

博士論文

Thesis for doctorate

年齢・時代区分データ解析のための環境効果モデル

Environment effect model for analyzing (age, period)-tabulated data

平成16年度

2004

華山 宣胤

NOBUTANE HANAYAMA

目次

1.	はじめに	1
1.1.	目的と意義	1
1.1.1.	目的	1
1.1.2.	意義	1
1.2.	年齢・時代区分データとこれまでの解析方法	2
1.2.1.	年齢・時代区分データ	2
1.2.2.	これまでの解析方法 (APCモデル)	3
1.3.	APCモデルの問題点と基本的アイデア	4
1.3.1.	APCモデルの問題点	4
1.3.2.	基本的なアイデア	4
1.4.	3つの提案モデルとその新奇性	5
1.4.1.	3つの提案モデル	5
1.4.2.	AEモデルの新奇性	5
1.4.3.	Simple-two-stage model (STSモデル)の新奇性	6
1.4.4.	超高齢者モデルの新奇性	6
1.5.	実データの解析結果の概要	7
1.5.1.	実データの解析結果を示す目的	7
1.5.2.	AEモデルを用いた女子労働力人口の解析	7
1.5.3.	AEモデルを用いた乳癌死亡率の解析	7
1.5.4.	STSモデルを用いた全癌死亡率の解析	8
1.5.5.	超高齢者統計の解析	8
1.6.	本論文の構成	9
1.6.1.	第2章	9
1.6.2.	第3章	9
1.6.3.	第4章	10
1.6.4.	第5章	10
2.	これまでの研究と本論文の立場	11
2.1.	レクシス・ダイアグラムと死亡・応答統計	11
2.1.1.	レクシス・ダイアグラム上の死亡強度・応答確率	11
2.1.2.	Piecewise constant intensity model (PCIモデル)	12
2.2.	年齢・時代区分データに基づく尤度	13

2.2.1.	死亡データに基づく尤度	13
2.2.2.	応答データに基づく尤度	17
2.3.	レクシス・ダイアグラムとAPCモデル	19
2.3.1.	レクシス・ダイアグラム上の年齢・時代・コホート効果	19
2.3.2.	“出生区分の重複”による識別問題の解釈ともう一つのAPCモデル	21
2.4.	Robertson-Boyle model への批判と本論文の立場	22
2.4.1.	Robertson-Boyle model への批判	22
2.4.2.	本論文の立場	23
3.	Age-environment model (AEモデル)	25
3.1.	背景と概要	25
3.1.1.	背景	25
3.1.2.	本節の概要	26
3.2.	AEモデル	27
3.2.1.	連続関数としてのAEモデル	27
3.2.2.	PCIモデルとしてのAEモデル	28
3.3.	APCモデルとの関係	29
3.3.1.	APCモデルと年齢, 時代, コホート効果の定義	29
3.3.2.	環境効果と時代効果, コホート効果との違い	29
3.3.3.	APCモデルの副モデルとしてのAEモデル	30
3.4.	最低影響年齢を導入したAEモデルとその推定可能関数	32
3.4.1.	最低影響年齢を導入したAEモデル	32
3.4.2.	推定可能関数	33
3.5.	日本の乳癌死亡データの解析	34
3.5.1.	解析結果の概要	34
3.5.2.	パラメータ推定値についての考察	36
3.6.	北欧4ヶ国の乳癌発生データの解析	37
3.6.1.	解析結果の概要	37
3.6.2.	パラメータ推定値についての考察	38
3.7.	女子労働力人口の解析	40
3.7.1.	解析結果の概要	40
3.7.2.	超過分散について	41
3.7.3.	パラメータの推定値についての考察	42
3.8.	結び	44

3.8.1.	結論.....	44
3.8.2.	リマーク (Remark)	45
4.	Simple-two-stage model (STSモデル)	47
4.1.	背景と概要.....	47
4.1.1.	背景.....	47
4.1.2.	概要.....	47
4.2.	STSモデルの導出とその推定可能関数.....	48
4.2.1.	連続死亡強度に対するSTSモデル.....	48
4.2.2.	PCIモデルに対するSTSモデル.....	50
4.3.	尤度関数と識別問題.....	51
4.3.1.	尤度関数	51
4.3.2.	識別問題	52
4.4.	最尤推定量の存在性と一致性.....	54
4.4.1.	尤度方程式.....	54
4.4.2.	最尤推定量の存在性と一致性.....	54
4.5.	全癌死亡データの解析	55
4.5.1.	解析結果の概要.....	55
4.5.2.	パラメータ推定値についての考察.....	58
4.6.	日米癌死亡データ解析結果の考察.....	61
4.6.1.	解析結果の概要.....	61
4.6.2.	パラメータ推定値についての考察.....	64
4.7.	結び.....	69
4.7.1.	結論.....	69
4.7.2.	リマーク (Remark)	69
5.	超高齢者モデル.....	71
5.1.	背景と概要.....	71
5.1.1.	目的.....	71
5.1.2.	背景.....	71
5.2.	一般パレート分布	73
5.2.1.	定義と性質.....	73
5.2.2.	γ が負か零か.....	74
5.2.3.	最尤推定	75
5.3.	年齢・時代区分データとそのモデル.....	77

5.3.1.	年齢・時代区分データ.....	77
5.3.2.	生存者数統計のモデル.....	78
5.3.3.	死亡者数統計のモデル.....	78
5.4.	実データの解析.....	81
5.4.1.	生存者数統計の解析.....	81
5.4.2.	死亡者数統計の解析.....	84
5.5.	結び.....	87
5.5.1.	結論.....	87
5.5.2.	リマーク (Remark).....	87
付録A	(3.3.3-aの証明).....	88
付録B	(3.4.2-aの証明).....	89
付録C	(Proposition 4.3.2 の証明).....	90
付録D	(Property 4.4.2-b の証明).....	92
付録E	(Property 4.4.2-c の証明).....	93
参考文献	96

表一覧

Table 3.5.1-a : 人口100,000あたりの乳癌死亡者数 (日本) .	35
Table 3.5.1-b : 人口年 (person-years at risk) (日本女性, 単位: 100人) .	35
Table 3.5.1-c : AE, AC, AP, APCモデルを日本女性の乳癌死亡データへ当てはめた結果の概要 .	35
Table 3.6.1-a : AE, APCモデルを北欧4ヶ国の女性の乳癌発生データへ当てはめた結果の概要 .	37
Table 3.7.1-a : 女子労働力人口 (日本, 単位: 1000人) .	40
Table 3.7.1-b : 年齢階級別女子人口 (単位: 万人) .	40
Table 3.7.1-c : AE, APC, AP, ACモデルを女子労働力人口データへ当てはめた結果の概要 .	41
Table 3.7.2-a : AEモデルの残差との新旧世代・年齢階級別の平均値 .	42
Table 4.5.1-a : 100,000人あたりの癌死亡者数と人口年 (person-years at risk) (男性, 日本) .	56
Table 4.5.1-b : 100,000人あたりの癌死亡者数と人口年 (person-years at risk) (女性, 日本) .	57
Table 4.5.1-c : STS, APCモデルを日本の癌死亡データへ当てはめた結果の概要 .	57
Table 4.5.2-a : 成長危険強度の推定値 (日本) .	58
Table 4.5.2-b : 環境危険強度の推定値 (日本) .	60
Table 4.6.1-a : 癌死亡者数と対応する人口年 (person-years at risk) (米国) .	62
Table 4.6.1-b : 癌死亡者数と対応する人口年 (person-years at risk) (日本) .	63
Table 4.6.1-c : 日米の癌死亡データへSTS, APCモデルを当てはめた結果の概要 .	63
Table 4.6.2-a : 環境危険強度の推定値 (米国) .	64
Table 4.6.2-b : 環境危険強度の推定値 (日本) .	65
Table 4.6.2-c : 成長危険強度の推定値 (米国) .	67
Table 4.6.2-d : 成長危険強度の推定値 (日本) .	67
Table 5.2.3-a : パラメータの値に依存する一般パレート分布の特徴 .	76
Table 5.4.1-a : 100歳以上生存者数 (男性) .	82
Table 5.4.1-b : 一般パレート分布に基づく多項モデルによるパラメータの推定結果 .	83
Table 5.4.2-a : 84-117歳の死亡者数 (男性) .	85
Table 5.4.2-b : 一般パレート分布に基づくPCIモデルによるパラメータ推定結果 .	86
Table 5.4.2-c : Coles (2001) の関数“gpd.fit()” による推定の結果 .	86

図一覧

Figure 1.2.1-a : 年齢・時代区分データ (死亡データの場合)	3
Figure 1.2.1-b : 年齢・時代区分データ (応答データの場合)	3
Figure 2.1.1-a : レクシス・ダイアグラム	11
Figure 2.1.1-b : レクシス・ダイアグラム (個人応答の場合)	12
Figure 2.1.2-a : レクシス・ダイアグラムとPiecewise Constant Intensity Model	13
Figure 2.2.1-a : PCIモデルを仮定したレクシス・ダイアグラム上の死亡者数	14
Figure 2.2.1-b : PCIモデルを仮定したレクシス・ダイアグラム上の独立でない領域	15
Figure 2.2.1-c : レクシス・ダイアグラム上で表した $\tau_k^{(m)}$	16
Figure 2.2.1-d : x_{ij}^* の人口年 $N_{ij}^{(py)}$ による近似	17
Figure 2.2.2-a : PCIモデルを仮定したレクシス・ダイアグラム上の応答数	18
Figure 2.2.2-b : レクシス・ダイアグラム上の出生時代区分が同じ集団の応答数	19
Figure 2.3.1-a : PCIモデルを仮定したレクシス・ダイアグラム上の年齢効果	20
Figure 2.3.1-b : PCIモデルを仮定したレクシス・ダイアグラム上の時代効果	20
Figure 2.3.1-c : PCIモデルを仮定したレクシス・ダイアグラム上のコホート効果	21
Figure 2.3.2-a : PCIモデルにおける出生時代区分の重複	22
Figure 2.3.2-b : 出生時代区分の重複を回避したコホート効果	22
Figure 2.4.1-a : 出生時代区分の重複を回避したコホート効果を用いたAPCモデルでの制約	23
Figure 2.4.2-a : レクシス・ダイアグラム上の出生時代区分が同じ集団の応答数	24
Figure 3.2.1-a : 連続死亡強度についてのAEモデルの考え方	28
Figure 3.2.2-a : PCIモデルについてのAEモデルの考え方	28
Figure 3.3.2-a : η_{ij} にAEモデルを仮定し, ξ_2 以外の全ての効果を0としたパターン	29
Figure 3.3.2-b : η_{ij} にAPCモデルを仮定し, β_2 以外の全ての効果を0としたパターン	30
Figure 3.3.2-c : η_{ij} にAPCモデルを仮定し, γ_2 以外の全ての効果を0としたパターン	30
Figure 3.4.1-a : $\rho = 2$ の場合に, $\phi = 1$ のAEモデルを η_{ij} に仮定し, ξ_2 以外は全てゼロとした パターン	32
Figure 3.4.1-b : $\rho = 0$ の場合に, $\phi = 1$ のAEモデルを η_{ij} に仮定し, ξ_2 以外は全てゼロとした パターン	33
Figure 3.5.2-a : “[10,15) ~ [15,20) 歳へ加齢する際に受ける環境効果”の推定値 (乳癌死亡, 日 本)	36
Figure 3.5.2-b : “1985-89年と同じ環境での年齢効果”の推定値 (乳癌死亡, 日本)	37
Figure 3.6.2-a : “[10,15)-[15,20) 歳へ加齢する際に受ける環境効果”の推定値 (乳癌発生, 北欧4 ヶ国)	38
Figure 3.6.2-b : “1988-92年と同じ環境での年齢効果”の推定値 (乳癌発生, 北欧4ヶ国)	39
Figure 3.6.2-c : “1888-92年と同じ環境での年齢効果”の推定値 (乳癌発生, 北欧4ヶ国)	39

Figure 3.7.3-a : $\mu = 0$, $\alpha_2 - \alpha_1 = 0$ の制約下での環境効果の推定値 (女子労働力人口). ...	43
Figure 3.7.3-b : $\mu = 0$, $\alpha_2 - \alpha_1 = 0$ の制約下での加齢効果 (年齢効果の1階差分) の推定値 (女子労働力人口).	44
Figure 4.2.1-a : 連続死亡強度についてのSTSモデルの考え方	49
Figure 4.2.2-a : PCIモデルについてのSTSモデルの考え方	51
Figure 4.3.2-a : レクシス・ダイアグラム上で表した a_k と b_k	53
Figure 4.5.2-a : 成長危険強度の推定値 (男性, 日本).	58
Figure 4.5.2-b : 成長危険強度の推定値 (女性, 日本).	59
Figure 4.5.2-c : 環境危険強度の推定値 (男性, 日本).	60
Figure 4.5.2-d : 環境危険強度の推定値 (女性, 日本).	61
Figure 4.6.2-a : 環境危険強度の推定値 (男性, 米国).	65
Figure 4.6.2-b : 環境危険強度の推定値 (女性, 米国).	65
Figure 4.6.2-c : 環境危険強度の推定値 (男性, 日本).	66
Figure 4.6.2-d : 環境危険強度の推定値 (女性, 日本).	66
Figure 4.6.2-e : 成長危険強度の推定値 (男性, 米国).	67
Figure 4.6.2-f : 成長危険強度の推定値 (女性, 米国).	68
Figure 4.6.2-g : 成長危険強度の推定値 (男性, 日本).	68
Figure 4.6.2-h : 成長危険強度の推定値 (女性, 日本).	68
Figure 5.1.2-a : 死亡者数統計から算出した世代別平均余命の推移	72
Figure 5.1.2-b : 世代別死亡者数の推移	72
Figure 5.4.1-a : 生存者数統計から得た死亡者数と死亡者数統計の比較	83
Figure 5.4.1-b : 生存者数統計から算出した世代別平均余命の推移	84

1. はじめに

この章では、本論文の目的と意義を明らかにする。そして、本論文の研究対象である年齢・時代区分データに頻繁に用いられてきた解析方法としてAge-period-cohort model (APCモデル)を紹介し、これまで指摘されているこのモデルの問題点と本論文の基本的アイデアを明らかにする。さらに、3つの提案モデルとその新奇性、実際のデータ解析結果の概要を述べ、最後に論文全体の構成を述べる。

1.1. 目的と意義

1.1.1. 目的

本論文の目的は、死亡率や罹病率、人々の価値観や嗜好に関する質問項目への応答率について長期に渡って年齢区分別に観測されたデータ（これらを年齢・時代区分データと呼ぶ）から、生活環境の効果を抽出するための統計モデルを提案することである。また、提案モデルに含まれるパラメータの識別問題および最尤推定量の存在性も議論する。さらに、提案モデルを用いた実データの解析結果を示し、実際のデータ解析において提案したモデルが有意義な結論を導出できることを示す。

1.1.2. 意義

最新の癌研究では、癌は、何年も前に発癌物質によって癌化の第1段階へ進まされた細胞が、何年も期間（長い場合には数十年）をかけて複数の変異段階を経て悪性腫瘍へ成長することによって引き起こされる病気であると考えられている（Voet, Voet, 1995, pp. 1183-1184 参照）。実際、塩化ビニルの生産工場の労働者が20年後に塩化ビニルが原因の肝悪性血管内被腫にかかったことや、第2次大戦中に小島で毒ガス兵器の生産に従事させられていた人々が、20年後になって肺癌に襲われたことが知られている。つまり、ある時点で癌で死亡したり癌と診断されても、その原因は何年も前の生活環境や体質に起因しているのである。特に、乳癌については、DDTなどを含む環境中から摂取されるエストロゲン様物質（bioavailable estrogen existing in the environment）や有機塩素系化合物への暴露が、発展途上国での乳癌死亡の増加に関わっているとして注目されている（Davis *et al.*, 1993, Wolff, Weston, 1997, 参照）。これらの事実から、癌罹病率や死亡率データの解析から過去の環境中の癌危険因子量を推定することは、今後の癌罹病や癌死亡の趨勢を予測するために重要である。

また、人間の価値観や嗜好は、生まれてからの体験の蓄積に因るところが大きい。例えば、若年の頃より将来社会に出て働くことを前提として教育を受けてきた女性と、家庭に入り育児

と家事へ従事することが普通であると教育されてきた女性とでは、仕事を持つことへの価値観は異なると考えられる。また、寿司ネタとしてトロを好むのは、子供の頃より家庭で洋風で油分の多い食事をするようになってからだと言われている。したがって、価値観や嗜好に関する調査への応答から生活環境の効果を抽出することは、人間の価値観や嗜好を決定した時代背景を知る上で有意義であると考えられる。

さらに、人間の寿命限界と生活環境との関係も注目されている。近年の老化に関する研究分野では、人の細胞は時間とともにフリーラジカル（活性酸素）と呼ばれる破壊分子により酸化され、次第に老化していくという説がある（D. Harman, 1956参照）。これは、抗酸化物質を摂取して活性酸素の働きを抑えることによって寿命を延ばせるという考えの根拠となっている。一方、寿命については、人の細胞には生物時計（バイオリジカル・クロック）がプログラムされていて、細胞が再生できる寿命が設定されているという説もある（L. Hayflick, 1961参照）。これは、どんなに健康を保っても、寿命には限界があるという考えの根拠となっている。そこで、高齢者（例えば90歳以上）の生存者数統計や死亡者統計から寿命限界の有無を議論することは、両説のどちらが正しいかを検証する一つの方法であろう。また、寿命分布に限界がある場合、世代別に限界値を推定しその変化を見ることによって、国民の生活環境の変化をうかがい知ることができるであろう。

このように、様々な分野において、データから環境の効果を抽出し、その有無や趨勢を議論することは有意義である。

1.2. 年齢・時代区分データとこれまでの解析方法

1.2.1. 年齢・時代区分データ

年齢・時代区分データは、長期に渡って年齢区分別に観測されたデータである。一般には年齢と時代の区間長は任意であるが、本論文では特に、年齢区分の間隔と時代区分の間隔が等しく5年である場合をあつかう。つまり、調査時点 P_j ($= P_1 + 5(j-1)$; $j=1, 2, \dots, J$) に年齢区分 $[A_{i-1}, A_i)$ ($A_i = A_0 + 5i$; $i=1, 2, \dots, I$) に対応してまとめられたデータ (Figure 1.2.1-a), または、時代区分 $[P_{j-1}, P_j)$ と年齢区分 $[A_{i-1}, A_i)$ に対応してまとめられたデータ (Figure 1.2.1-b) をあつかう。両方の図で、観測されたデータは各セルに対応する。

時点 P_j に観測されるデータは、世論調査やマーケティング調査の質問項目への応答率などを想定している。また、時代区分 $[P_{j-1}, P_j)$ でまとめられたデータは、死亡強度 (death intensity) や発病強度 (incidence intensity) を想定している。 $[A_{i-1}, A_i)$ と P_j に対応する応答率を π_{ij} , $[A_{i-1}, A_i)$ と $[P_{j-1}, P_j)$ に対応する死亡/罹病強度を μ_{ij} で表すこととする。

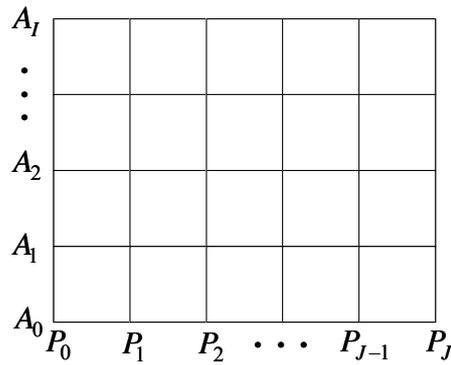


Figure 1.2.1-a : 年齢・時代区分データ（死亡データの場合）.

