96	0.01	0.69	1.0
		1.17	0.00	0.00	1.09	1.16	0.01	0.87	1.2
		0.89	0.00	0.01	0.85	0.89	0.01	0.70	0.9
		0.49	0.00	0.01	0.47	0.50	0.01	0.39	0.5
		0.01	-0.00	0.01	0.02	0.02	0.01	0.01	0.0
		-0.30	-0.00	0.01	-0.29	-0.28	0.00	-0.24	-0.3
		-0.59	-0.00	0.00	-0.56	-0.58	0.00	-0.46	-0.6
		-0.79	-0.00	-0.00	-0.75	-0.77	-0.01	-0.59	-0.8
		-0.89	-0.00	-0.01	-0.86	-0.88	-0.01	-0.65	-0.9
	-0.98	-0.00	-0.02	-0.94	-1.02	-0.03	-0.74	-1.0	
	μ^P	0.00	0.99	-0.01	1.01	-0.00	0.95	1.11	1.0
		0.00	0.00	-0.00	0.02	0.00	0.01	0.12	0.0
		0.00	-0.49	0.00	-0.48	0.01	-0.48	-0.44	-0.5
		-0.00	0.49	0.01	0.48	0.00	0.50	0.47	0.5
		-0.00	-0.00	0.00	-0.02	-0.00	0.01	-0.10	0.0
		-0.00	-0.99	-0.00	-1.01	-0.01	-0.99	-1.17	-1.0
	μ^C	-0.01	0.00	-0.71	-0.06	-0.69	-0.73	-1.16	-1.0
		-0.01	0.00	-0.49	-0.06	-0.50	-0.48	-0.88	-0.5
		-0.01	0.00	-0.23	-0.05	-0.26	-0.22	-0.58	-0.2
		-0.01	0.00	-0.02	-0.05	-0.05	-0.01	-0.31	0.0
		-0.01	0.00	0.16	-0.04	0.14	0.17	-0.05	0.2
		-0.00	0.00	0.27	-0.03	0.25	0.28	0.11	0.3
		-0.00	0.00	0.35	-0.01	0.34	0.36	0.27	0.4
		0.00	0.00	0.36	-0.00	0.35	0.37	0.35	0.4
		0.00	0.00	0.35	0.01	0.35	0.36	0.42	0.4
		0.01	0.00	0.26	0.03	0.27	0.27	0.41	0.3
		0.01	-0.00	0.17	0.04	0.18	0.16	0.39	0.2
		0.01	-0.00	0.06	0.05	0.09	0.06	0.36	0.1
		0.01	-0.00	-0.04	0.05	-0.02	-0.05	0.33	0.0
		0.01	-0.00	-0.20	0.06	-0.17	-0.21	0.23	-0.2
	0.01	-0.01	-0.30	0.05	-0.29	-0.32	0.12	-0.4	
	Smallest ABIC		49.95	48.49	48.39	74.30	72.05	72.67	93.86

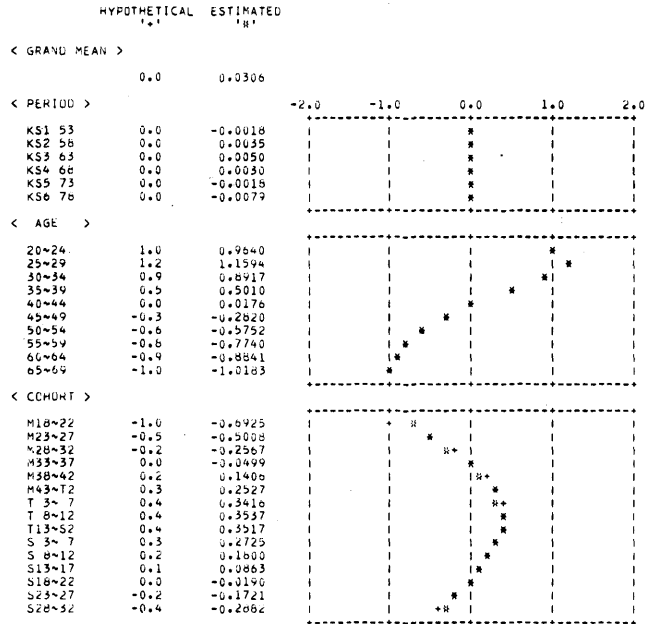


Figure 4. Hypothetical and estimated parameters for Case-5 in Table 2.
 図 4. 仮設パラメータと推定パラメータ (表2のケース5).

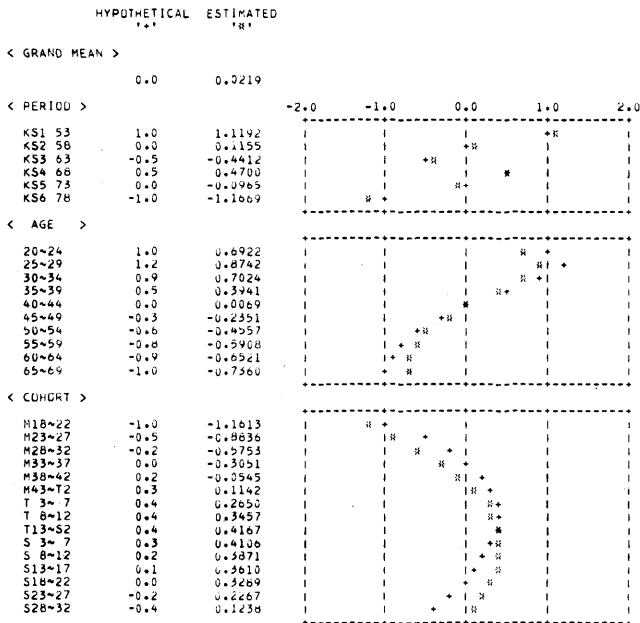


Figure 5. Hypothetical and estimated parameters for Case-7 in Table 2.
 図 5. 仮設パラメータと推定パラメータ (表2のケース7).

に再現しないのは、パラメータの漸進的变化の条件がパラメータの推定においてバイアスをかける方向に働いているためである。(ABICによる適当な超パラメータの選択は、推定パラメータの分散と偏りのバランスを考えることである。)

たとえばケース5のように時代効果が存在しない場合、 $\sigma_p^2 \rightarrow 0$, $\sigma_A^2, \sigma_c^2 \rightarrow \infty$ とすれば、厳密に $\mu_A = 0$ という制約を置いたコウホート・モデルに対応し、正確に仮設データを再現するはずである。

ケース3, 5, 6はいずれもコウホート効果が存在するが、両端のコウホート効果の再現はあまりよくない。これは、両端のコウホート効果の対応するコウホート表中のセルがそれぞれただ1つであり、データへの適合よりも1次階差の2乗和への寄与を小さくした方が全体としての最大化に有利だからである。

年齢・時代・コウホートの3効果が存在するケース7では、推定パラメータは仮設パラメータの変化をおおよそ跡づけている。しかし、その再現性は必ずしもよくない。

その理由の1つは、 σ_A^2 等の組合せが限定されているためと考えられる。たとえば、ケース7の超パラメータの組合せよりも $\sigma_A^2/\sigma^2=2$, $\sigma_p^2/\sigma^2=16$, $\sigma_c^2/\sigma^2=1$ という超パラメータの組合せに対しては $ABIC=93.70$ と減少し、推定パラメータの再現もケース7よりもよくなる。

もう1つの本質的理由は、設定パラメータの値は任意に与えたものであり、パラメータの変化がゆるやかという条件を意識していない、ということである。したがって、 σ_A^2 等の組合せを細かく追求しても、正確に仮設パラメータを再現できるとは限らない。

むしろ、任意に設定したパラメータにかなり近い形を再現することの方に留意すべきであると思う。漸進的变化の条件によって、データから、研究者が自然に受け入れることのできるパラメータの変化を抽出できる、といえよう。

最後に、選択された超パラメータについてみてみよう。仮設パラメータにおいて存在しないと仮定された効果に対応する超パラメータの値はいずれも $\sigma^2/16$ になっている。変化の小さい効果に対しては、小さい値の超パラメータを選んでも、最適化の基準に対して不利にならないからである。このことから、選択された超パラメータの大きさが、効果の大小の目安となることが示唆される。

4.2 国民性調査データ

国民性調査は、1953年(昭和28年)に第1次調査を実施して以来、5年ごとに継続して調査が行なわれている。最新の調査は、1978年の第6次調査である。30年以上にもわたり、計画的にはほぼ同一の項目について継続している全国的な意識調査は、世界にも例をみない。(国民性調査の詳細については、国民性調査委員会 [9], [10] を参照。)コウホート分析の格好の適用対象である。

国民性調査の調査間隔は5年であるから、年齢区分を5歳刻みにとって、ある項目の正反応比率を年齢×時代の表に整理すれば、標準コウホート表が得られる。この標準コウホート表に対して、本稿のベイズ型コウホート・モデルを適用した。ただし、国民性調査の標本抽出法が(3.3)式の仮定とは一致しないので、あくまで形式的な適用である。しかし、この不一致が推定パラメータの形状に本質的な違いを与えるとは考えていない。それぞれの標本抽出法に対応するモデルも考えるべきだが、本稿のような簡単な標本抽出法に基づくモデルであっても、そこから得られる分析結果がデータ解釈に対して与える利得は大きい、と思う。

国民性データのコウホート分析は、長く続けられている項目について、あらかじめ男と女の標本に分けて行なった。ここでは、その中から、「学歴」と「政党支持」の結果について示すことにする(その他の項目の分析結果については、鈴木・中村・田中 [7] をみよ)