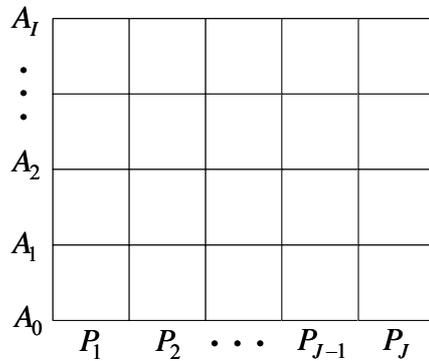


Figure 1.2.1-b : 年齢・時代区分データ（応答データの場合）.

1.2.2. これまでの解析方法（APCモデル）

年齢・時代区分データの解析では、年齢、時代そしてコホートの3つの効果を含むAge-period-cohort model（APCモデル）が頻繁に用いられている（Osmond, Gardener, 1982, 中村, 1982, Holford, 1983, 赤井・渋谷, 1984, Kupper *et al.*, 1985, Clayton, Schiffllers, 1987 b, Robertson, Boyle, 1989, Tango, Kurashina, 1987, Robertson, Boyle, 1998 参照）。APCモデルは $\log\{\pi_{ij}/(1-\pi_{ij})\}$ または $\log \mu_{ij}$ （両方を η_{ij} で表す）を

$$\eta_{ij} = \alpha_i + \beta_j + \gamma_{j-i} \quad (1.2.2-a)$$

と表すモデルである。ここで、 α_i は年齢区分 $[A_{i-1}, A_i)$ に対応した年齢効果、 β_j は時点 P_j または時代区分 $[P_{j-1}, P_j)$ に対応した時代効果、 γ_{j-i} は $j-i$ ($=1-I, 2-I, \dots, J-1$) 番目のコホートに対応するコホート効果である。

コホートとは「地理的に、もしくは、他のなんらかの方法で画された全住民のうちで、一定の時期に人生における同一の重大な出来事を体験した人々」(Glenn, 1977, pp.8) で、コホート

効果とは「コホート成員であることが原因となって生じる効果」(Glenn, 1977, p.11)と定義される。コホートの定義における「一定の時期」は様々で、結婚期の体験を考えた場合を「結婚コホート」、教育を受けた期間の体験を考えた場合を「教育コホート」と呼ぶことがある(Glenn, 1977, p.8)。そして、出生時を考えた場合を「出生コホート」(Glenn, 1977, p.8)と呼び、この場合、その効果を「時代区分 $[P_{j-i}, P_{j-i+1})$ に起きた出来事を、その時点の年齢が $[A_0, A_1)=[0, 5)$ である幼児が体験することに因る効果」と考えることができる。

1.3. APCモデルの問題点と基本的アイデア

1.3.1. APCモデルの問題点

APCモデルを用いたデータ解析では、1つの大きな問題点がある。(1.2.2-a)式において、任意に与えられた α_i , β_j , γ_{j-i} ($i=1, \dots, I; j=1, \dots, J$)に対して、 $\alpha_i^* = \alpha_i + a - ic$, $\beta_j^* = \beta_j - a + b + jc$, $\gamma_{j-i}^* = \gamma_{j-i} - b - (j-i)c$ を考えると、方程式の系

$$\alpha_i^* + \beta_j^* + \gamma_{j-i}^* = \alpha_i + \beta_j + \gamma_{j-i}, \quad i=1, \dots, I; j=1, \dots, J, \quad (1.3.1-a)$$

が任意の $a, b, c \in R$ に対して成り立つ、つまり、 $\alpha_i, \beta_j, \gamma_{j-i}$ はモデル(1.2.2-a)において識別不能なパラメータである。APCモデルにおける識別問題は様々な方法で表現されているが、(1.3.1-a)の表現はOsmond, Gardener (1989)によるものである。識別問題については、方程式の系(1.2.2-a)の行列表現を用いて、Kupper *et al.* (1985), Holford (1983), 赤井・渋谷 (1985)も議論している。

この問題について、ベイズ型モデル(中村, 1982, Berzuini, Clayton, 1994)、“出生時代区分の重複”を回避する方法(Robertson, Boyle, 1989)、推定可能関数のみで議論すべきという立場(Hirotsu, 1988, Holford, 1983, Tango, 1988)、病理学的見地から年齢効果に対して更なるモデルを仮定する方法(Holford, Zhang, McKay, 1994)など、様々な解決策が提案されている。

1.3.2. 基本的なアイデア

本論文で提案する方法は、死亡データの解析においては、上で述べた病理学的見地から年齢効果に対して更なるモデルを仮定する方法の範疇に入る。つまり、新しいモデルを提案し、そのモデルを用いた解析がAPCモデルを用いたものに比べて遜色のないものであり、新しいモデルを導入することが有意義であることを主張するものである。ただし、提案する3つのモデルに共通するキーワードは、環境効果である。環境効果は、年齢・時代区分データ解析では、本論文で新しく導入される概念で、「過去に人々が共通に体験した出来事の効果」と定義されるものである。その特徴は、体験の直後からその後何年にも渡って効果が現れることである。

「人々が共通に体験する出来事」や「過去の体験」という表現から、環境効果はAPCモデルにおける時代効果やコホート効果と同等のものであると思われるかもしれない。しかし、環境効果は以下の点で時代効果およびコホート効果と異なる。まず、環境効果は対応する時代区分／調査時点が過ぎても影響を持ち続ける効果であるが、時代効果はある時代区分／調査時点においてのみ影響を及ぼし、その後の時代区分／次の調査時点では影響を及ぼさない。また、環境効果はその時代区分／調査時点に生きる全ての人々（調査対象者）に影響を与えるが、コホート効果は世代の固有の効果で時代区分／調査時点とは無関係である。このような意味で、環境効果は新しい概念であり、時代効果やコホート効果では代用できない。

1.4. 3つの提案モデルとその新奇性

1.4.1. 3つの提案モデル

本論文では、Age-environment model (AEモデル)、Simple-two-stage model (STSモデル)、そして超高齢者統計に一般パレート分布を仮定したモデル（超高齢者モデル）の3つのモデルを提案する。全てのモデルは、年齢・時代区分データに対して定義されるが、AEモデルは、世論調査やマーケティング調査の質問項目への応答率および乳癌死亡率への適用を、STSモデルは全癌（全ての部位に関する癌）の死亡率への適用を想定している。また、超高齢者モデルでは、死亡者数統計と生存者数統計の両方を扱う。

統計モデルとして、応答率に対するAEモデルは一般線形模型（McCullagh, Nelder, 1983 参照）のロジスティック回帰モデルに、乳癌死亡率に対するAEモデルは一般線形模型のポアソン回帰モデルのクラスに、STSモデルは非線形ポアソン回帰モデルのクラスに属する。また、超高齢者モデルは、生存者数統計の場合は一般パレート分布関数の差分を用いた多項モデルとして、死亡者数統計の場合、一般パレート分布のハザード関数を用いたPiecewise-constant-intensity model（PCIモデル、Keiding, 1990参照）として提案される。

各モデルは、既存の統計モデルのクラスに属するが、それぞれの新奇性は以下に述べる通りである。

1.4.2. AEモデルの新奇性

AEモデルは、年齢・時代区分データの解析に環境効果を導入した最も単純なモデルで、APCモデルに含まれる時代効果やコホート効果と環境効果の関係式から、AEモデルは、計画行列の列空間の意味でAPCモデルに含まれ、環境効果は時代効果とコホート効果にある制約を与えたものに相当することを確認される。しかし、AEモデルでは、推定可能関数の中から、人間の年齢を固定した場合の環境効果の趨勢や、環境を一定と仮定した場合の年齢効果の趨勢を表すクラスを見出すことができる。一方、APCモデルでは、推定可能関数の中から、年齢・時代・コ

ホートの各効果の曲率（"curvature", Tango, 1988）や線形成分からの偏差（"deviations from linearity", Holford, 1983）を見出すことは出来るが、趨勢を表現できる推定可能関数のクラスは見出されていない。

また、AEモデルは最低影響年齢の設定によって異なる計画行列の列空間を持ち、最低影響年齢に依るモデルのクラスとして定義される。そして、実際のデータ解析では、データへ最も良い当てはまりを示す最低影響年齢が採択される。このように、AEモデルは、単に環境効果が導入されていること以外にも新奇性を持つ。（第3章参照。）

1.4.3. Simple-two-stage model（STSモデル）の新奇性

STSモデルは、分子生物学で研究されている癌腫瘍発生・成長モデルを参考にして導出される。そこでは、癌死亡強度を、人間が生まれてから現時点に至る期間に暴露され体内に摂りこまれた環境中の危険因子と、摂りこまれた危険因子が体内で末期癌にまで成長してしまう危険強度の畳込み和で表す。これまで同様のデータ解析に用いられたAPCモデルをはじめとする様々なモデルのほとんどは、一般線形模型に属すが、STSモデルは、癌死亡危険強度を2種類の危険強度の畳込み和で表すため、一般線形模型で用いられる結合関数を用いても線形化することができない。このため、STSモデルは非線形ポアソン回帰モデルのクラスに属する。このことが、STSモデルを用いた年齢・時代区分別の癌死亡データ解析の数学的な新奇性である。

非線形ポアソンモデルに含まれるパラメータの最尤推定量の存在性と一般性には保証されていない。そこで、STSモデルにおける最尤推定量の存在性と一般性は、本論文の証明によって明らかにされる。（第4章参照。）

1.4.4. 超高齢者モデルの新奇性

年齢・時代区分データは、年齢区分と時代区分で統合されたデータで、離散データの1つのタイプである。一方、寿命分布の限界を議論するために用いる極値理論は、連続データについて議論されている。そこで、超高齢者モデルの提案では、極値理論を年齢・時代区分データの解析と結び付けることに新奇性がある。

また、生存者数統計の解析に用いられる多項モデルでは、一般パレート分布関数から算出した年齢区分内の生存確率を用いる。しかし、これまでの超高齢者統計の研究では、このような方法で導いたモデルを使用した例は見当たらない。

1.5. 実データの解析結果の概要

1.5.1. 実データの解析結果を示す目的

本論文で提案する3つのモデルは、統計モデルとして、数学的に明らかにすべき性質を備えている点で、それ自身の理論的研究にも十分な意義がある。これに加えて、これらのモデルを用いた実際のデータ解析では、これまでとは異なる角度から新しい事実を見出し予測を行うことを可能にする。

このことを示すため、本論文では、各モデルに実際のデータを当てはめ、モデルに含まれる効果を表すパラメータの推定結果を示し、これまでの手法による解析結果との比較を行う。また、頻繁に年齢・時代区分データへ当てはめられるAPCモデルと比べて、残差尤度比およびAICの意味で、データに対する優れた当てはまりを示すことも紹介する。

1.5.2. AEモデルを用いた女子労働力人口の解析

個人応答についてのAEモデルを用いたデータ解析の例として、女子労働力人口の解析を行う。労働力人口は就業者と失業者の総数で、その全人口に占める割合は労働力率と呼ばれる。失業者は、非就業者で職安へ登録しているもの、つまり就労の意思を有するものを指すので、ある個人が労働力人口に数えられるか否かは、本人の就労意欲の有無によって決まると考えられる。特に男女雇用均等法の施行以前の女性の就労は、「平等でない労働環境の下でも働きたい」という個人の強い価値観の現われと考えられる。また、「1970年代のウーマンリブ運動」や「女性の高学歴化」等の時代環境の影響も強く受けていると考えられる。そこで、このようなデータにAEモデルを当てはめることは1つの自然な方法である。

データ解析の結果、「1970年代の女性の社会進出」、「1980年代の女性の高学歴化」そして「1980年代の女性の職場環境の改善」に対応する環境効果が抽出された。これらは、女子労働力率を上げる効果であるが、一方、「1971~75年のオイルショックに伴う不況」や「1990年代の女子大生の就職難」に対応する、労働力率を下げる効果も抽出された。また、最低影響年齢の推定から、10歳以上の年齢時に体験した時代環境が、女子の就労の意思決定に影響を及ぼすと判断された。(第3章参照。)

1.5.3. AEモデルを用いた乳癌死亡率の解析

死亡率についてAEモデルを用いたデータ解析の例として、乳癌死亡データの解析を行う。乳癌は女性にとって重大な病気の1つで、近年、特に発展途上国において発生率や死亡率が急激に増加している。原因としては、初潮から閉経までの生理期間などの年齢に関わる要因が知られているが、従来知られている要因では全体の1/3のケースしか説明できないと言われている。近年、DDTなどを含む環境中から摂取されるエストロゲン様物質(bioavailable estrogen existing

in the environment) や有機塩素系化合物への暴露が、発展途上国での乳癌死亡の増加に関わっているとして注目されている (Davis *et al.*, 1993, Wolff and Weston, 1997, 参照)。そこで、乳癌死亡データへAEモデルを当てはめ、乳癌リスクの趨勢に影響を及ぼしている過去の環境要因を明らかにすることは、有意義であると考えられる。

日本人の乳癌死亡率データ解析の結果、1945年前後の「第2次世界大戦」や1970年代の「高度経済成長期」に関連する環境効果が抽出され、最低影響年齢は10歳と推定された。また、北欧4ヶ国（ノルウェー、スウェーデン、デンマーク、フィンランド）における乳癌発生率データの解析の結果、フィンランドにおいてヘルシンキ・オリンピックが開催された1952年以降悪化した環境効果などを抽出することができた。さらに、2つのデータ解析において、AEモデルはAICの意味でAPCモデルよりデータによく当てはまることが示された。(第3章参照。)

1.5.4. STSモデルを用いた全癌死亡率の解析

STSモデルは、分子生物学の分野で明らかにされている癌の発生・成長過程を参考に構築されたモデルであるが、転移については細かい記述を行っていない。つまり、原発性の癌であることを前提としている。そこで、全部位を対象とした癌死亡（全癌死亡）データへの当てはめが最も自然である。

日本人の全癌死亡データ解析の結果、男性について、昭和初期の「世界恐慌」から「敗戦」に至る時代の環境効果や、1970年代の「高度経済成長期」の環境効果を抽出した。一方、女性については、男性の場合のような「世界恐慌」から「敗戦」に至る時代に関する環境効果の趨勢を見出すことはできなかった。このことは、女性の方が男性に比べて逆境に強いということを示唆している。また、日米両国のデータ解析結果の比較から、体内に摂り込まれた発癌因子が末期癌へ成長する危険強度の経時変化について日米共通の趨勢が見られた。このことは、癌に対する抵抗力/抑止力は、人類に共通の経時変化を見せることを示唆している。(第4章参照。)

1.5.5. 超高齢者統計の解析

これまで、日本においては、毎年平均年齢が上昇し、また、100歳以上の超高齢者人口が増加している。しかし、この場合の平均年齢や平均余命とは、当該年度の死亡者の平均年齢であり、現在生存している人々のその後の余命の期待値を表すものではない。寿命に関してより精密な分析を行うには、出生時期を同じくする人々、つまり世代について追跡的な分析を行う必要があり、その意味で、年齢・時代区分データとしての分析は重要である。

厚生労働省は1964年以来毎年9月に100歳以上の高齢者名簿、いわゆる長寿番付を発表している。100歳以上の各人の寿命が同一のパレート分布に従うと仮定し、1885-87年の世代の男性生

存者数データに、一般パレート分布から算出された毎年の死亡確率を用いた多項モデルを用いた解析から、「寿命分布に限界が無い」という仮説は棄却され、寿命分布の限界は134歳と推定された。死亡者数統計については、厚生労働省（厚生省）が毎年発表している人口動態統計があるが、85歳以上（年度によって90歳以上）の原因別死亡者数が1つの年齢区分にまとめられている。しかし、同省収蔵の保管第3表を閲覧すれば原因別各歳死亡者数が得られる。このデータにおいて、1980-83年の各年度に95歳である世代について、PCIモデルを当てはめた結果から、「寿命分布に限界が無い」という仮説は棄却され、寿命分布の限界は125-27歳と推定された。（第5章参照。）