[1] 学 歴

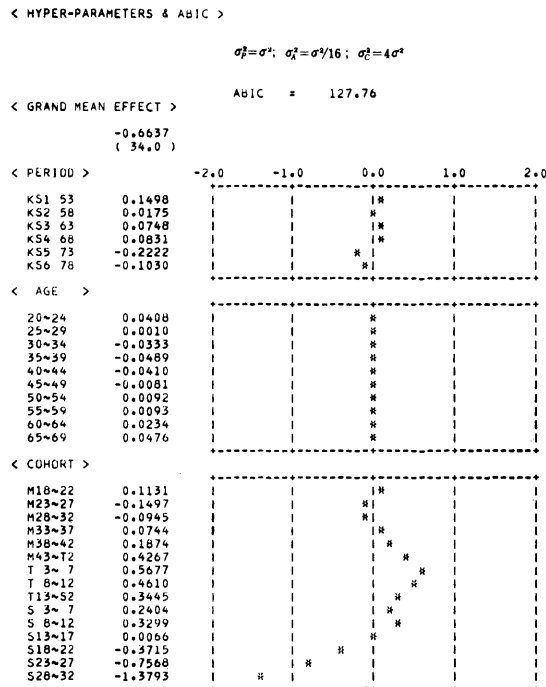
学歴は20代前半でほとんど決定し、以後変わらないはずのものであるから、コウホート効果のみの変数であると予想できる(ただし、学歴によって死亡率に差がある場合、また学歴によって調査での捕捉率が異なる場合は話が別である)。この意味で学歴は、実際のデータに対してコウホート・モデルがうまく働くかどうかを判断するための材料を提供してくれる。国民性調査データの、〈中学卒〉、〈高校卒〉、〈大学卒〉の標準コウホート表にそれぞれベイズ型コウホート・モデルをあてはめた分析結果が Figure 6, Figure 7, Figure 8 である(ここでは、男の標本についてのみ示した)。

これらの結果をみると、どの学歴区分でもコウホート効果は大きく、コウホート効果のピークが次々にずれて現われていることがわかる。日本社会の高学歴化は、若いコウホートを順次高学歴化することによって達成されたのである。

コウホート効果に対して、年齢効果はほとんど現われない。また、時代効果も、〈大学卒〉は別にして現われ方は小さい。

時代効果は、実際的には、本来の時代効果のほかに個々の調査の質の違い(標本抽出法、調査法、コーディングの若干の相違に起因する)を反映するとみるべきである。〈大学卒〉の時代効果の変動は、上述のような意味の違いを反映していると考えられるが、詳しくは不明である。

以上、1つの例にすぎないが、国民性調査の学歴変数の分析結果から、実際のデータについても提案モデルがうまく働くということがいえる。



DATA SOURCE: Surveys on the Japanese National Characters.
Figure 6. Estimated effects for educational attainment: for males for junior highschool completion, based on Bayesian cohort model.

図 6. 学歴〈中学卒〉のベイズ型モデルによるコウホート分析(男)。

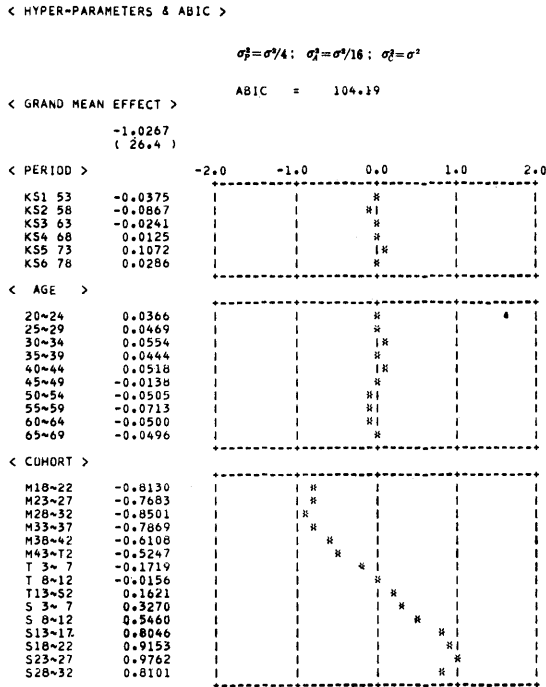


Figure 7. Estimated effects for educational attainment: for males for highschool completion, based on Bayesian cohort model.

図 7. 学歴〈高校卒〉のベイズ型モデルによるコウホート分析(男).

DATA SOURCE: Surveys on the Japanese National Characters.

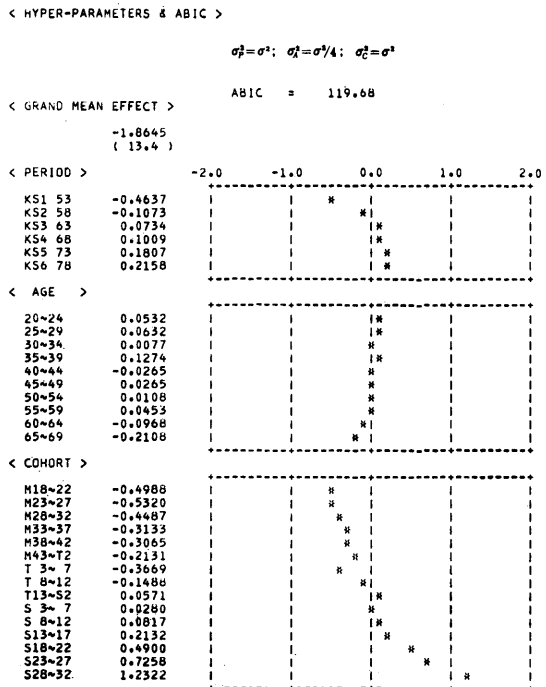


Figure 8. Estimated effects for educational attainment: for males for college completion, based on Bayesian cohort model.

図 8. 学歴〈大学卒〉のベイズ型モデルによるコウホート分析(男).

DATA SOURCE: Surveys on the Japanese National Characters.

[2] 政党支持

政治意識の領域では、政治的態度の保守化が年齢を重ねることによるのか、それとも世代差の反映なのか、という点が争われてきた。コウホート分析における識別問題のために、これまでは結着をつけるのが不可能だったわけである。そこで、本稿のベイズ型コウホート・モデルを国民性調査の政党支持変数に用いることによって、この争点について考えてみよう。

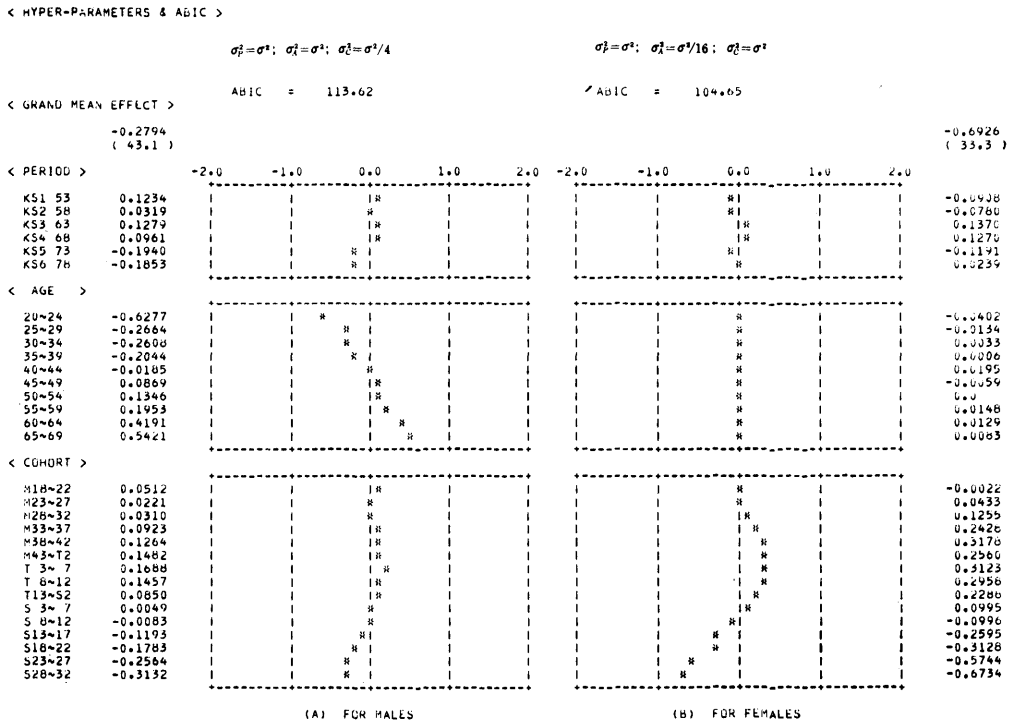
Figure 9 と Figure 10 はそれぞれ〈自民党支持〉と〈社会党支持〉の分析結果である。これらの結果をみてわかることだが、政党支持が年齢だけ、あるいはコウホートだけで決まるといった単純な構造はしていない。

〈自民党支持〉と〈社会党支持〉をそれぞれ男女別に分析した結果からいえることは、まず第1に、いずれの場合もコウホート効果は確かに存在することである。ただし、男の〈自民党支持〉のコウホート効果は小さい。

第2に、年齢効果は男に見られるのに、女には見られないことである。これは〈自民党支持〉でも〈社会党支持〉でも同じである。ただし、男の〈社会党支持〉の年齢効果は20代前半がやや大きいだけである。しかし、男の〈自民党支持〉の年齢効果はめざましい。男の場合、年齢を重ねるにつれ社会的地位が上昇するためであろう。

第3に、政党支持の変化は年齢・コウホートによる以外に時代的趨勢による部分も大きい、ということである。とくに、〈社会党支持〉の時代効果の変化は男女ともに大きい。

このように複雑な諸効果の構造を、従来の方法で抽出するのは困難であった。ここでは国民



DATA SOURCE: Surveys on the Japanese National Characters.
Figure 9. Estimated effects for party identification: for Liberal Democratic Party, based on Bayesian cohort model.

図 9. 政党支持〈自民党〉のベイズ型モデルによるコウホート分析。

< HYPER-PARAMETERS & ABIC >

$$\sigma_1^2 = \sigma^2; \sigma_2^2 = \sigma^2/4; \sigma_3^2 = \sigma^2$$

$$\sigma_1^2 = 4\sigma^2; \sigma_2^2 = \sigma^2/16; \sigma_3^2 = \sigma^2$$

< GRAND MEAN EFFECT >

ABIC = 108.97

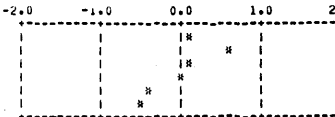
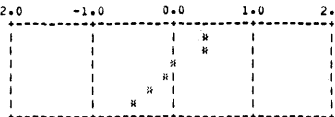
ABIC = 110.02

< PERIOD >

-1.5470
(17.6)

-1.7449
(14.9)

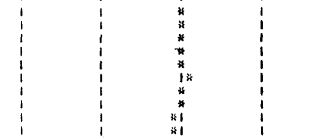
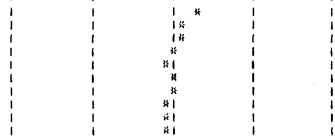
K51 53 0.3538
K52 58 0.4455
K53 63 0.0430
K54 68 -0.0684
K55 73 -0.2545
K56 78 -0.5194