1.6. 本論文の構成

最初に述べたように、本論文の目的は、年齢・時代区分データから、生活環境の効果を抽出するための3つの統計モデルを提案し、それらの持つ数学的性質を明らかにし、さらに、提案したモデルが実データ解析において有意義であることを示すことである。そのための準備として、まず、年齢・時代区分データを解析するための基本となる理論を紹介する。次に、これまでこのデータの解析に頻りに用いられてきたAPCモデルを紹介し、その問題点を明らかにする。そして、解析対象であるデータの種類（女子労働力人口、乳癌死亡、全癌死亡、超高齢者生存者/死亡者数）に応じて3つの新しい統計モデルを提案する。さらに、提案モデルを用いた場合とAPCモデルを用いた場合の実データの解析結果を比較し、また、パラメータの推定値から得られる所見を述べる。（ただし、超高齢者生存者/死亡者数の解析については、APCモデルとの比較は行わない。）

各章の内容は次のとおりである。

1.6.1. 第2章

第2章では、年齢と時代（時間）に依存する死亡強度と応答率を視覚的に表現するための道具としてレクシス・ダイアグラムを紹介し、年齢・時代区分データを解析するための統計モデルを考える上で基本となるPiecewise-constant-intensity model（PCIモデル）を導入する。そして、PCIモデルを用いた場合の最尤推定を議論する。また、Robertson, Boyle（1989）によるAPCモデルの識別問題の解釈と彼等が提案するもう一つのAPCモデルをレクシス・ダイアグラムを用いて紹介する。最後に、彼等の提案するモデルに対する批判を紹介するとともに、本論文の立場を明らかにする。

1.6.2. 第3章

第3章では、環境効果を導入したモデルであるAEモデルを定義し、識別問題の議論から推定可

能関数を明らかにする．そして，日本人についての女子労働力人口および乳癌死亡データ，北
欧4ヶ国の乳癌発生データへAEモデルを当てはめた結果から，AEモデルがAPCモデルよりも
AICの意味でデータへ良く当てはまること，および，AEモデルに含まれるパラメータの推定値
から得られる知見を示す．

1.6.3. 第4章

第4章では，STSモデルを提案する．STSモデルは非線形ポアソン回帰モデルのクラスに属し，
一般に性質が明らかになっている一般線形模型や一般加法模型のクラスに属さない．そこで，
尤度関数と尤度方程式を示し，最尤推定量の存在性と一致性を示す．また，モデルの識別問題
を議論し推定可能関数を示す．そして，日本人に関する全癌死亡データへSTSモデルを当ては
めた結果から，残差尤度比とAIC両方の意味で，STSモデルがAPCモデルよりもデータへ良く当
てはまること，および，STSモデルに含まれるパラメータの推定値から得られる知見を示す．

1.6.4. 第5章

第5章では，超高齢者の生存者数統計および死亡者数統計の解析を行う．まず，一般パレート分
布に関する一般的性質を示し，超高齢者データへ一般パレート分布を仮定することの妥当性を
議論する．そして，実際の生存者数統計および死亡者数統計のデータ解析から，「寿命分布に限
界がない」という仮説を否定し，寿命分布の限界の推定値を示す．

2. これまでの研究と本論文の立場

この章では、年齢と時代（時間）に依存する死亡強度と応答率を視覚的に表現するための道具としてレクシス・ダイアグラムを紹介し、年齢・時代区分データを解析するための統計モデルを考える上で基本となるPiecewise-constant-intensity model（PCIモデル）を導入する。そして、PCIモデルを用いた場合の最尤推定を議論する。また、Robertson, Boyle（1989）によるAPCモデルの識別問題の解釈と彼等が提案するもう一つのAPCモデルをレクシス・ダイアグラムを用いて紹介する。最後に、彼等の提案するモデルに対する批判を紹介するとともに、本論文の立場を明らかにする。

2.1. レクシス・ダイアグラムと死亡・応答統計

2.1.1. レクシス・ダイアグラム上の死亡強度・応答確率

レクシス・ダイアグラムは、19世紀に活躍した保険学者レクシス（Wilhelm Lexis, 1837～1914）が考案した図示法で、人間集団に属する各個人が、生まれてから死ぬまでを（年齢，時間）平面上に表したもので、Figure 2.1.1-aのように表される図式である。（Keiding, 1990参照。）図において横軸は時間，縦軸は年齢を表して、45度の斜線はある個人が横軸上の端点で出生し、右上の端点で死亡したことを表している。

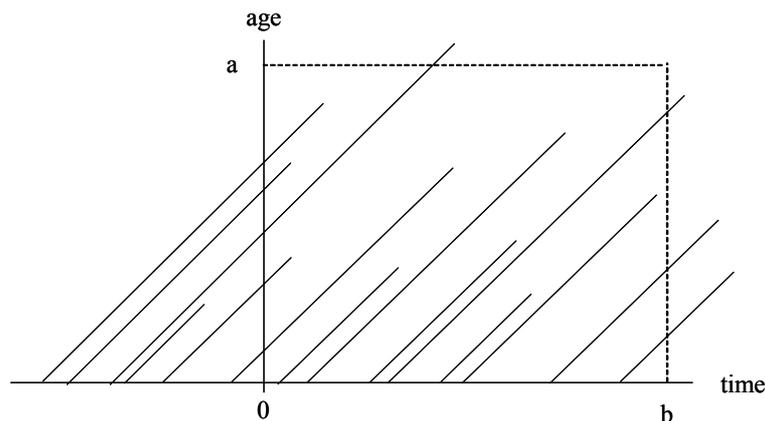


Figure 2.1.1-a : レクシス・ダイアグラム。

レクシス・ダイアグラムは人間の生誕と死亡を表したものであるが、価値観や嗜好を問う質

問項目への応答を扱う場合，Figure 2.1.1-bのように表すことができる．図において，斜実線は，“Yes”と応答する期間を，斜点線は“No”と応答する期間を表している．つまり，実斜線の時点で調査が行われれば，“Yes”と応答し，さらに，点線の時点で調査をすれば，“No”と応答することを表している．

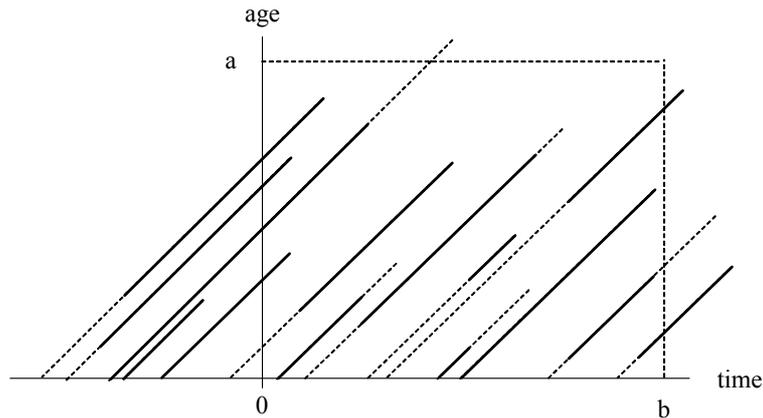


Figure 2.1.1-b：レクシス・ダイアグラム（個人応答の場合）．

年齢を $u \in [0, a)$ ，時代（時間）を $v \in [0, b)$ で表すとき，縦横の点線と縦軸・横軸で囲まれた長方形 $[0, a) \times [0, b)$ は，観測される範囲を表している．そして，時点 v で u 歳である人が続く微小時間 $[v, v + dx)$ に死亡する強度を $\mu^{all}(u, v) \in [0, \infty]$ で表す．しかし，死因には，様々な分類があるため，興味の対象である原因による死亡強度を $\mu(u, v)$ ，興味の対象外の原因による死亡強度を $\lambda(u, v)$ で表す．このとき，1人の人間の死亡が2つ以上の死因に分類されることは無いため，

$$\mu^{all}(u, v) = \mu(u, v) + \lambda(u, v)$$

である．また，ある調査に対して“Yes”と応答する確率を $\pi(u, v) \in [0, 1]$ で表す．

2.1.2. Piecewise constant intensity model（PCIモデル）

人口動態調査で得られる年齢・時代区分データは，時代区分 $[P_{j-1}, P_j)$ ($P_j = P_0 + 5j$; $j = 0, 1, \dots, J-1$) と年齢区分 $[A_{i-1}, A_i)$ ($A_i = A_0 + 5i$; $i = 1, 2, \dots, I$) で集計（統合）されたデータである．(Figure 1.2.1-a参照．) また，一定の年数間隔で行われる世論調査やマーケティング調査で得られる年齢・時代区分データは，調査時点 P_{j-1} に年齢区分 $[A_{i-1}, A_i)$ で集計（統合）されたデータである．(Figure 1.2.1-b参照．) このように，年齢・時代区分データは

$I \times J$ 個に集計（統合）された離散データの1つであり，このデータをレクシス・ダイアグラム上で連続関数として与えられた死亡強度 $\mu(u, v)$ あるいは応答確率 $\pi(u, v)$; $(u, v) \in [0, a) \times [0, b)$ についてのモデルを直接当てはめることはできない．

このような不都合に対処するため，レクシス・ダイアグラム上の死亡強度 $\mu(u, v)$ あるいは応答確率 $\pi(u, v)$ について，次のような仮定を置く Piecewise constant intensity model (PCIモデル) を導入する．

$$\begin{cases} \mu(u, v) = \mu_{ij} \\ \lambda(u, v) = \lambda_{ij} \end{cases} \text{ if } (u, v) \in W_{ij} \equiv \{(u, v) \mid (u, v) \in [5(i-1), 5i) \times [5(j-1), 5j)\}$$

(2.1.2-a)

$$\pi(u, v) = \pi_{ij} \text{ if } (u, v) \in \tilde{W}_{ij} \equiv \{(u, v) \mid (u, v) \in [5(i-1), 5i) \times [5(j-1), 5j)\}$$

(2.1.2-b)

ただし， a と b が $a = 5I$ と $b = 5J$ (I と J は正の整数) と表されるとする．(PCIモデルについては，例えば，Keiding, 1990, Berzuini, Clayton, 1994, Robertson, Boyle, 1998 を参照．) Figure 2.1.2-a はレクシス・ダイアグラム上の断片領域 W_{ij} と \tilde{W}_{ij} を示している．

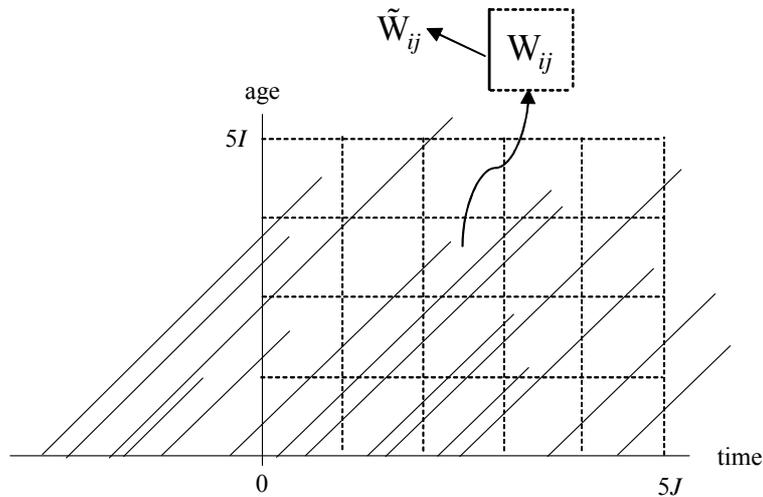


Figure 2.1.2-a : レクシス・ダイアグラムと Piecewise Constant Intensity Model .

2.2. 年齢・時代区分データに基づく尤度

2.2.1. 死亡データに基づく尤度

$X_{ij}^{(n)}$ ($n = 1, \dots, N_{ij}$) をレクシス・ダイアグラム上の断片 W_{ij} 上で生存する n 番目の個人の領域内の生存時間を表す確率変数とする．そして， $D_{ij}^{(n)}$ を，その n 番目の個人が，興味の対象

である原因で W_{ij} 内で死亡した ($D_{ij}^{(n)} = 1$) か否 ($D_{ij}^{(n)} = 0$) かを表す、表示関数とする。ただし、仮定 (2.1.2-a) から、

$$\Pr\{X_{ij}^{(n)} \in [x, x + dx) \mid X_{ij}^{(n)} \geq x\} = (\mu_{ij} + \lambda_{ij})dx. \quad (2.2.1-a)$$

である。このとき、確率変数の組 $(X_{ij}^{(n)}, D_{ij}^{(n)})$ に関する観測値の組 $(x_{ij}^{(n)}, d_{ij}^{(n)})$ に基づく尤度への寄与は

$$\mu_{ij}^{d_{ij}^{(n)}} \exp[-\{\mu_{ij} + \lambda_{ij}\} x_{ij}^{(n)}].$$

である。したがって、観測値の組 $(x_{ij}^{(n)}, d_{ij}^{(n)})$ の集まりに基づく尤度は、 $d_{ij}^{\bullet} \equiv \sum_{n=1}^{N_{ij}} d_{ij}^{(n)}$ 、 $x_{ij}^{\bullet} \equiv \sum_{n=1}^{N_{ij}} x_{ij}^{(n)}$ とすると、

$$L^* \equiv \prod_{i=1}^I \prod_{j=1}^J \mu_{ij}^{d_{ij}^{\bullet}} \exp[-\{\mu_{ij} + \lambda_{ij}\} x_{ij}^{\bullet}], \quad (2.2.1-b)$$

に比例する (Keiding, 1975 参照)。ただし、 L^* は部分尤度である。このとき、実際に与えられる (必要となる) データは $(d_{ij}^{\bullet}, x_{ij}^{\bullet})$; $i = 1, \dots, I$; $j = 1, \dots, J$ である。そして、 d_{ij}^{\bullet} は W_{ij} 内の斜線の端点の数、 x_{ij}^{\bullet} は W_{ij} 内の斜線の総延長である (Figure 2.2.1-a 参照)。

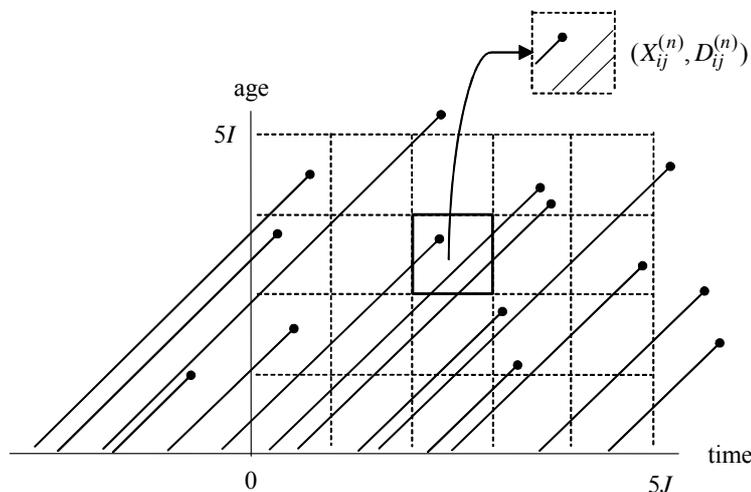


Figure 2.2.1-a : PCIモデルを仮定したレクシス・ダイアグラム上の死亡者数。

部分尤度関数 (2.2.1-b) はPCIモデルの研究でよく知られていて、その形は $D_{ij}^{\bullet} \equiv \sum_{n=1}^{N_{ij}} D_{ij}^{(n)}$ が $E[D_{ij}^{\bullet}] = x_{ij}^{\bullet} \mu_{ij}$ であるようなポアソン分布に従うと仮定した場合の部分尤度と同様である。このことから、 μ_{ij} に関するモデルはポアソン回帰モデルと呼ばれる (例えば Keiding, 1990, Berzuini, Clayton, 1994, Robertson, Boyle, 1998 を参照)。しかし、 $X_{ij}^{\bullet} \equiv \sum_{n=1}^{N_{ij}} X_{ij}^{(n)}$ とおくと、

$(X_{ij}^{\bullet}, D_{ij}^{\bullet})$ は $(X_{i-1,j-1}^{\bullet}, D_{i-1,j-1}^{\bullet})$, $(X_{i-1,j}^{\bullet}, D_{i-1,j}^{\bullet})$, $(X_{i,j-1}^{\bullet}, D_{i,j-1}^{\bullet})$ と独立ではなく, 厳密には, L^* はポアソン回帰モデルに関する部分尤度とはいえない (Figure 2.2.1-b 参照). また, $(X_{ij}^{(n)}, D_{ij}^{(n)})$ は明らかに *i.i.d.* ではないため, *i.i.d.* の仮定の下で一般的に示された最尤推定量の存在性と一致性 (Lehmann, Casella, 1998 参照) は保証されない. ところが, *i.i.d.* を仮定しなくても, Lehmann, Casella (1998) と同様の議論を L^* および μ_{ij} に関するモデル $\mu_{ij} = \mu_{ij}(\theta)$; $\theta = (\theta_1, \theta_2, \theta_3, \dots, \theta_p)'$ について展開することにより, 最尤推定量の存在性と一致性を示すことができる. ただし, 本章での θ は任意のモデルにおけるパラメータベクトルを表すが, 第3章ではAEモデルのパラメータベクトル $\theta = (\mu, \alpha_1, \dots, \alpha_I, \xi_{2-I-\phi}, \dots, \xi_J)'$, 第4章ではSTSモデルのパラメータベクトル $\theta = (\psi_1, \dots, \psi_I, \xi_{2-I}, \dots, \xi_J)'$ に相当する.

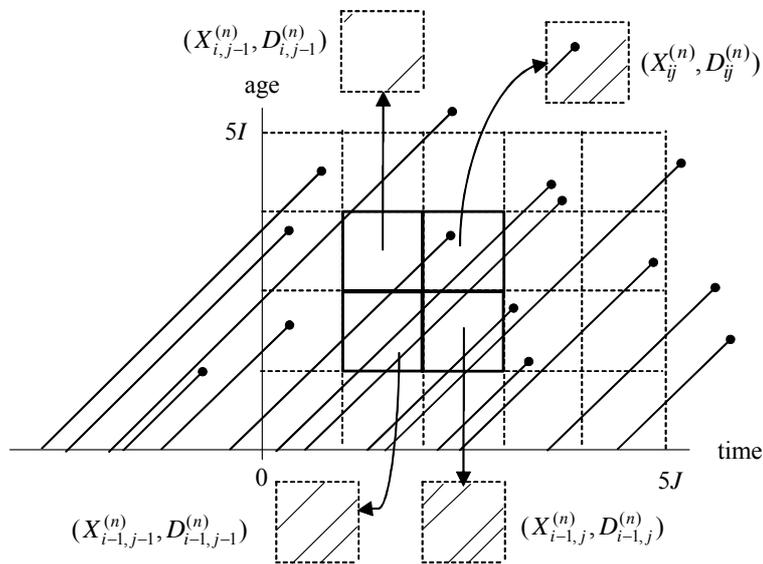


Figure 2.2.1-b : PCIモデルを仮定したレクシス・ダイアグラム上の独立でない領域.