0.1431
0.6112
0.0924
0.1115
-0.3605
-0.4977

< AGE >

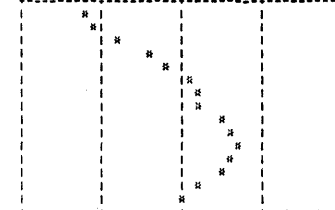
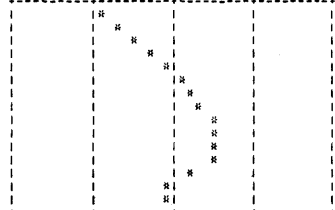
20~24 0.2864
25~29 0.1332
30~34 0.0703
35~39 -0.0457
40~44 -0.0530
45~49 -0.0126
50~54 -0.0285
55~59 -0.0825
60~64 -0.1347
65~69 -0.1351



-0.0066
0.0151
0.0399
0.0235
0.0362
0.0504
0.0197
-0.0234
-0.0660
-0.0688

< COHORT >

M18~22 -0.9365
M23~27 -0.6903
M28~32 -0.5489
M33~37 -0.3352
M38~42 -0.1074
M43~47 0.0684
T 3~ 7 0.2366
T 8~12 0.3241
T13~17 0.4625
S 3~ 7 0.5497
S 8~12 0.4525
S13~17 0.5029
S18~22 0.2033
S23~27 -0.0634
S28~32 -0.1182



-1.1933
-1.1087
-0.7603
-0.4036
-0.1790
0.1440
0.1757
0.2017
0.5356
0.6015
0.7166
0.6496
0.5085
0.1573
-0.0459

(A) FOR MALES

(B) FOR FEMALES

DATA SOURCE: Surveys on the Japanese National Characters.

Figure 10. Estimated effects for party identification: for Japan Socialist Party, based on Bayesian cohort model.

図 10. 政党支持（社会党）のベイズ型モデルによるコウホート分析。

性調査の政党支持変数についての分析であって、出所が別のデータについても分析する必要があるが、その場合も本稿のモデルを用いて容易に分析を実行することができるのである。

おわりに

ここまで、コウホート分析における識別問題を克服して、年齢・時代・コウホート効果の分離を可能にする新しいベイズ型コウホート・モデルについて述べた。このモデルは、パラメータの漸進的变化という比較的ゆるやかな条件を取込み、ABICの最小化という基準によって最適モデルを選択している。

パラメータの漸進的变化という条件は、コウホート分析を可能にするというだけでなく、欠測年のあるデータの分析も可能にしてくれる。

たとえば国民性調査には第2次調査だけ実施されていない項目がいくつかある。このような途中に欠測年のあるデータの分析も、本稿のモデルをそのまま用いることができる。欠測年に対応するセルがないものとして推定するだけでよい。パラメータの漸進的变化の条件があるため、欠測年に対応するパラメータが自然に推定できる。完全に揃っているデータについて、途中の1回分の調査を除外して実験してみたが、パラメータの形状には大きな変化が現われなかった。その意味でも、提案コウホート・モデルは柔軟にできているといえる。

ところで、本稿の提案モデルのひとつの問題点はパラメータの漸進的变化という仮定の妥当性であろう。この条件の考え方は、データに対して同程度に適合するならば、無理に大きな変

動をするパラメータよりも変動の小さいパラメータの方がその解釈も自然で容易である、ということである。もし大きな変動をするパラメータの中でデータへの適合が極めてよいものがあるれば、それは対応する超パラメータを大きくすることによって選択モデルの候補にすることができる。しかし、あくまでも適合度の改善の程度とパラメータの無理のない解釈とのバランスの上でモデル選択を考えなければならないのである。パラメータの漸進的変化の条件は、ABICと組み合わせることによって非常に多くの場面で妥当性をもつものとする。

最後に、モデルの拡張の方向について整理しておこう。

i) 一般コウホート表の分析モデル。

本稿で述べたように標準コウホート表に対するモデルは可能となった。既存のデータについて年齢変数の再コード化などによって標準コウホート表が作れるならば、そうするのがよい。また、今後企画する調査については、将来の分析のために年齢は細かい区分で記録しておくべきである。

一方、これまで蓄積されている調査データは標準コウホート表の形式を意識していないわけであるから、調査間隔も一定でなく、年齢区分も粗い形でしか残されていないことが多い。過去のデータの活用という意味で、一般コウホート表の分析モデルの開発は急務である。これについては稿を改めて論ずる予定である。

ii) 多様な標本抽出法への対応。

本稿では、各セル頻度が2項分布に従うと仮定したが、現実のデータは一般に複雑な標本抽出法によって得られる。それらの方法に対応できるモデルを開発する必要がある。

iii) 数量データに対する分析モデル。

これまではセル比率に対するコウホート・モデルであったが、たとえば平均出生児数の変化を分析するためには、各セルに数量が入るモデルを作らなければならない。この場合、正規分布をする不規則性を考慮すれば拡張は容易である。