以下の議論のために, 次の記号法を導入する. M_k を時代区分 $(5(k-1), 5(k+1))$ に生まれた人数, $b_k^{(m)}$ ($k=1-I, \dots, J-1$; $m=1, \dots, M_k$) を時代区分 $(5(k-1), 5(k+1))$ に生まれた m 番目の個人の誕生時点とする. また, 時代区分 $(5(k-1), 5(k+1))$ に生まれた m 番目の個人が, 死亡せずに $W_{i,k+i}$ を通過した場合の長さを表すために

$$\tau_k^{(m)} = \begin{cases} b_k^{(m)} - 5(k-1) & \text{if } b_k^{(m)} \in (5(k-1), 5k) \\ 5(k+1) - b_k^{(m)} & \text{if } b_k^{(m)} \in [5k, 5(k+1)) \end{cases}$$

を導入する. Figure 2.2.1-clは W_{ij} , $b_{j-i}^{(m)}$, $\tau_{j-i}^{(m)}$ をレクシス・ダイアグラム上で図示したものである.

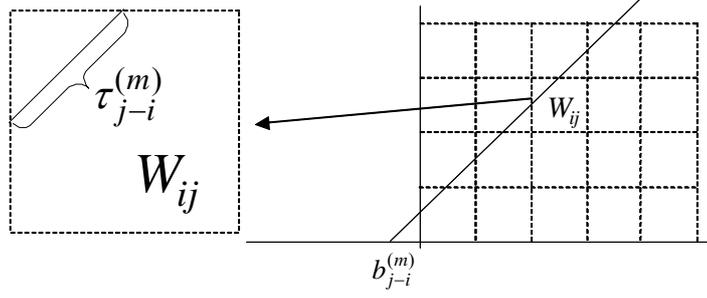


Figure 2.2.1-c : レクシス・ダイアグラム上で表した $\tau_k^{(m)}$.

このとき次の命題を与える .

Proposition 2.2.1: $\mu_{ij}(\boldsymbol{\theta})$, M_k および $\tau_k^{(m)}$ が次のCondition 2.2.1-aからCondition 2.2.1-dを満たすならば ,

$$\left. \frac{\partial \log L^*}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \mathbf{0}$$

を満たす解 $\hat{\boldsymbol{\theta}}$ が , $\min\{M_{1-l}, \dots, M_{J-1}\} \rightarrow \infty$ のとき , 1に近づく確率で存在し , $\boldsymbol{\theta}_0$ への一致性を持つ .

Condition 2.2.1-a : すべての l, m, n, i, j に対して次式を満たす定数 G_{ij} と H_{ij} が存在する .

$$\sup_{\boldsymbol{\theta} \in Q_a(\boldsymbol{\theta}_0)} \left| \frac{\partial^3 \mu_{ij}(\boldsymbol{\theta})}{\partial \theta_l \partial \theta_m \partial \theta_n} \right| \leq G_{ij} , \quad \sup_{\boldsymbol{\theta} \in Q_a(\boldsymbol{\theta}_0)} \left| \frac{\partial^3 \log \mu_{ij}(\boldsymbol{\theta})}{\partial \theta_l \partial \theta_m \partial \theta_n} \right| \leq H_{ij} ,$$

ただし , $\boldsymbol{\theta}_0$ は $\boldsymbol{\theta}$ の真値 , $Q_a(\boldsymbol{\theta}_0)$ は十分に小さい半径 $a > 0$ の $\boldsymbol{\theta}_0$ の近傍とする .

Condition 2.2.1-b : 任意の $\boldsymbol{\theta} \in Q_a(\boldsymbol{\theta}_0)$ に対し , 次の行列が正値定符号である .

$$\sum_{i=1}^I \sum_{j=1}^J \frac{\partial \mu_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \mu_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$$

Condition 2.2.1-c : 次式を満たす k が存在しない .

$$\lim_{M_k \rightarrow \infty} \frac{1}{M_k} \sum_{m=1}^{M_k} \tau_k^{(m)} = 0 .$$

Condition 2.2.1-d : $\min\{M_{1-l}, \dots, M_{J-1}\} \rightarrow \infty$ のとき , 次式を満たす $h_0 > 0$ が存在する .

$$\frac{\max\{M_{1-l}, \dots, M_{J-1}\}}{\min\{M_{1-l}, \dots, M_{J-1}\}} < h_0 .$$

ところで、実際の調査において各個人の生存時間が正確に観測されることはない。このため、 x_{ij}^\bullet の値は、通常、 W_{ij} に対応する人口年（person-years at risk, $N_{ij}^{(py)}$ で表す）によって近似される（例えば、Keiding, 1990 参照）。人口年とは、連続した期間（多くの場合5年）の各年の人口（生存者数）を累積した値で、 W_{ij} に対応する人口年 $N_{ij}^{(py)}$ は、Figure 2.2.1-c における○の数である。

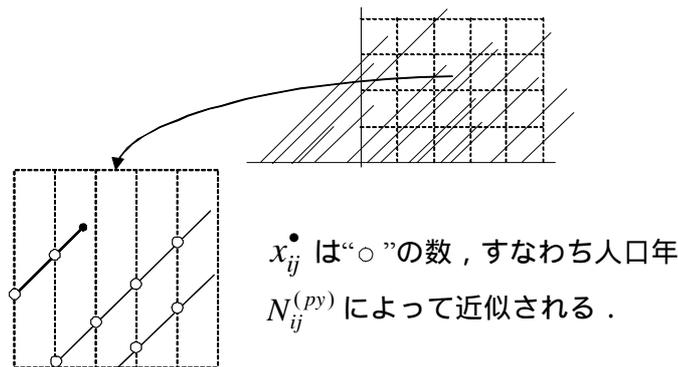


Figure 2.2.1-d : x_{ij}^\bullet の人口年 $N_{ij}^{(py)}$ による近似。

2.2.2. 応答データに基づく尤度

$R_{ij}^{(n)}$ を、レクシス・ダイアグラム上の断片 \tilde{W}_{ij} で生存する個人 n ($=1, \dots, \tilde{N}_{ij}$) が調査に対して “Yes” と応答する ($R_{ij}^{(n)} = 1$) か否 ($R_{ij}^{(n)} = 0$) を表す、表示関数とする。このとき、確率変数 $R_{ij}^{(n)}$ の観測値 $r_{ij}^{(n)}$ に基づく尤度への寄与は

$$\pi_{ij}^{r_{ij}^{(n)}} (1 - \pi_{ij})^{1 - r_{ij}^{(n)}}.$$

である。したがって、観測値 $r_{ij}^{(n)}$ の集まりに基づく尤度は

$$\tilde{L}^* \equiv \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{r_{ij}^\bullet} (1 - \pi_{ij})^{\tilde{N}_{ij} - r_{ij}^\bullet} \quad (2.2.2-a)$$

に比例する、ただし、 $r_{ij}^\bullet \equiv \sum_{n=1}^{\tilde{N}_{ij}} r_{ij}^{(n)}$ で、 $\tilde{N}_{ij} - r_{ij}^\bullet$ は、 \tilde{W}_{ij} で生存し、調査に対して “No” と応答した人数である。このとき、実際に与えられる（必要となる）データは $(r_{ij}^\bullet, \tilde{N}_{ij})$; $i=1, \dots, I$; $j=1, \dots, J$ である。そして、 r_{ij}^\bullet は \tilde{W}_{ij} と斜実線の交点の数、 \tilde{N}_{ij} は \tilde{W}_{ij} と斜実線との交点および斜点線との交点の数の和である（Figure 2.2.2-a参照）。

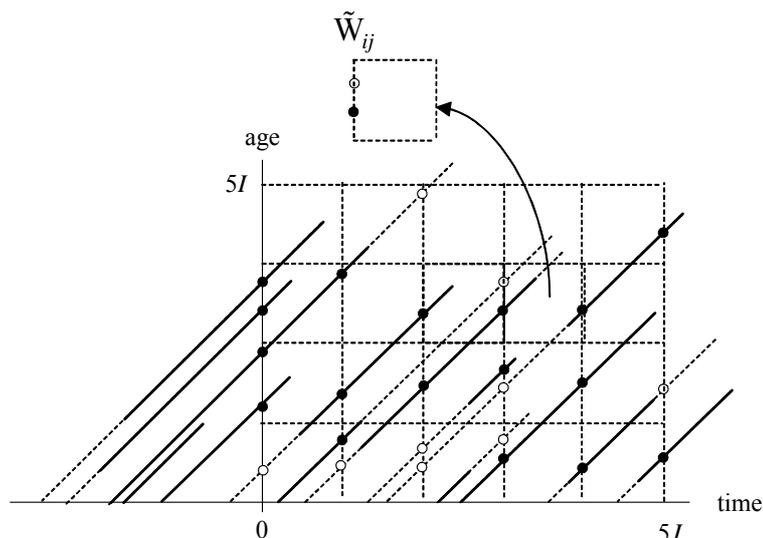


Figure 2.2.2-a : PCIモデルを仮定したレクシス・ダイアグラム上の応答数 ,

● = “yes” , ○ = “no” .

調査対象の人々の応答が互いに独立であると仮定すると, R_{ij}^{\bullet} ($\equiv \sum_{n=1}^{\tilde{N}_{ij}} R_{ij}^{(n)}$) と $R_{i-1,j-1}^{\bullet}$ は独立ではないが, 死亡データの場合と異なり, R_{ij}^{\bullet} と $R_{i-1,j}$ または $R_{i,j-1}$ は独立である. そこで, 例えば $I=4$; $J=5$ のとき, データセット $\{r_{12}^{\bullet}, r_{23}^{\bullet}, r_{34}^{\bullet}, r_{45}^{\bullet}\}$ (Figure 2.2.2-b参照) の生起確率は

$$\begin{aligned}
 & P(r_{12}^{\bullet}, r_{23}^{\bullet}, r_{34}^{\bullet}, r_{45}^{\bullet}) \\
 &= P(r_{12}^{\bullet})P(r_{23}^{\bullet} | r_{12}^{\bullet})P(r_{34}^{\bullet} | r_{12}^{\bullet}, r_{23}^{\bullet})P(r_{45}^{\bullet} | r_{12}^{\bullet}, r_{23}^{\bullet}, r_{34}^{\bullet}) \\
 &= \binom{\tilde{N}_{12}}{r_{12}^{\bullet}} \pi_{12}^{r_{12}^{\bullet}} (1-\pi_{12})^{\tilde{N}_{12}-r_{12}^{\bullet}} \binom{\tilde{N}_{23}}{r_{23}^{\bullet}} \pi_{23}^{r_{23}^{\bullet}} (1-\pi_{23})^{\tilde{N}_{23}-r_{23}^{\bullet}} \\
 &\quad \times \binom{\tilde{N}_{34}}{r_{34}^{\bullet}} \pi_{34}^{r_{34}^{\bullet}} (1-\pi_{34})^{\tilde{N}_{34}-r_{34}^{\bullet}} \binom{\tilde{N}_{45}}{r_{45}^{\bullet}} \pi_{45}^{r_{45}^{\bullet}} (1-\pi_{45})^{\tilde{N}_{45}-r_{45}^{\bullet}}
 \end{aligned} \tag{2.2.2-b}$$

である. このことから, \tilde{L}^* がデータセット $\{r_{ij}^{\bullet}\}_{i=1,j=1}^{I,J}$ に基づく部分尤度であることが確かめられる.

部分尤度 \tilde{L}^* は, 形式上は, $R_{ij}^{\bullet} \equiv \sum_{n=1}^{\tilde{N}_{ij}} R_{ij}^{(n)}$ が $E[R_{ij}^{\bullet}] = \tilde{N}_{ij}\pi_{ij}$ であるような2項分布に従うと仮定した場合の部分尤度と同様である. このことから, π_{ij} に関するモデルは2項回帰モデルに帰着することが分かる. 2項回帰モデルは, 一般にアンケート調査などへの回答率に対して仮定されるモデルであるが, 年齢・時代区分で与えられる応答率に関するモデルも, このモデルに帰着することが, 以上の議論から明らかになった.

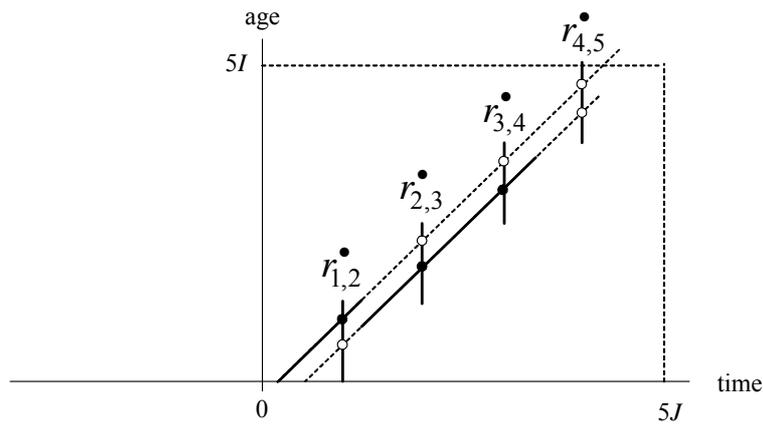


Figure 2.2.2-b：レクシス・ダイアグラム上の出生時代区分が同じ集団の応答数．

Figure 2.2.2-b から分かるように，データセット $\{r_{ij}^{\bullet}\}_{i=1, j=1}^{I, J}$ は各調査時点での応答数であり，調査時点以外，すなわち

$$W_{ij} - \tilde{W}_{ij} = \{(u, v) \mid (u, v) \in [5(i-1), 5i) \times (5(j-1), 5j)\}$$

での応答の変化（例えば，調査時点間 $(5(j-1), 5j)$ に，“Yes” “No” “Yes” と回答が変わったこと）は観測されない．このため，このようなデータの解析では，調査と調査の間の時点に起きた出来事の影響を見逃す可能性があることに注意が必要である．逆に，短い期間しか影響を及ぼさないような出来事であっても，調査時点あるいはその直前に起きた場合には観測データへ影響を及ぼすことにも注意が必要である．このような問題は，aggregation biasと呼ばれ，世論調査などを，どの程度の時間（年数）間隔で行うべきかという問題である．

2.3. レクシス・ダイアグラムとAPCモデル

2.3.1. レクシス・ダイアグラム上の年齢・時代・コホート効果

年齢・時代区分データの解析では，年齢効果，時代効果そしてコホートの3つの効果を含むAge-period-cohort model（APCモデル）が頻繁に用いられている．APCモデルは， $\log\{\pi_{ij}/(1-\pi_{ij})\}$ または $\log \mu_{ij}$ （両方を η_{ij} で表す）を（1.2.2-a）式のように

$$\eta_{ij} = \alpha_i + \beta_j + \gamma_{j-i}$$

と表すモデルである．ここで， α_i は年齢区分 $[A_{i-1}, A_i)$ に対応した年齢効果， β_j は時点 P_j または時代区分 $[P_{j-1}, P_j)$ に対応した時代効果， γ_{j-i} （ $j-i=1-I, 2-I, \dots, J-1$ ）は $j-i$ （ $=1-I, 2-I, \dots, J-1$ ）番目のコホートに対応するコホート効果であるが，出生コホートの考えを採用すれば，時代区分 $[P_{j-i}, P_{j-i+1})$ に起きた出来事を，その時点の年齢が $[A_0, A_1)=[0, 5)$ である幼児が体験することに因る効果である．

3つの効果を，PCIモデル (2.1.2-a) の仮定の下で，レクシス・ダイアグラム上の断片 W_{ij} を用いて以下のように表現することができる．ただし，具体的に示すために， $I = 5$; $J = 6$ の場合を考える．まず， η_{ij} ($i = 1, 2, \dots, 5$; $j = 1, 2, \dots, 6$) が (1.2.2-a) のように表され，さらに，APCモデルに含まれるパラメータが $\alpha_1 = \alpha_2 = \alpha_4 = \beta_1 = \dots = \beta_5 = \gamma_{-3} = \dots = \gamma_4 = 0$ である場合を考えよう．この場合，年齢効果 α_3 は

$$\{W_{31}, W_{32}, W_{33}, W_{34}, W_{35}\}$$

に共通した効果である．(Figure 2.3.1-a 参照．)

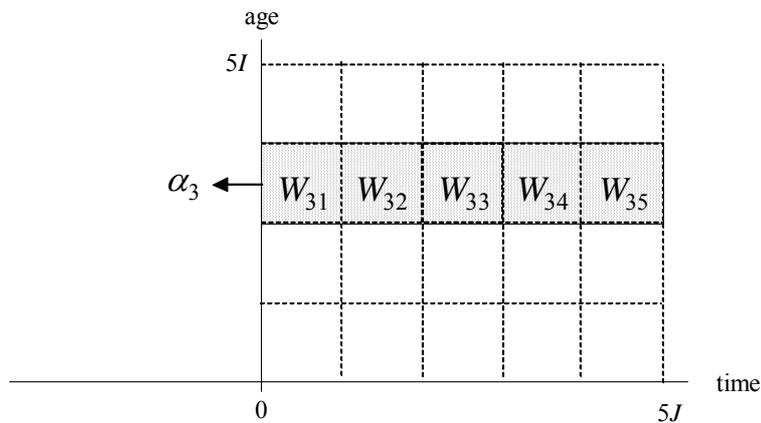


Figure 2.3.1-a : PCIモデルを仮定したレクシス・ダイアグラム上の年齢効果．

つぎに，APCモデルに含まれるパラメータが $\alpha_1 = \dots = \alpha_4 = \beta_1 = \beta_2 = \beta_4 = \beta_5 = \gamma_{-3} = \dots = \gamma_4 = 0$ である場合を考えよう．この場合，時代効果効果 β_3 は

$$\{W_{13}, W_{23}, W_{33}, W_{43}\}$$

に共通した効果である．(Figure 2.3.1-b 参照．)

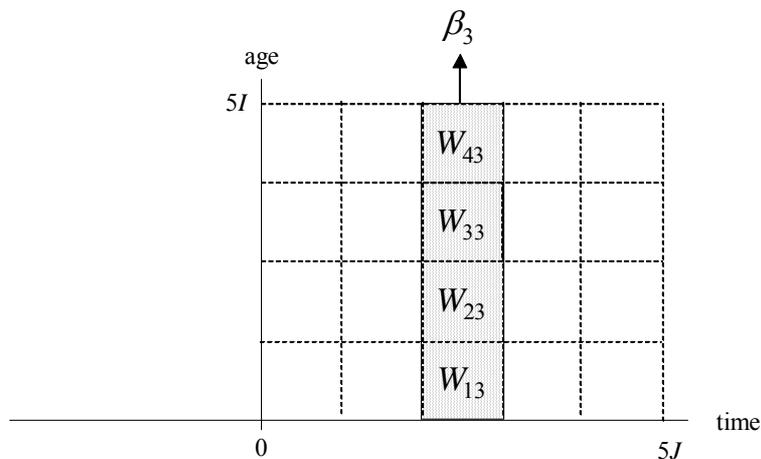


Figure 2.3.1-b : PCIモデルを仮定したレクシス・ダイアグラム上の時代効果．

最後に，APCモデルに含まれるパラメータが $\alpha_1 = \dots = \alpha_4 = \beta_1 = \dots = \beta_5 = \gamma_{-3} = \dots = \gamma_0 = \gamma_2 = \dots = \gamma_4 = 0$ である場合を考えよう．この場合，コホート効果 γ_1 は

$$\{W_{12}, W_{23}, W_{34}, W_{45}\}$$

に共通した効果である．(Figure 2.3.1-c 参照．)

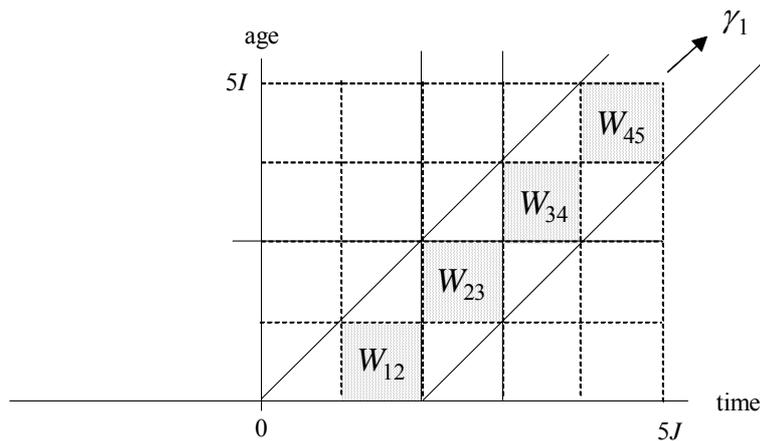


Figure 2.3.1-c : PCIモデルを仮定したレクシス・ダイアグラム上のコホート効果．

2.3.2. “出生区分の重複”による識別問題の解釈ともう一つのAPCモデル

前述のように，APCモデルでは，パラメータの識別問題が存在する．Osmond, Gardener (1989) は (1.3.1-a) を用いて識別問題を表現したが，Robertson, Boyle (1989) は，“コホート効果に対応する出生時代区分の重複”による説明を行っている．これは，次のように説明することができる．例えばコホート効果 γ_1 に対応するレクシス・ダイアグラム上の断片群 $W_{1,2}, W_{2,3}, W_{3,4}, W_{4,5}$ で生存しうる個人の出生時代区分は $[P_0, P_2)$ であるが，これは隣接するコホート効果 γ_2 に対応するレクシス・ダイアグラム上の断片 $W_{1,3}, W_{2,4}, W_{3,5}$ で生存しうる個人の出生時代区分 $[P_1, P_3)$ と重複している．一方，年齢効果と時代効果については，例えば， α_3 と β_3 について，対応する年齢・時代区分は $[A_2, A_3)$ と $[P_2, P_3)$ で，隣接する効果の対応する年齢・時代区分と重複しない．(Figure 2.3.2-a 参照．)

Robertson, Boyle (1989) は，この出生時代区分がAPCモデルの識別問題の原因と考え，断片領域 W_{ij} を上三角部分 W_{ij}^0 と下三角部分 W_{ij}^1 に分割し，出生時代区分の重複しないコホート効果を用いた独自のAPCモデルを提案している．(Figure 2.3.2-b参照．) Robertson, Boyle (1989) のモデルでは， W_{ij}^0 と W_{ij}^1 内でのデータを振り分けるために，個人毎に観測されたデータを用いているが， W_{ij} で統合されているデータについては，モデルを当てはめることができない．

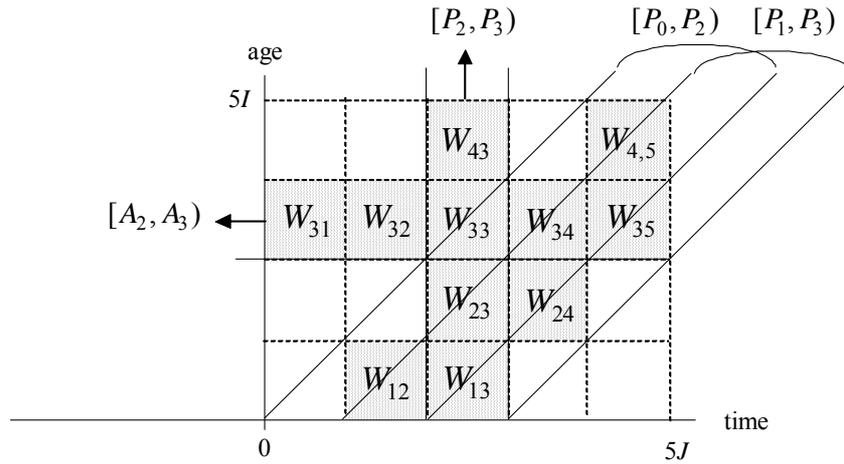


Figure 2.3.2-a : PCIモデルにおける出生時代区分の重複 .

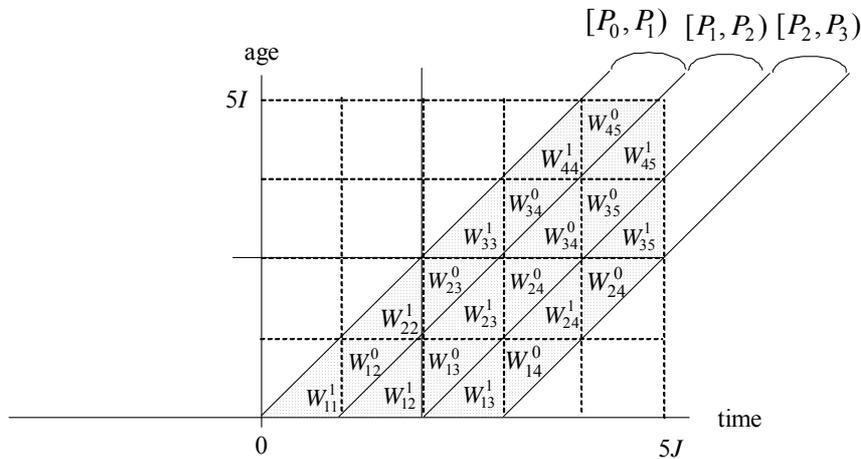


Figure 2.3.2-b : 出生時代区分の重複を回避したコホート効果 .

2.4. Robertson-Boyle model への批判と本論文の立場

2.4.1. Robertson-Boyle model への批判

Robertson, Boyle (1989) のモデルは、様々な研究者から批判を受けている。例えば、Osmond, Gardener (1989) は識別問題の表現方法 (1.3.1-a) を連続変数 $u, v \in R$ について、

$$\begin{aligned} \alpha^*(u) &= \alpha(u) + a - cu, \\ \beta^*(v) &= \beta(v) - a + b + cv, \\ \gamma^*(v-u) &= \gamma(v-u) - b - (v-u)c \end{aligned}$$

を考えると、

$$\mu(u, v) = \alpha(u) + \beta(v) + \gamma(v - u) = \alpha^*(u) + \beta^*(v) + \gamma^*(v - u) \quad (2.4.1-a)$$

が任意の $a, b, c \in R$ に対して成り立つことを示し, APCモデルにおけるパラメータの識別問題が, 連続関数としてモデルを考えた場合にも起きる根本的な問題であると考えた. そして, Robertson, Boyle (1989) の指摘する “出生時代区分の重なり” は, 本質的な問題ではないと主張している.

さらに, レクシス・ダイアグラム上の断片 W_{ij}^0 と W_{ij}^1 の重心位置と対応させた年齢・時代・コホートの各効果を考え, Robertson, Boyle (1989) のモデルが, 隣り合う効果が等しいという制約を課したものに過ぎない, という批判を展開している.(Figure 2.4.1-a参照.) また, Tango (1988) は, Robertson, Boyleが1987年に同じモデルを用いてスコットランドの肺癌と喉頭癌の発生率データを解析した論文に対して, Osmond, Gardener (1989) と同様の批判を行っている.

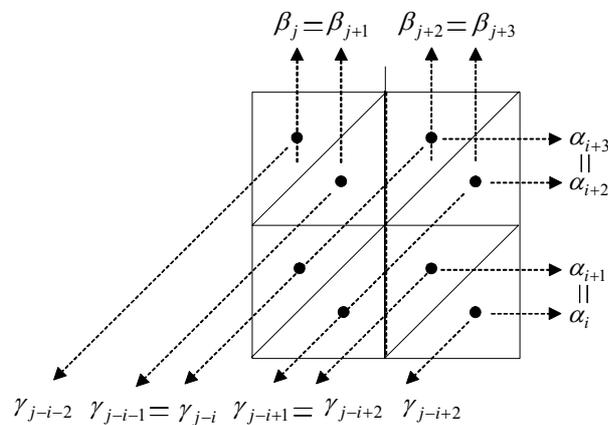


Figure 2.4.1-a : 出生時代区分の重複を回避したコホート効果を用いたAPCモデルでの制約 .

2.4.2. 本論文の立場

本論文では, Osmond, Gardener (1989) やTango (1988) の立場を取る. その理由の1つは, 価値観や嗜好に関する質問項目への応答について観測されるレクシス・ダイアグラム上の領域 \tilde{W}_{ij} に関しては, Robertson, Boyle (1989) の指摘する “出生時代区分の重なり” が起きないからである.(Figure 2.4.2-a参照.) つまり, \tilde{W}_{ij} で観測されその5年後に生存している人々は $\tilde{W}_{i+1, j+1}$ 上で観測され他の領域で観測されないため, データセット $\{\dots, (r_{ij}^\bullet, N_{ij}), (r_{i+1, j+1}^\bullet, N_{i+1, j+1}), \dots\}$ は $[P_{j-i-1}, P_{j-i})$ に生まれたコホート集団を追跡したものである. もう1つの理由は, 本論文で提案されるAE, STSの2つのモデルはレクシス・ダイアグラム上の確率過程としての人間の死亡や応答に対して導入したモデルを, 年齢・時代区分データへ適応できる形へ変えたものだからである. つまり, モデルの根本はOsmond, Gardener (1989) 同様, 連続関数としての死亡強度や応答率にあるのである.

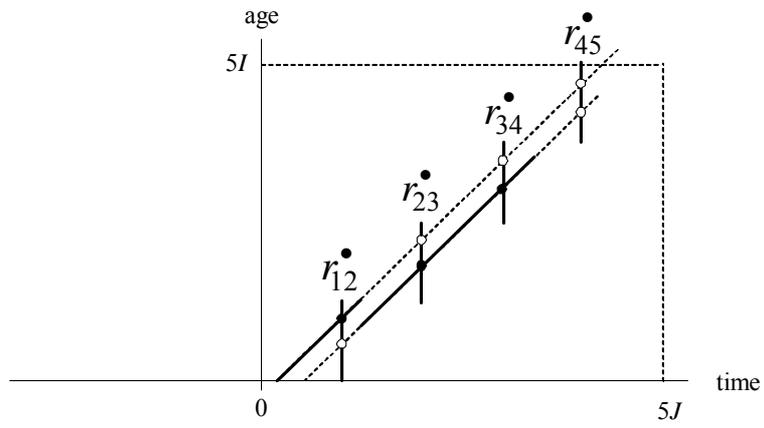


Figure 2.4.2-a : レクシス・ダイアグラム上の出生時代区分が同じ集団の応答数 .

3. Age-environment model (AEモデル)

本章では、環境効果を導入したモデルであるAEモデルを定義し、識別問題の議論から推定可能関数を明らかにする。そして、乳癌死亡データ、北欧4ヶ国の乳癌発生データ、日本人についての女子労働力人口へAEモデルを当てはめた結果から、AEモデルがAICの意味でAPCモデルよりデータへ良く当てはまること、および、AEモデルに含まれるパラメータの推定値から得られる知見を示す。

3.1. 背景と概要

3.1.1. 背景

乳癌は女性にとって重大な病気の1つで、近年、特に発展途上国において発生率や死亡率が急激に増加している。原因としては、初潮から閉経までの生理期間などの年齢に関わる要因が知られているが、従来知られている要因では全体の1/3のケースしか説明できないと言われている。近年、DDTなどを含む環境中から摂取されるエストロゲンや有機塩素系化合物への暴露が、発展途上国での乳癌死亡の増加に関わっているとして注目されている (Davis *et al.*, 1993, Wolff, Weston, 1997, 参照)。そこで本章では、年齢時代区分で与えられた乳癌死亡率データの解析から、年齢に伴って変化するリスクと環境に伴って変化するリスクの趨勢を投影することを目的の1つとする。

労働力人口は就業者と失業者の総数で、その全人口に占める割合は労働力率と呼ばれる。失業者は、非就労者で職安へ登録しているもの、つまり就労の意思を有するものを指すので、ある個人が労働力人口に数えられるか否かは、本人の就労意欲の有無によって決まると考えられる。特に男女雇用機会均等法の施行以前の女性の就労は、「平等でない労働環境の下でも働きたい」という個人の強い価値観の顕われと考えられる。また、「1970年代のウーマンリブ運動」や「女性の高学歴化」等の時代環境の影響も強く受けられていると考えられる。また、従来から知られているように、「新卒時の就職」、「結婚、出産、子育てに伴う離職」そして「子供がある程度の年齢に達した際の仕事への復帰」などの年齢の影響が大きいと考えられる。したがって、このようなデータにAEモデルを当てはめることは1つの自然な方法である。そこで、本章のもう1つの目的は、年齢・時代区分別に女子労働力人口データの解析から、時代環境の効果と加齢による効果を抽出することである。

調査対象として、ある国の国民全体のように同じ居住地域に住む人々を選んだ場合、時代環境の変化は世代間の差異として現れる。つまり、APCモデルの考え方で言えば、コホート効果が時代環境の影響を反映していると考えられる。しかし、Glenn (1977) によるコホート効果の定義 (副節1.2.2参照) によれば、コホート効果は特定の年齢層が体験する時代環境の

効果で、その時代に生きる全ての人々が共通に体験する環境の効果ではない。そこで本章では、年齢・時代区分データを解析するために、APCモデルに替わる新しいモデルを提案し、ある時代環境を体験した全ての人々に共通に現れる効果を抽出することを目指す。

3.1.2. 本節の概要

次の3.2節では、年齢・時代区分別に与えられた死亡強度、または、アンケート調査項目への応答率に対して、環境効果のアイデアを新しく導入する。そして、年齢・時代区分別の死亡強度や応答率に対するモデルとして、環境効果と年齢効果をパラメータとして持つ Age-environment model (AEモデル) を提案する。

年齢・時代区分別に与えられる癌死亡率の研究では、対数を取った癌死亡強度を年齢効果、時代効果そしてコホート効果の線形関数で表す age-period-cohort model (APCモデル) が頻繁に用いられている。(Osmond, Gardener, 1982, Holford, 1983, Kupper *et al.*, 1985, Clayton, Schifflers, 1987 b, Robertson, Boyle, 1989, Robertson, Boyle, 1998 参照。) AEモデルで用いられる環境効果とAPCモデルで用いられる時代効果とコホート効果は共に時代区分と関連する効果である(副節3.3.1参照)。このため、AEモデルは何らかの意味でAPCモデルと関係を持つものと思われるだろう。そこで、副節3.3.2では環境効果と時代効果またはコホート効果の違いについて述べる。しかしながら、モデルをベクトルと行列で表現した場合、計画行列の列空間の意味では、AEモデルはAPCモデルの“副モデル”の1つであることを副節3.3.3で示す。

AEモデルでは、人々が生まれてから現在に至るまでに暴露された全ての環境リスクに影響されると仮定している。しかしながら、一般に知られているように、生理の始まっていない女性の乳房は発達していない。事実、10歳未満の女児が乳癌で死亡することは極めて希である。このことから、ある年齢以下の女性については発癌物質への暴露が影響を与えないと考えることもできる。このような事実に対処するため、第3.4節では「発癌物質への暴露が影響を及ぼす最低年齢」(最低影響年齢)を導入したAEモデルを新たに提案する。また、モデルに含まれるパラメータ(効果)の推定可能関数(Scheffé, 1959参照)を明らかにする。