謝 辞

国民性調査委員会の先達諸氏は、30年にもわたる貴重な調査データの蓄積を通して、コウホート分析のための豊穡な研究土壌を与えて下さった。赤池弘次氏には、強力なデータ解析の枠組を与えていただくとともに、日々多くの示唆と激励のことばを受けた。石黒真木氏には、モデル構築上の貴重なヒントと有益な助言をいただいた。鈴木達三第6研究部長には、日常的な討論からコウホート分析の研究へ導びいていただき、種々の教示・協力・激励を受けた。さらに、レフェリーの方々には適切なコメントをいただいた。ここに記して心より感謝いたします。

付録 A1. パラメータの推定手順

3.3節で言及したパラメータ推定の具体的手順は以下のとおりである。

Step 1: 初期値 $\mu^{(1)}$ を求める。

$m^{(0)}=n$ として(3.14)の関係から p を求め、 $\bar{\Sigma}(W)$ を決めた後、 $\bar{D}'\bar{\Sigma}^{-1}\bar{D}\mu^{(1)}=\bar{D}'m^{(0)}$ より $\mu^{(1)}$ を求める。 $t=1$ として Step 2 へ。

Step 2: 収束判定を行なう。

$q^{(t)}=D\mu^{(t)}$ の関係から $q^{(t)}$ 、そして $p^{(t)}$ を求め、(3.14) から新しい $m^{(t)}$ を決める。

$m^{(t-1)}$ と $m^{(t)}$ より収束判定をし、更新する必要がある Step 3 へ、そうでなければ終

了.

Step 3: μ を更新する.

$$\bar{\Sigma}(\cdot, W), \text{ また } \alpha = D_s \mu^{(t)} \text{ を計算する. ステップ幅 } \Delta \mu = (\bar{D}' \bar{\Sigma}^{-1} \bar{D})^{-1} \bar{D}' \begin{bmatrix} n - m^{(t)} \\ S^{-1} z - S^{-1} \alpha \end{bmatrix}$$

を求め, $\mu^{(t+1)} = \mu^{(t)} + \Delta \mu$ により μ を更新する. (ステップ幅の決め方については付録 A2 を参照). $t+1$ を t として Step 2 へ.

付録 A2. ステップ幅の決め方

一般に指数分布族では, 観測値を \mathbf{x} , パラメータを μ とするとき, 尤度 L を,

$$(A.1) \quad L(\mu | \mathbf{x}) = h(\mathbf{x}) \exp\{\gamma(\mu) + \beta'(\mu) \mathbf{t}(\mathbf{x})\},$$

と書くことができる.

この L について, フィッシャーのスコア・ベクトル \mathbf{s} と情報行列 I は,

$$(A.2) \quad \mathbf{s}(\mu) = \frac{\partial \ln L}{\partial \mu} = \left(\frac{\partial \theta}{\partial \mu} \right) \Sigma^{-1} [\mathbf{t}(\mathbf{x}) - \theta],$$

$$(A.3) \quad I(\mu) = \text{Cov}[\mathbf{s}(\mu)] = \left(\frac{\partial \theta}{\partial \mu} \right) \Sigma^{-1} \left(\frac{\partial \theta}{\partial \mu} \right)',$$

で表わすことができる. ここに, θ と Σ とはそれぞれ $\mathbf{t}(\mathbf{x})$ の期待値と分散である:

$$(A.4) \quad \theta = E[\mathbf{t}(\mathbf{x})],$$

$$(A.5) \quad \Sigma = \text{Cov}[\mathbf{t}(\mathbf{x})].$$

最尤方程式 $\partial \ln L / \partial \mu = \mathbf{0}$ をガウス・ニュートン法で解くこととスコア法で解くことは同等であり, 重み付き最小2乗法をくり返し適用すればよい. その時のステップ幅 $\Delta \mu$ は,

$$(A.6) \quad \Delta \mu = I^{-1}(\mu) \mathbf{s}(\mu),$$

と与えられる (高根 [8], pp. 244-246).

本稿のベイズ型コウホート・モデルでは, L と $f \cdot \pi$ が対応する. したがって, 対数尤度に対応する $\ln f \cdot \pi$ が,

$$(A.7) \quad \begin{aligned} \ln f \cdot \pi &= \sum_t \binom{N_t}{n_t} + \sum_t n_t \ln p_t + \sum_t (N_t - n_t) \ln(1 - p_t) \\ &\quad + \ln\{(2\pi)^{-(I+J-2)} |S|^{-1/2}\} - \frac{1}{2} (\mathbf{z} - \alpha)' S^{-1} (\mathbf{z} - \alpha) \\ &= \sum_t n_t \ln p_t / (1 - p_t) + \mathbf{z}' S^{-1} \alpha \\ &\quad + \sum N_t \ln(1 - p_t) - \frac{1}{2} \alpha' S^{-1} \alpha + \ln h(\mathbf{n}, \mathbf{z}) \\ &= \mathbf{n}' \mathbf{q} + \mathbf{z}' S^{-1} \alpha + \gamma(\mu) + \ln h(\mathbf{n}, \mathbf{z}) \end{aligned}$$

と書けるから,

$$(A.8) \quad \mathbf{t} = \begin{pmatrix} \mathbf{n} \\ S^{-1} \mathbf{z} \end{pmatrix}$$

とおけばよいことがわかる. こうすれば, (A.4), (A.5) はそれぞれ,

$$(A.9) \quad \theta = \begin{pmatrix} \mathbf{m} \\ S^{-1}\mathbf{a} \end{pmatrix},$$

$$(A.10) \quad \bar{\Sigma} = \begin{pmatrix} W^{-1} & 0 \\ 0 & S^{-1} \end{pmatrix} = \bar{\Sigma}^{-1},$$

となる.