第3.5節では、APCモデルの他に、age-period model (APモデル) と age-cohort model (ACモデル) を参照モデルに加えて、最低影響年齢を導入したAEモデルと参照モデルを日本の乳癌死亡データへ当てはめた解析結果を示す。解析結果から、AEモデルはAPモデルおよびACモデルより、残差尤度比の意味で、データへより良く当てはまることが分かる。そして、AICの意味では、最低影響年齢が0, 5, 10, 15歳の各AEモデルがAPCモデルより良い当てはまりを示すことが分かる。また、全てのAEモデルの中で、最低影響年齢が10歳のAEモデルが最も小さいAIC値を示すことから、最低影響年齢は10歳と推察される。さらに、環境効果と年齢効果の推定値の趨勢から得られる知見(findings)を与える。

最近の研究では、Rostgaard *et al.* (2001) がAPCモデルをデンマーク、フィンランド、ノルウ

エーそしてスウェーデンの年齢・時代区分別の乳癌発生率データへ当てはめ、コホート効果と時代効果の4ヶ国間比較を行っている。かれらは十分に解析結果を検討し、いくつかの重要な知見を与えているが、AEモデルを当てはめてみることも興味深い。そこで、第3.6節では、AEモデルを北欧4ヶ国の同じデータへ当てはめる。また、日本のデータの場合と同様に、AEモデルはAICの意味でAPCモデルよりデータへ良く当てはまったことを示す。さらに、4ヶ国間の比較から得られたいくつかの知見を与える。

第3.7節では、年齢・時代区分別に観測された女子労働力人口データへAEモデル、APモデル、ACモデルそしてAPCモデルを当てはめた解析結果を示す。解析結果からAEモデルは、AICの意味で、APモデルやACモデルより良い当てはまりを示すことが分かる。また、最低影響年齢は10歳と推察され、「10歳以降で体験する時代環境が20歳以降の就労の意思に影響を及ぼす」ことが分かる。

3.2. AEモデル

3.2.1. 連続関数としてのAEモデル

年齢 $u \in [0, a)$ と時点 (年) $v \in [0, b)$ の乳癌死亡強度を $\mu(u, v)$, ある調査に対して “ Yes ” と応答する確率を $\pi(u, v)$ とし, $\log \mu(u, v)$ と $\log [\pi(u, v) / \{1 - \pi(u, v)\}]$ の両方を $\eta(u, v)$ で表そう。そして、時点 v に年齢が u であるような個人と、時点 v に年齢が $(u - du)$ であるような個人を考えよう。このとき、これら2人の違いは、年齢だけでなく、年長者の方は時間区分 $[v - u, v - u + du)$ における環境リスクに暴露されているが、年少者の方は暴露されていないことに気づくであろう。この考えに基づいて、年長者に関する乳癌死亡強度と年少者に関する乳癌死亡強度との間に $\eta(u, v) - \eta(u - du, v) = \{\xi(v - u) + \delta(u)\} du$ を仮定する。(Figure 3.2.1-a 参照。) この仮定の下で乳癌死亡強度関数は

$$\eta(u, v) = \mu_c + \alpha(u) + \int_{v-u}^v \xi(t) dt, \quad (3.2.1-a)$$

と表される。ただし、 μ_c は総平均、 $\alpha(u) = \int_0^u \delta(t) dt$ は年齢区分 $[u, u + du)$ に関する年齢効果、そして $\xi(t)$; $t \in (-a, b)$ はこの論文で新しく導入される時間区分 $[t, t + dt)$ における環境リスクへの暴露に関する環境効果である。式 (3.2.1-a) において、環境効果の積分 $\int_{v-u}^v \xi(t) dt$ は $v - u$ および v の2変数に対する関数で、 $v - u$ だけの関数ではないことに気づくが、このことによって環境効果がコホート効果として知られるものと異なることが分かる。以上のモデルが、本論文で提案するAEモデルの基本的なアイデアを表している。

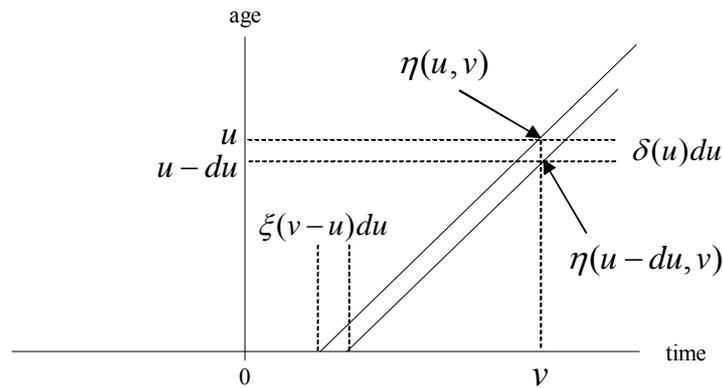


Figure 3.2.1-a : 連続死亡強度についてのAEモデルの考え方 .

3.2.2. PCIモデルとしてのAEモデル

この章で解析するデータは、年齢区分 $[A_{i-1}, A_i)$; $i = 1, \dots, I$ と時代区分 $[P_{j-1}, P_j)$; $j = 1, \dots, J$, について与えられる乳癌死亡者数と人口年 (person-years), および、年齢区分 $[A_{i-1}, A_i)$ と調査時点 P_{j-1} について与えられる女子の労働力人口と総人口である . ただし、 $A_0 = 5\phi$, $A_i = A_0 + 5i$, $P_j = P_0 + 5j$, $\phi (= 0, \dots, I-1)$ は、 $5\phi (= A_0)$ がデータの与えられている最低年齢を表す定数である . このとき、乳癌死亡者数と人口年に副節2.2.1における $(d_{ij}^{\bullet}, N_{ij}^{(py)})$ を、そして、女子の労働力人口と総人口に副節2.2.2における $(r_{ij}^{\bullet}, \tilde{N}_{ij})$ を対応させると、これらのデータにPCIモデルを当てはめることができる .

そこで、(3.2.1-a) と同じアイデアに基づいて $\log \mu_{ij}$ と $\log \{ \pi_{ij} / (1 - \pi_{ij}) \}$ (両方を η_{ij} で表す) に対して以下のようなモデルを仮定する .

$$\eta_{ij} = \mu_c + \alpha_i + \sum_{k=j-i-\phi+1}^j \xi_k . \quad (3.2.2-a)$$

(レクシス・ダイアグラムを用いた説明は、Figure 3.2.2-a を参照 .) このモデルを age-environment model (AEモデル) と呼ぶ .

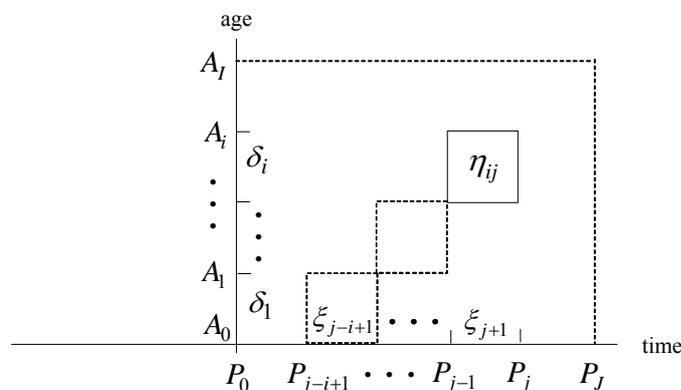


Figure 3.2.2-a : PCIモデルについてのAEモデルの考え方 .

3.3. APCモデルとの関係

3.3.1. APCモデルと年齢，時代，コホート効果の定義

第1章の副節1.2.2で述べたように，年齢・時代区分データの解析では，年齢効果，時代効果そしてコホート効果の3つの効果を含むAge-period-cohort model (APCモデル)が頻繁に用いられているAPCモデルは， η_{ij} ($i = 1, \dots, I$; $j = 1, \dots, J$)を

$$\eta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{j-i}$$

と表すモデルである．ここで， α_i は年齢区分 $[A_{i-1}, A_i)$ に対応した年齢効果， β_j は時点 P_j または時代区分 $[P_{j-1}, P_j)$ に対応した時代効果である．また，「出生コホート」(Glenn, 1977, p.8)の考えを導入すると， γ_{j-i} は「時代区分 $[P_{j-i-\phi}, P_{j-i-\phi+1})$ に起きた出来事を，その時点の年齢が $[A_{-\phi}, A_{1-\phi}) = [0, 5)$ である幼児が体験することに因る効果」である．

AEモデルで用いられる環境効果とAPCモデルで用いられる時代効果とコホート効果は共に時代区分と関連する効果である．このため，AEモデルは何らかの意味でAPCモデルと関係を持つものと思われるだろう．そこで，次の副節3.3.2では環境効果と時代効果またはコホート効果の違いについて述べる．しかしながら，モデルをベクトルと行列で表現した場合，計画行列の列空間の意味では，AEモデルはAPCモデルの“副モデル”の一つであることを副節3.3.3で示す．

3.3.2. 環境効果と時代効果，コホート効果との違い

ここでは，環境効果と時代効果，コホート効果との違いを説明するために， $I = 3$ ， $J = 4$ ， $\phi = 1$ の場合について具体的に説明する．まず，以下の3つのパターンを考えよう．

- (1) η_{ij} にAEモデルを仮定し， ξ_2 以外の全ての効果を0としたパターン． (Figure 3.3.2-a)
- (2) η_{ij} にAPCモデルを仮定し， β_2 以外の全ての効果を0としたパターン． (Figure 3.3.2-b)
- (3) η_{ij} にAPCモデルを仮定し， γ_2 以外の全ての効果を0としたパターン． (Figure 3.3.2-c)

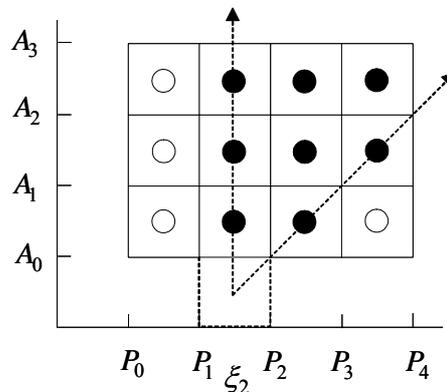


Figure 3.3.2-a: η_{ij} にAEモデルを仮定し， ξ_2 以外の全ての効果を0としたパターン，

●は環境効果が有効な領域，○は環境効果が無効な領域．

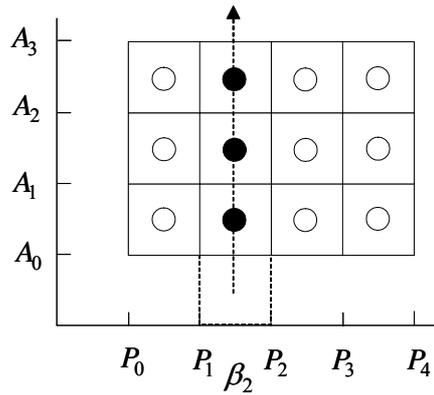


Figure 3.3.2-b : η_{ij} に APCモデルを仮定し, β_2 以外の全ての効果を 0 としたパターン, ●は環境効果が有効な領域, ○は環境効果が無効な領域.

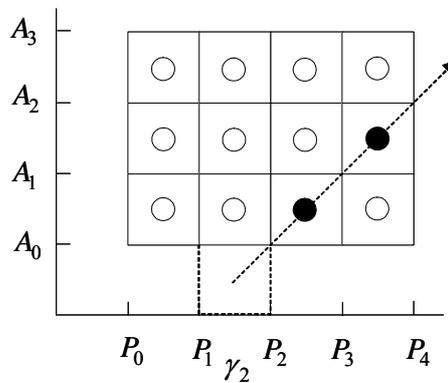


Figure 3.3.2-c : η_{ij} に APCモデルを仮定し, γ_2 以外の全ての効果を 0 としたパターン. (●は環境効果が有効な領域, ○は環境効果が無効な領域.)

このとき, 上の3つのパターンを見ながら, ξ_2 と β_2 または γ_2 の違いを次のように説明する.

(a) パターン(1)と(2)を比べると, ξ_2 は時点 P_2 の後も影響を持ち続けているが, 一方, β_2 は時代区分 $[P_1, P_2)$ においてのみ影響を持ち, P_2 以後は影響を持たない.

(b) パターン(1)と(3)を比べると, ξ_2 は時代区分 $[P_1, P_2)$ に生存する全ての人々に影響を与えるのに対して, γ_2 は時代区分 $[P_1, P_2)$ において年齢が $[A_{-\phi}, A_{1-\phi}) = [0, 5)$ であるような人々にだけ影響を与える.

このように, 環境効果と時代効果またはコホート効果は異なる効果であることが分かる.

3.3.3. APCモデルの副モデルとしてのAEモデル

まず, 次の表記法を導入しよう.

$$\begin{aligned} \boldsymbol{\eta} &= (\eta_{11}, \dots, \eta_{1J}, \dots, \eta_{I1}, \dots, \eta_{IJ})', \\ \boldsymbol{\theta} &= (\mu, \alpha_1, \dots, \alpha_I, \xi_{2-I-\phi}, \dots, \xi_J)', \\ \mathbf{X} &= (\mathbf{x}_{11}, \dots, \mathbf{x}_{1J}, \dots, \mathbf{x}_{I1}, \dots, \mathbf{x}_{IJ})'; \quad \mathbf{x}_{ij} = (1, \mathbf{0}_{i-1}, \mathbf{1}, \mathbf{0}_{I-i}, \mathbf{0}_{j-1+I-i}, \mathbf{1}_{i+\phi}, \mathbf{0}_{J-j})', \\ \mathbf{0}_k &= (\underbrace{0, 0, \dots, 0}_{k \text{ elements}}), \quad \mathbf{1}_k = (\underbrace{1, 1, \dots, 1}_{k \text{ elements}}), \\ \boldsymbol{\pi} &= (\alpha_1, \dots, \alpha_I, \beta_1, \dots, \beta_J, \gamma_{1-I}, \dots, \gamma_{J-1})', \\ \mathbf{Z} &= (\mathbf{z}_{11}, \dots, \mathbf{z}_{1J}, \dots, \mathbf{z}_{I1}, \dots, \mathbf{z}_{IJ})'; \\ \mathbf{z}_{ij} &= (1, \mathbf{0}_{i-1}, \mathbf{1}, \mathbf{0}_{I-i}, \mathbf{0}_{j-1}, \mathbf{1}, \mathbf{0}_{J-j}, \mathbf{0}_{I-i+j-1}, \mathbf{1}, \mathbf{0}_{J+i-j-1})'. \end{aligned}$$

この表記法によって，AEモデルとAPCモデルを，それぞれ

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\theta} \quad , \quad \boldsymbol{\eta} = \mathbf{Z}\boldsymbol{\pi}$$

と表すことができる．そして，上式で用いられている計画行列 \mathbf{X} と \mathbf{Z} について

$$\mathbf{X} = \mathbf{Z}\mathbf{F}$$

という関係を得る．ただし，

$$\begin{aligned} \mathbf{F} &\equiv (\mathbf{f}^{(\mu)}, \mathbf{f}_1^{(\alpha)}, \dots, \mathbf{f}_I^{(\alpha)}, \mathbf{f}_{2-I-\phi}^{(\xi)}, \dots, \mathbf{f}_J^{(\xi)}); \\ \mathbf{f}^{(\mu)} &= (1, \mathbf{0}_I, \quad \mathbf{0}_J, \quad \mathbf{0}_{I+J-1} \quad)' \\ \mathbf{f}_i^{(\alpha)} &= (0, \mathbf{0}_{i-1}, \mathbf{1}, \mathbf{0}_{I-i}, \quad \mathbf{0}_J, \quad \mathbf{0}_{I+J-1} \quad)' \\ \mathbf{f}_k^{(\xi)} &= \begin{cases} (0, \mathbf{0}_I, \quad \mathbf{0}_J, \quad \mathbf{1}_{k-1+I+\phi}, \mathbf{0}_{J-k-\phi})'; & k = 2-I-\phi, \dots, 1, \\ (0, \mathbf{0}_I, \quad -\mathbf{1}_{k-1}, \mathbf{0}_{J-k+1}, \quad \mathbf{1}_{k-1+I+\phi}, \mathbf{0}_{J-k-\phi})'; & k = 2, \dots, J-\phi-1, \\ (0, \mathbf{0}_I, \quad -\mathbf{1}_{k-1}, \mathbf{0}_{J-k+1}, \quad \mathbf{1}_{I+J-1} \quad)'; & k = J-\phi, \dots, J. \end{cases} \end{aligned}$$

である．計画行列の関係式 $\mathbf{X} = \mathbf{Z}\mathbf{F}$ は， $I = 3$ ， $J = 4$ ， $\phi = 1$ とすると，具体的に次のように書くことができる．

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

このとき，関係式 $\mathbf{X} = \mathbf{Z}\mathbf{F}$ から

$$\begin{cases} R(\mathbf{X}) \subset R(\mathbf{Z}) & J - \phi \geq 3 \\ R(\mathbf{X}) = R(\mathbf{Z}) & J - \phi \leq 2, \end{cases} \quad (3.3.3-a)$$

であることが分かり、このことから、計画行列 X の階数は、 $J - \phi \geq 3$ のとき、 X の列ベクトルの数より2少ない数すなわち $2I + J + \phi - 2$ で、 $J - \phi \leq 2$ のときは、 $2I + 2J - 4$ であることが分かる。ただし、 $R(\bullet)$ は行列の列空間を表す。(証明は、付録Aを参照。)

このように、環境効果は時代効果やコホート効果と異なるものと考えられるにも関わらず、計画行列の列空間の意味では、AEモデルはAPCモデルに含まれ、APCモデルの副モデル (submodels) の1つと考えることができる。

3.4. 最低影響年齢を導入したAEモデルとその推定可能関数

3.4.1. 最低影響年齢を導入したAEモデル

第3.1節で述べたように、乳癌に関する発癌性物質への暴露は、ある年齢以下の女兒には影響を与えないかもしれない。このような事実に対処するため、次のようなもう一種類のAEモデルを提案する。

$$\eta_{ij} = \mu + \alpha_i + \sum_{k=j-i-\phi+\rho+1}^j \xi_k; \quad \rho = 0, 1, \dots, \phi \quad (3.4.1-a)$$