さらに, $\partial \mathbf{m} / \partial \mu = D'W^{-1}$, $\partial S^{-1}\mathbf{a} / \partial \mu = D_s'S^{-1}$ と計算できるから,

$$(A.11) \quad \frac{\partial \theta}{\partial \mu} = [D'W^{-1} \quad D_s'S^{-1}] = \bar{D}' \bar{\Sigma}^{-1},$$

となる. 結局,

$$(A.12) \quad \mathbf{s}(\mu) = (\bar{D}' \bar{\Sigma}^{-1}) \bar{\Sigma} \begin{pmatrix} n-m \\ S^{-1}\mathbf{z} - S^{-1}\mathbf{a} \end{pmatrix},$$

$$(A.13) \quad I(\mu) = (\bar{D}' \bar{\Sigma}^{-1}) \bar{\Sigma} (\bar{\Sigma}^{-1} \bar{D}) = \bar{D}' \bar{\Sigma}^{-1} \bar{D},$$

であるから, 求めるステップ幅 $\Delta \mu$ は,

$$(A.14) \quad \Delta \mu = (\bar{D}' \bar{\Sigma}^{-1} \bar{D})^{-1} \bar{D} \begin{pmatrix} n-m \\ S^{-1}\mathbf{z} - S^{-1}\mathbf{a} \end{pmatrix},$$

となる.

付録 B. ABIC の導出

$\ln n_i / (N_i - n_i)$ が漸近的に平均 $\ln p_i / (1 - p_i)$, 分散 $\{N_i p_i (1 - p_i)\}^{-1}$ に従うことを利用すれば, (3.3) の f は,

$$(B.1) \quad f \sim (2\pi)^{-L/2} |W|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{q})' W^{-1} (\mathbf{x} - \mathbf{q}) \right\},$$

と置換えることができる. ただし, $\mathbf{x}' = (x_1, \dots, x_L)$, $x_i = \ln n_i / (N_i - n_i)$.

したがって,

$$\begin{aligned} f \cdot \pi &\sim C_1 \begin{vmatrix} W & 0 \\ 0 & S \end{vmatrix}^{-1/2} \exp \left\{ -\frac{1}{2} \begin{pmatrix} \mathbf{x} - \mathbf{q} \\ \mathbf{z} - \mathbf{a} \end{pmatrix}' \begin{pmatrix} W & 0 \\ 0 & S \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{x} - \mathbf{q} \\ \mathbf{z} - \mathbf{a} \end{pmatrix} \right\} \\ &= C_1 |\bar{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \bar{D}\mu)' \bar{\Sigma}^{-1} (\mathbf{y} - \bar{D}\mu) \right\}, \end{aligned}$$

と表わせる. ここで, $\mathbf{y}' = (\mathbf{x}' \mathbf{z}')$, $C_1 = (2\pi)^{-(L/2+I+J-2)}$ とおいた.

上式に, 最尤推定量 $\hat{\boldsymbol{\rho}}$ を (重み付き最小 2 乗推定量と考えて) 用いれば,

$$(B.2) \quad f \cdot \pi \sim C_1 |\bar{\Sigma}|^{-1/2} \exp \left[-\frac{1}{2} \{ (\mathbf{y} - \bar{D}\hat{\boldsymbol{\rho}})' \bar{\Sigma}^{-1} (\mathbf{y} - \bar{D}\hat{\boldsymbol{\rho}}) + (\bar{D}\hat{\boldsymbol{\mu}} - \bar{D}\hat{\boldsymbol{\rho}})' \bar{\Sigma}^{-1} (\bar{D}\hat{\boldsymbol{\mu}} - \bar{D}\hat{\boldsymbol{\rho}}) \} \right],$$

と分離できるから,

$$(B.3) \quad \int f \cdot \pi d\mu \sim C_1 |\bar{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \bar{D}\hat{\boldsymbol{\rho}})' \bar{\Sigma}^{-1} (\mathbf{y} - \bar{D}\hat{\boldsymbol{\rho}}) \right\} \\ \times \frac{1}{C_2} |(\bar{D}' \bar{\Sigma}^{-1} \bar{D})^{-1}|^{1/2}$$

$$\begin{aligned} & \times \int C_2 |(\bar{D}'\bar{\Sigma}^{-1}\bar{D})^{-1}|^{-1/2} \exp\left[-\frac{1}{2}(\mu-\boldsymbol{\rho})'\bar{D}'\bar{\Sigma}^{-1}\bar{D}(\mu-\boldsymbol{\rho})\right] d\mu \\ & = C_3 |\bar{\Sigma}|^{-1/2} |\bar{D}'\bar{\Sigma}^{-1}\bar{D}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y}-\bar{D}\boldsymbol{\rho})'\bar{\Sigma}^{-1}(\mathbf{y}-\bar{D}\boldsymbol{\rho})\right], \end{aligned}$$

となる。ここで、 $C_2=(2\pi)^{-M/2}$, $C_3=C_1/C_2$.