ただし、 ρ は 5ρ が最低影響年齢を表す定数である。

ここで、Figure 3.4.1-aとFigure 3.4.1-bを見ながら (3.4.1-a) を (3.2.2-a) と比べると、計画行列の列空間が等しいという意味で、最低影響年齢が 5ρ であるAEモデルと $A_0 = 5(\phi - \rho)$ である場合のAEモデルが同等であることが分かる。したがって、最低影響年齢が 5ρ である場合のAEモデルの計画行列 X の階数は、 $J - (\phi - \rho) \geq 3$ のとき、 $2I + J + \phi - \rho - 2$ に等しく、 $J - (\phi - \rho) \leq 2$ のときは、 $2I + 2J - 4$ に等しい。

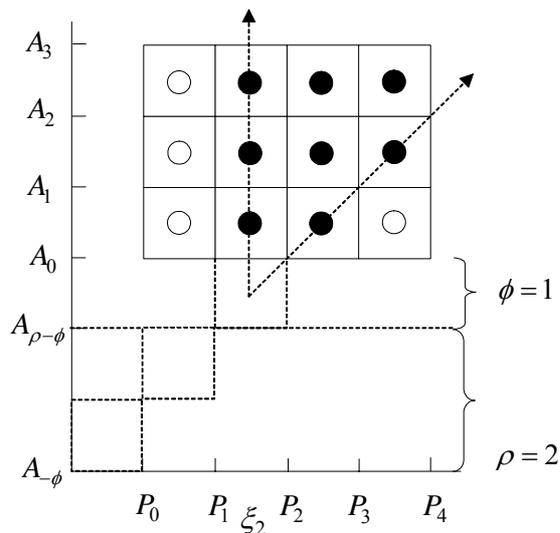


Figure 3.4.1-a : $\rho = 2$ の場合に、 $\phi = 1$ のAEモデルを η_{ij} に仮定し、
 ξ_2 以外は全てゼロとしたパターン、
 ●は環境効果が有効な領域、○は環境効果が無効な領域。

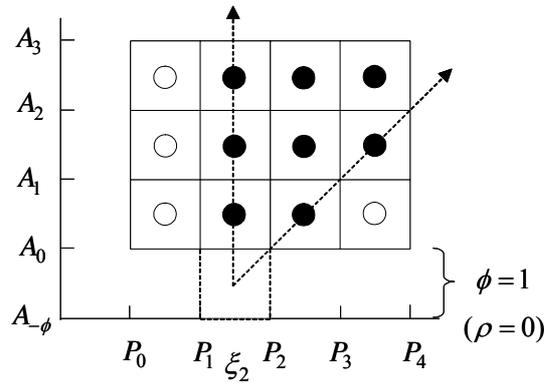


Figure 3.4.1-b : $\rho = 0$ の場合に , $\phi = 1$ の AE モデルを η_{ij} に仮定し ,
 ξ_2 以外は全てゼロとしたパターン ,
 ●は環境効果が有効な領域 , ○は環境効果が無効な領域 .

3.4.2. 推定可能関数

AEモデルの計画行列 X の階数は X に含まれる列ベクトルの数よりも小さいため , 明らかに各パラメータ (効果) は推定不可能 , つまり , 方程式の系 $\eta = X\theta$ においてパラメータベクトル θ の各成分は一意には定まらない . 通常 , このような問題が発生した場合 , θ に制約を置き , その制約の下でパラメータが推定される . この場合 , θ の最尤推定値を $\hat{\theta}$ とすると , $\hat{\eta} \equiv X\hat{\theta}$ は一意であるが , $\hat{\theta}$ の各成分は θ に対する制約の置き方に依存して値が変わってしまう . Scheffé (1959) は , 線形不偏推定量についてこの問題を議論し , $c \in R(X')$ を用いた関数 $c'\hat{\theta}$ の値は , 制約の置き方に依らず一意に定まることを示したが , 部分尤度 L^* に基づく最尤推定値についても , 同様の関数が一意に定まることが知られている (Holford, 1983 参照) . そこで , 以下では , $c'\theta ; c \in R(X')$ を AEモデルの推定可能関数と呼び , AEモデルの全ての推定可能関数を明らかにするとともに , それらの中から , 年齢効果と環境効果の趨勢を把握するのに適すると思われるものを見出し , 紹介する .

まず , $J - (\phi - \rho) \geq 3$ の場合 , 全ての推定可能関数は

$$\begin{cases} \alpha_2 - \alpha_1 + \xi_k & k = 2 - I - (\phi - \rho), \dots, J \\ \mu + \alpha_i + (i + \phi - \rho)\xi_J & i = 2, \dots, I \end{cases} \quad (3.4.2-a)$$

の線形結合によって表される . (証明は付録Bを参照 .) そして , 推定可能関数 (3.4.2-a) の線形結合の中で , $i (= 2, \dots, I)$ を固定した場合の推定可能関数のクラス

$$\xi_k^{(i)} \equiv \alpha_i - \alpha_{i-1} + \xi_k : k = 2 - I - (\phi - \rho), \dots, J$$

を , 年齢が区分 $[A_{i-2}, A_{i-1})$ から $[A_{i-1}, A_i)$ へ移る過程の女性に影響を与える環境効果の趨勢を表しているとして解釈可能なものとして導入する . また , $k (= 2 - I - (\phi - \rho), \dots, J)$ を固

定した場合のもう一つの推定可能関数のクラス

$$\alpha_i^{(k)} \equiv \mu + \alpha_i + (i + \phi - \rho)\xi_k; \quad i=1, \dots, I$$

を，環境中のリスクが $[P_{k-1}, P_k)$ における場合から変化しないという仮定の下での年齢効果の趨勢を表していると解釈可能なものとして導入する．

$J - (\phi - \rho) \leq 2$ の場合には， $\mathbf{X} = \mathbf{ZF}$ という関係と $R(\mathbf{X}) = R(\mathbf{Z})$ という事実から，

$$\begin{cases} \xi_{2-I-(\phi-\rho)} = \gamma_{1-I} \\ \xi_{k-(\phi-\rho)+1} = \gamma_{k-1} - \gamma_k & k = 2-I, \dots, J-1 \\ \xi_k = \beta_k - \beta_{k-1} & k = 2, \dots, J \\ \xi_J = \beta_J \end{cases} \quad (3.4.2-b)$$

という対応関係を得ることができる．これによって， $J - (\phi - \rho) \leq 2$ の場合のAEモデルの推定可能関数は (3.4.2-b) とAPCモデルでの推定可能関数から直ぐに求めることができる．そして，この場合には，これまでのAPCモデルに関する研究において解釈可能なものとして導入されている推定可能関数（例えば，Holford, 1983, Tango, Kurashina, 1987 参照）を用いることを推奨する．

3.5. 日本の乳癌死亡データの解析

3.5.1. 解析結果の概要

年齢時代区分別に与えられた日本における乳癌死亡率と対応する人口年をTable 3.5.1-aとTable 3.5.1-bに示す．そして，各モデルをデータへ当てはめた結果の概要を残差尤度比（deviance）とAICによって，Table 3.5.1-cに示す．これらの表から以下のような知見を得ることができる．

(i) ρ の値をどの値（ $\rho = 0, \dots, 5$ ）に設定しても，AEモデルは残差尤度比とAICの両方の意味でACモデルやAPモデルより良い当てはまりを示している．

(ii) $\rho = 0, \dots, 3$ に設定したAEモデルは，AICの意味で，わずかではあるが，APCモデルより良い当てはまりを示している．

(iii) $\rho = 2$ に設定したAEモデルが，AICの意味で，他の全てのモデルより良い当てはまりを示している．したがって，最低影響年齢は10歳と推察される．

さらに，次の副節では，最低影響年齢を10歳に設定した場合のAEモデルの環境効果と年齢効果の推定値から得られる知見を与える．

Table 3.5.1-a : 人口100,000あたりの乳癌死亡者数 (日本).

		Time period							
Age group	<i>j</i>	1	2	3	4	5	6	7	8
<i>i</i>		1950-54	1955-59	1960-64	1965-69	1970-74	1975-79	1980-84	1985-89
1	25 - 29	0.47	0.44	0.38	0.46	0.55	0.68	0.85	0.70
2	30 - 34	1.76	1.69	1.69	1.75	2.32	2.52	2.73	2.76
3	35 - 39	3.88	4.01	3.90	4.11	4.46	4.80	5.29	6.01
4	40 - 44	7.16	6.59	6.57	6.80	7.82	8.27	8.69	10.00
5	45 - 49	8.89	8.51	9.61	9.96	11.72	12.51	13.03	14.31
6	50 - 54	10.95	10.49	10.80	12.36	14.64	16.57	17.43	18.79
7	55 - 59	12.83	11.36	11.51	12.99	15.02	17.77	20.31	21.48
8	60 - 64	13.35	12.03	10.67	12.67	14.51	16.43	19.41	20.86
9	65 - 69	13.74	12.55	12.03	12.10	13.85	16.49	17.88	19.01
10	70 - 74	14.40	15.81	13.87	12.65	14.04	15.56	17.47	18.85

Table 3.5.1-b : 人口年 (person-years at risk) (日本女性, 単位: 100人).

		Time period							
Age group	<i>j</i>	1	2	3	4	5	6	7	8
<i>i</i>		1950-54	1955-59	1960-64	1965-69	1970-74	1975-79	1980-84	1985-89
1	25 - 29	178912	198499	207687	218858	231675	261975	205727	191697
2	30 - 34	150630	176597	195039	207881	219129	232942	261392	205949
3	35 - 39	135180	148776	174368	194170	207059	219866	232354	260571
4	40 - 44	122820	132576	146328	172097	191504	207502	218622	231251
5	45 - 49	101777	120077	129728	143432	169578	191731	205461	216934
6	50 - 54	89314	98672	116457	126261	142031	168580	189411	202740
7	55 - 59	72734	85378	94480	111355	121710	138706	165203	186089
8	60 - 64	61283	68148	80720	88884	106789	119410	134896	161235
9	65 - 69	49268	54010	61354	72564	82324	102078	113381	129554
10	70 - 74	38043	40489	45292	51856	64084	74660	93140	105552

Table 3.5.1-c : AE, AC, AP, APCモデルを日本女性の乳癌死亡データへ当てはめた結果の概要.

Model	Deviance	d.f.	Difference of AIC from APC's
AE ($\rho = 5$)	241.3	54	128.3
AE ($\rho = 4$)	149.5	53	38.4
AE ($\rho = 3$)	108.0	52	-1.0
AE ($\rho = 2$)	101.1	51	-5.9
AE ($\rho = 1$)	103.4	50	-1.7
AE ($\rho = 0$)	102.8	49	-0.3
APC	101.1	48	0
AC	300.0	54	186.9
AP	296.1	63	165.0

3.5.2. パラメータ推定値についての考察

$\hat{\zeta}_k^{(2)}$ の推定値 :

Figure 3.5.2-aに $\hat{\zeta}_k^{(2)}$ ($k = -11, \dots, 8$) の推定値を示す。ただし, k は1945 + 5*k* 年から1949 + 5*k* までの時代区分に対応している。図から, $\hat{\zeta}_k^{(2)}$ の推定値には, 1985-99年, 1940-44年そして1970-74年に対応した3つの目だったピークを見つけることができる。

これらのピークを, 次のように解釈できる。1つ目は, 1884年の日清戦争と1904年の日露戦争の準備のための殖産興業政策による産業革命の局面で発生した公害と関連すると解釈できる。また, 2番目のピークは, 1929年から1931年の世界恐慌に引き続き第2次大戦に至るまでの貧困な食料事情による免疫力の低下に関わるものと解釈できる。さらに, 3番目のピークは, 1945年から1960年の復興景気とそれに続く1960年から1975年の高度経済成長期の公害と関連するものであると解釈できる。

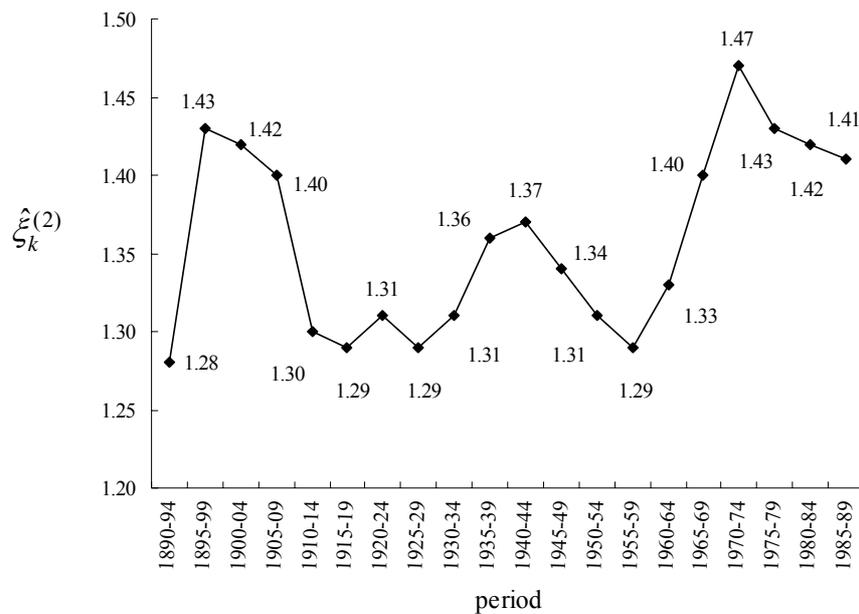


Figure 3.5.2-a : “[10,15) ~ [15,20) 歳へ加齢する際に受ける環境効果”の推定値 (乳癌死亡, 日本)。

$\alpha_i^{(8)}$ の推定値 :

Figure 3.5.2-bに $\alpha_i^{(8)}$ ($i = 1, \dots, 10$) の推定値を示す。ただし, i は25 + 5*i* 歳から29 + 5*i* 歳までの年齢区分に対応している。図から, 年齢効果が*i* について線形に増加していることが分かる。このことは, 日本の女性については, 乳癌死亡リスクが, 加齢に伴って指数関数的に増加することを示唆していると考えられる。

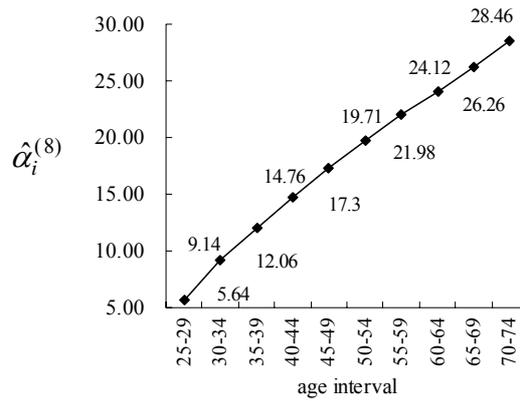


Figure 3.5.2-b : “1985-89年と同じ環境での年齢効果”の推定値（乳癌死亡，日本）。

3.6. 北欧4ヶ国の乳癌発生データの解析

3.6.1. 解析結果の概要

ここで解析する北欧諸国の女性の乳癌発生率データは，Rostgaard *et al.* (2001) から得たものである．AEモデルとAPCモデルをこれらのデータへ当てはめた解析結果の概要を残差尤度比とAICについてTable 3.6.1-a に示す．表から，スウェーデン，ノルウェー，フィンランドについては $\rho = 2, 3$ に対応するAEモデルが，デンマークについては $\rho = 2$ に対応するAEモデルがAICの意味でAPCモデルより良い当てはまりを示していることが分かる．（各国の解析結果について，最小のAICを * で示す．）日本におけるデータの解析結果を参考にすると， $\rho = 2$ つまり最低影響年齢を10歳と推測するのが妥当と思われる．環境効果と年齢効果の推定値の趨勢から得られる知見は，次の副節で与える．

Table 3.6.1-a : AE，APCモデルを北欧4ヶ国の女性の乳癌発生データへ当てはめた結果の概要．

Model	Denmark			Sweden		
	Deviance	d.f.	Difference of AIC from APC's	Deviance	d.f.	Difference of AIC from APC's
AE ($\rho = 6$)	200.3	50	125.1	338.9	50	37.6
AE ($\rho = 5$)	141.1	49	67.9	309.2	49	9.8
AE ($\rho = 4$)	96.8	48	25.6	297.6	48	0.2
AE ($\rho = 3$)	72.8	47	3.6	292.4	47	-3.0 *
AE ($\rho = 2$)	66.3	46	-0.9 *	291.7	46	-1.7
APC	65.2	45	0	291.4	45	0

Model	Norway			Finland		
	Deviance	d.f.	Difference of AIC from APC's	Deviance	d.f.	Difference of AIC from APC's
AE ($\rho = 6$)	96.2	50	40.2	218.1	50	78.7
AE ($\rho = 5$)	62.8	49	8.8	149.8	49	12.4
AE ($\rho = 4$)	52.2	48	0.1	138.5	48	3.0
AE ($\rho = 3$)	49.5	47	-0.6	131.0	47	-2.4 *
AE ($\rho = 2$)	46.7	46	-1.4 *	129.4	46	-2.0
APC	46.1	45	0	129.4	45	0

3.6.2. パラメータ推定値についての考察

$\xi_k^{(2)}$ の推定値 :

Figure 3.6.2-aに $\xi_k^{(2)}$ ($k = -13, \dots, 7$) の推定値を示す．ただし, k は1953 + 5 k 年から1953 + 5 k + 4年までの時代区分に対応している．図から, 1947年以前のフィンランドにおける環境効果が他の3ヶ国に比べて小さいが, 1948-52年の時代区分以降, 他の3ヶ国と同様の値になっていることが分かる．また, 1938-42年の時代区分に対応する環境効果が著しく低いレベルであることが分かる．

ここで, 1948-52年の時代区分はヘルシンキ・オリンピックが開催された年を, また, 1938-42年の時代区分は「冬戦争の奇跡」と呼ばれる年を含んでいることは興味深い．

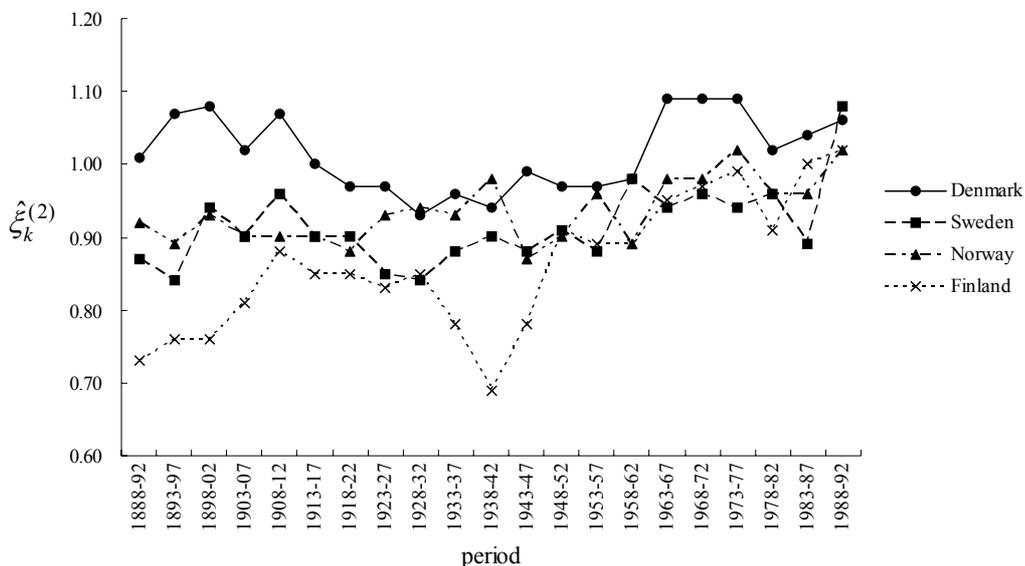


Figure 3.6.2-a : “[10,15)-[15,20) 歳へ加齢する際に受ける環境効果”の推定値 (乳癌発生, 北欧4ヶ国).