これより、

$$\begin{aligned} \text{(B.4)} \quad \text{ABIC} &= -2 \ln \int f \pi d\mu \\ &\sim \ln |\bar{\Sigma}| + \ln |\bar{D}'\bar{\Sigma}^{-1}\bar{D}| \\ &\quad + (\mathbf{y}-\bar{D}\boldsymbol{\rho})'\bar{\Sigma}^{-1}(\mathbf{y}-\bar{D}\boldsymbol{\rho}) + \ln C_3 \\ &= \ln |W| + \ln |S| + \ln |D'W^{-1}D + D_s'S^{-1}D_s| \\ &\quad + (\mathbf{x}-D\boldsymbol{\rho})'W^{-1}(\mathbf{x}-D\boldsymbol{\rho}) \\ &\quad + (\mathbf{z}-D_s\boldsymbol{\rho})'S^{-1}(\mathbf{z}-D_s\boldsymbol{\rho}) + \ln C_3, \end{aligned}$$

である。

ところで、(B.1)の対応において、 $p_i=m_i/N_i$, $\mathbf{q}=D\boldsymbol{\rho}$ とし、両辺の自然対数をとれば、

$$\begin{aligned} \text{(B.5)} \quad & \sum n_i \ln m_i/n_i + \sum (N_i-n_i) \ln (N_i-m_i)/(N_i-n_i) + C_4 \\ & \sim -\frac{1}{2} \ln |W| - \frac{1}{2} (\mathbf{x}-D\boldsymbol{\rho})'W^{-1}(\mathbf{x}-D\boldsymbol{\rho}) \end{aligned}$$

という関係がある (C_4 は適当な定数項)。

これを (B.4) に用いれば、

$$\begin{aligned} \text{(B.6)} \quad \text{ABIC} &\sim 2 \sum n_i \ln n_i/m_i + 2 \sum (N_i-n_i) \ln (N_i-n_i)/(N_i-m_i) \\ &\quad + \boldsymbol{\rho}'D_s'S^{-1}D_s\boldsymbol{\rho} \\ &\quad + \ln |S| + \ln |D'W^{-1}D + D_s'S^{-1}D_s| \\ &\quad + \ln C_3 - \frac{1}{2} C_4, \end{aligned}$$

となる。ここに、 $\mathbf{z}'=(0, \dots, 0)$ を用いた。

ABIC の比較に際して、定数項は共通だから、結局 (3.16) の ABIC を評価すればよい。

参 考 文 献

- [1] Akaike, H. (1980). Likelihood and the Bayes procedure, *Bayesian Statistics* (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith), Valencia, University Press.
- [2] Bradley, E.L. (1973). The equivalence of maximum likelihood and weighted least squares estimates in the exponential family, *J. Amer. Statist. Ass.*, **68**, 199-200.
- [3] Fienberg, S.E. and Mason, W.M. (1979). Identification and estimation of age-period-cohort models in the analysis of discrete archival data, *Sociological Methodology 1979* (ed. K.F. Schuessler), Jossey-Bass.
- [4] Glenn, N.D. (1977). *Cohort Analysis*, Sage Publications.
- [5] Ishiguro, M. and Akaike, H. (1981). A Bayesian approach to the trading-day adjustment of monthly data, *Time Series Analysis: Proceedings of the International Conference held at Houston, August 1980* (eds. O.D. Anderson and M.R. Perryman), Amsterdam, North-Holland.
- [6] Mason, W.M., Winsborough, H.H. and Poole, W.K. (1973). Some methodological issues in cohort analysis of archival data, *American Sociological Review*, **38**, 242-258.
- [7] 鈴木達三, 中村 隆, 田中さえ子 (1982). 継続調査の二次分析, 統計数理研究所研究リポート No.

56.

- [8] 高根芳雄 (1980). 多次元尺度法, 東京大学出版会.
- [9] 統計数理研究所国民性調査委員会 (1975). 第3日本人の国民性, 至誠堂.
- [10] 統計数理研究所国民性調査委員会 (1982). 第4日本人の国民性, 出光書店.
- [11] 安田三郎 (1969). 社会統計学 (旧版), 丸善.
- [12] 吉田 潤 (1981). 発達差・世代差・時代差～コーホート分析をめぐって～, NHK 文研月報, **31** (No. 6).

A Bayesian Cohort Model for Standard Cohort Table Analysis

Takashi Nakamura

(The Institute of Statistical Mathematics)

A new Bayesian cohort model which resolves the identification problem in standard cohort table analysis is proposed in this paper.

Many authors have pointed out that there exists an identification problem in the cohort analysis, and that the age, period, and cohort effects cannot be separated. However, using a quasi-Bayesian framework of statistical inference introduced by Akaike (1980, Likelihood and the Bayes procedure, *Bayesian Statistics* (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith), Valencia, University Press), it is possible to construct a Bayesian cohort model which overcomes the identification problem, on the assumption that the successive parameters should change gradually. The model selection scheme based on "Akaike's Bayesian Information Criterion (ABIC)" is also described.

The usefulness of the proposed model is demonstrated by fitting it first to synthetic data, then to data sets from the Japanese National Character Studies.