$\alpha_i^{(7)}$ と $\alpha_i^{(-13)}$ の推定値 :

Figure 3.6.2-bとFigure 3.6.2-cに $\alpha_i^{(7)}$ と $\alpha_i^{(-13)}$ ($i = 1, \dots, 11$) の推定値を示す．ただし, i は $20 + 5i$ 歳以上 $24 + 5i$ 歳以下の年齢区分に対応している．両方の図から, Figure 3.6.2-cのフィンランドの場合を除いて, 年齢効果は上に凸の増加曲線を描いていることが分かる．一方, Figure 3.6.2-cのフィンランドの場合は45-49歳の年齢区分以降わずかに減少している．また, 2つの図の比較から, 環境中のリスクが1888-92年の時代区分のものと同一であると仮定した場合には, フィンランドの女性についての年齢効果が4ヶ国の中で最も低いレベルである．一方, 100年後の1988-92年と同一であると仮定すると, スウェーデンに次いで最も高いレベルであることが分かる．

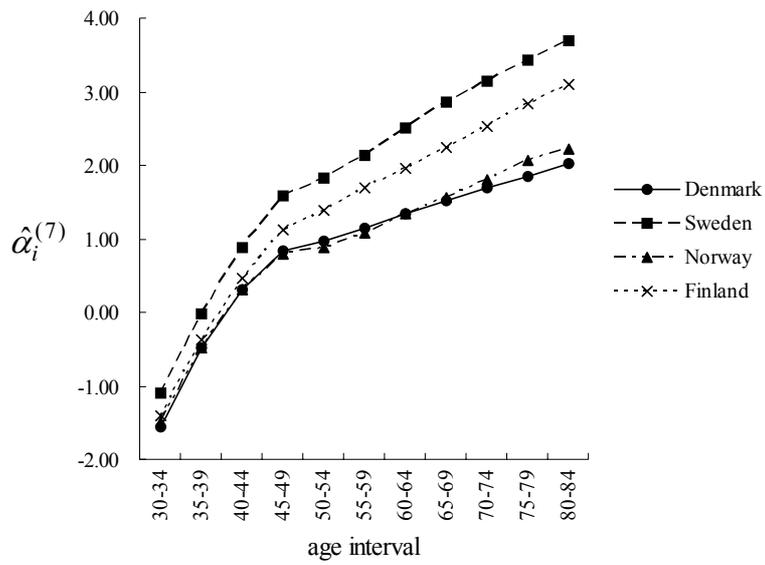


Figure 3.6.2-b : “1988-92年と同じ環境での年齢効果”の推定値（乳癌発生，北欧4ヶ国）.

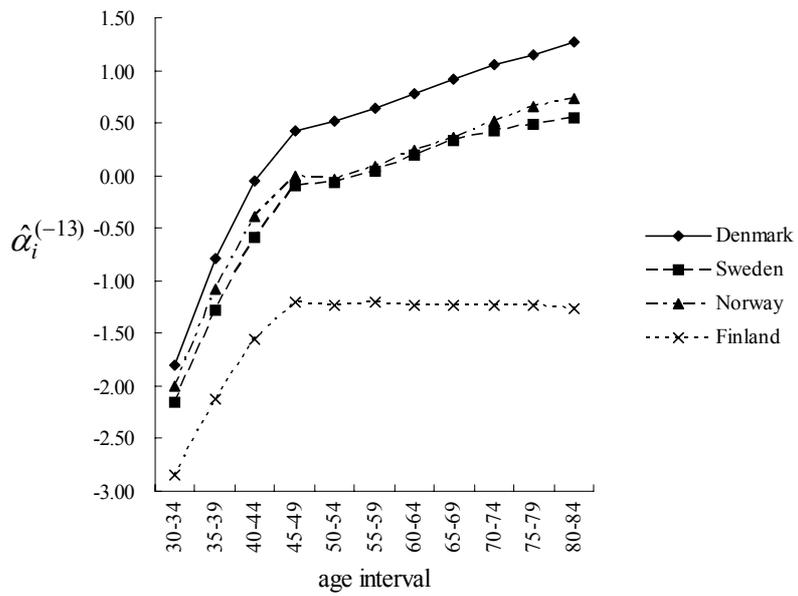


Figure 3.6.2-c : “1888-92年と同じ環境での年齢効果”の推定値（乳癌発生，北欧4ヶ国）.

3.7. 女子労働力人口の解析

3.7.1. 解析結果の概要

Table 3.7.1-a に1960年から1995年の5年毎の20歳以上64歳以下の女子労働力人口（年平均値）を，また，Table 3.7.1-b に対応する女子人口を5歳階級別に示す．ただし，20歳以上のデータであるから， $\phi = 4$ である．

Table 3.7.1-a：女子労働力人口（日本，単位：1000人）.

年齢階級別女子労働力人口		単位：千人							
年齢区分	年度								
	1960	1965	1970	1975	1980	1985	1990	1995	
20 - 24	277	325	374	301	273	289	326	361	
25 - 29	217	204	208	226	223	210	245	287	
30 - 34	216	205	201	204	255	229	200	213	
35 - 39	200	226	234	227	268	317	283	234	
40 - 44	171	204	235	245	268	313	366	314	
45 - 49	162	162	199	227	261	282	327	373	
50 - 54	124	140	153	182	216	244	268	302	
55 - 59	94	104	116	126	156	182	212	229	
60 - 64	68	68	77	89	97	116	138	153	

Table 3.7.1-b：年齢階級別女子人口（単位：万人）.

年齢階級別女子人口		単位：万人							
年齢区分	年度								
	1960	1965	1970	1975	1980	1985	1990	1995	
20 - 24	419	457	532	448	385	401	428	485	
25 - 29	411	421	455	534	446	383	394	434	
30 - 34	377	411	417	460	532	449	382	401	
35 - 39	327	375	407	419	458	529	445	388	
40 - 44	274	323	366	408	416	455	528	448	
45 - 49	256	270	318	369	404	414	452	529	
50 - 54	216	249	264	315	364	397	408	450	
55 - 59	184	207	237	259	309	357	393	405	
60 - 64	149	172	196	234	251	301	350	386	

Table 3.7.1-c に $\rho = 0$ のAEモデル（0歳以上で体験した時代環境の効果を含むAEモデル）， $\rho = 1$ （5歳以上で体験した時代環境の効果を含むAEモデル）， \dots ， $\rho = 4$ のAEモデル（20歳以上で体験した時代環境の効果を含むAEモデル），APモデル，ACモデルそしてAPCモデルを当てはめた解析結果の概要を示す．表中のAICの比較から以下のことが分かる．

(i) 全てのモデルの中でAPCモデルはAICが最も小さく， $\rho = 2$ のAEモデルは，APCモデルに次いでAICが小さい．

(ii) 全てのAEモデル ($\rho = 0, 1, \dots, 4$) のAICは, ACモデルより小さい.

(iii) $\rho = 0, 1, \dots, 3$ に対応するAEモデルは, APモデルよりAICが小さいが, $\rho = 4$ のAEモデルはAPモデルよりAICが大きい.

各AEモデルの残差尤度比の値を比較すると, $\rho = 2$ に対応するAEモデルが最も良い当てはまりを示している. このことから, 10歳以上の年齢時に体験した時代環境が女子の就労の意思に影響を及ぼすと判断することができる.

Table 3.7.1-c: AE, APC, AP, ACモデルを女子労働力人口データへ当てはめた結果の概要.

モデル	計画行列の階数	AIC	APCモデルとのAICの差	残差尤度比	残差の自由度
AE ($\rho = 4$)	24	348063847	314817	801 *E+3	48
AE ($\rho = 3$)	25	347881582	132552	619 *E+3	47
AE ($\rho = 2$)	26	347773270	24240	511 *E+3	46
AE ($\rho = 1$)	27	347828177	79147	566 *E+3	45
AE ($\rho = 0$)	28	347792996	43966	531 *E+3	44
APC	30	347749030	0	539 *E+3	42
AP	16	347916056	167026	487 *E+3	56
AC	24	348199432	450402	654 *E+3	48

3.7.2. 超過分散について

Table 3.7.1-c では, 全てのモデルにおいて, 残差尤度比の値が5%の有意水準で有意になっている. これは, 一般に2値の応答変数の応答確率について2項回帰モデルを当てはめた場合, 仮定した分布より応答変数の変動が大きく, モデルとの残差が過大になる, 超過分散と呼ばれる現象によるものと考えられる. 超過分散は, 今回使用したデータのように, サイズが 9×8 と比較的大きく, 各セルに対応する母集団が100万人単位であるようなデータにモデルを当てはめた場合に起きやすい問題である. 同様の問題はHolford (1983), Robertson, Boyle (1986), Tango (1987) などでも議論されている.

一般にこのような問題が発生した場合, 2項分布モデルに対してはベータ2項分布を, ポアソン回帰モデルに対しては特殊・ポアソン分布 (Wedderburn, 1974 およびBreslow, 1984参照) を用いる方法が提案されている. ところが, Robertson, Boyle (1998) は, 癌罹患率データに彼等のAPCモデルを当てはめた場合の残差に規則性があることを示しながら, 特殊・ポアソン分布を用いることが問題の解決に至らないと主張している. 彼等が示したのは, 年齢・時代区分データを古いコホートに対応するものと新しいコホートに対応するものの2つのグループに分け, グループ別に年齢階級毎の残差の平均をとった値である. そして, それらの変動が全く逆の動きをしていることを示し, このような場合には特殊・ポアソン分布の使用が適切ではないと主張したのである.

AEモデルはAPCモデルと関係が深いので、同様の現象が起きている可能性があると考えられる。そこで、Robertson, Boyle (1998) と同様に、年齢・時代区分データを古いコホートに対応するものと新しいコホートに対応するものの2つのグループにわけ、それぞれのグループでの残差（すなわち $r_{ij}^* / \tilde{N}_{ij} - \pi_{ij}(\hat{\theta})$ ）の年齢階級平均値を算出した。（Table 3.7.2-a 参照。）表において「*」でマークされた箇所に注目すると、同じ年齢階級では、古いコホートに対応するものと新しいコホートに対応するものの符号が逆になっていることが分かる。これは、Robertson, Boyle (1998) が癌罹患率データに彼等のAPCモデルを当てはめた場合の現象と同様の規則性である。したがって、我々の場合にも、ベータ2項分布を使用することは問題の解決には至らないと判断した。

Table 3.7.2-a：AEモデルの残差との新旧世代・年齢階級別の平均値。

年齢区分	年度								平均	
	1960	1965	1970	1975	1980	1985	1990	1995		
新しいコホート	20 - 24	-0.036	0.011	0.023	0.011	0.007	-0.005	-0.010	-0.001	0.000
	25 - 29		-0.020	-0.049	-0.041	-0.011	0.009	0.043	0.053	-0.002 *
	30 - 34			0.005	-0.016	-0.006	-0.004	-0.028	-0.033	-0.014 *
	35 - 39				0.001	-0.006	0.001	0.002	-0.039	-0.008 *
	40 - 44					0.003	0.017	0.007	0.007	0.008 *
	45 - 49						0.019	0.025	0.016	0.020 *
	50 - 54							0.016	0.017	0.017 *
	55 - 59							0.011		0.011 *
古いコホート	25 - 29	0.017								0.002 *
	30 - 34	0.068	0.014							0.012 *
	35 - 39	0.021	0.012	0.007						0.007 *
	40 - 44	-0.018	-0.013	0.000	-0.003					-0.007 *
	45 - 49	-0.015	-0.035	-0.012	-0.002	0.004				-0.015 *
	50 - 54	-0.025	-0.027	0.005	0.020	-0.010	0.003			-0.011 *
	55 - 59	-0.008	0.001	0.000	0.032	0.000	-0.024	-0.012		-0.006 *
	60 - 64	0.000	0.008	0.024	0.041	0.017	-0.015	-0.044	-0.031	0.000

3.7.3. パラメータの推定値についての考察

環境効果：

Figure 3.7.3-a に $\rho = 2$ の場合のAEモデル（ただし $\mu = 0$, $\alpha_2 - \alpha_1 = 0$ の制約を与えた）に含まれた環境効果の推定値を示す。（ただし、環境効果は値が大きいほど女子労働力率を大きくする。）これらから、以下のことが推察される。

(i) 1976～1990年の時代区分に対応する環境効果の推定値が他の時代区分のそれに比べて大きい。「1970年代の女性の社会進出」、「1980年代の女性の高学歴化」そして「1980年代の女性の職

場環境の改善」等の影響と思われる。

(ii) 1991～1995年の時代区分に対応する環境効果の推定値がその前の時代区分に対応する環境効果に比べて減少している。バブル経済崩壊に伴う不況に因る「女子大生の就職難」の影響と思われる。労働力人口は就業人口と失業人口の合計であるから、就職難、つまり求人の有無とは本来関係ないはずの統計値である。ところが、実際には、自分の望む職種に就けず離職した人々や、就職に夢を持たずに職を探さない人々が職業安定所（職安）に登録することは稀である。そこで、いわゆる「就職氷河期」の体験が、就業人口だけでなく労働力人口全体に影響すると推察される。

(iii) 1971～1975年の時代区分に対応する環境効果の推定値が前後の時代区分のそれに比べて小さい。この時代区分の間に起きた第1次、2次オイルショックに伴う不況の影響と思われる。

(iv) 1926～1935年および1946～1950年の時代区分に対応する環境効果の推定値が前後の時代区分のそれに比べて小さい。「世界恐慌」および「第2次大戦後の経済混乱」の影響と思われる。

(v) 1906～1910年の時代区分に対応する環境効果は「1960年に60～64歳の人間集団」つまり「1901～1905年生まれ世代」だけに関わる環境効果である。この1901～1905年に対応する環境効果が大きな値として推定されているのは、この集団の労働力率が「1965年に60～64歳の人間集団」や「1970年に60～64歳の人間集団」の労働力率と比べて大きいからである。このことは、労働者としては高年齢である60～64歳の年齢層が、1960年以降に日本の経済が発達するにつれて、働く必要が無くなった過程を表していると考えられる。

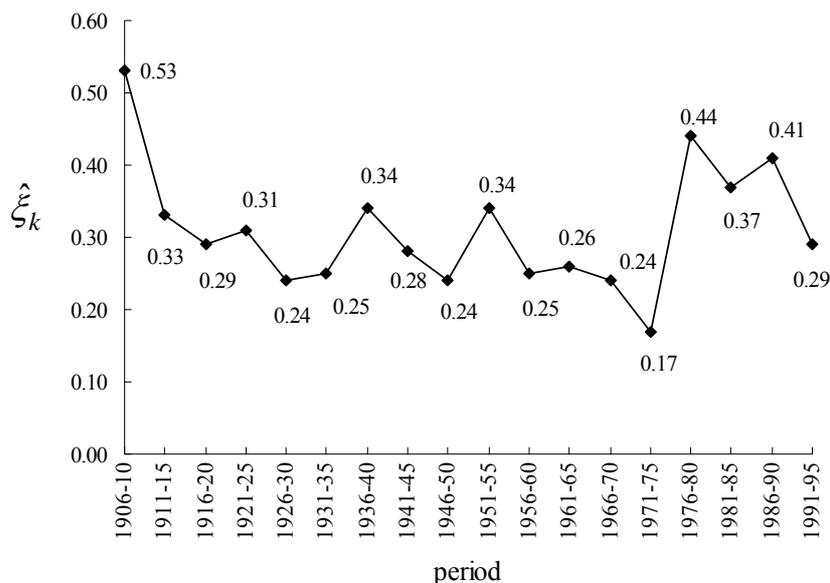


Figure 3.7.3-a : $\mu = 0$, $\alpha_2 - \alpha_1 = 0$ の制約下での環境効果の推定値 (女子労働力人口)。

年齢効果：

Figure 3.7.3-b に加齢効果（年齢効果の1階差分， $\mu = 0$ ， $\alpha_2 - \alpha_1 = 0$ を仮定）の推定値を示す．図から，加齢効果の推定値は，「M字カーブ」を呈している．これは，一般に知られている「20代前半の新卒採用 結婚，出産そして子育てに伴う離職 子育て終了による復職」による労働力率の変化傾向と一致している．

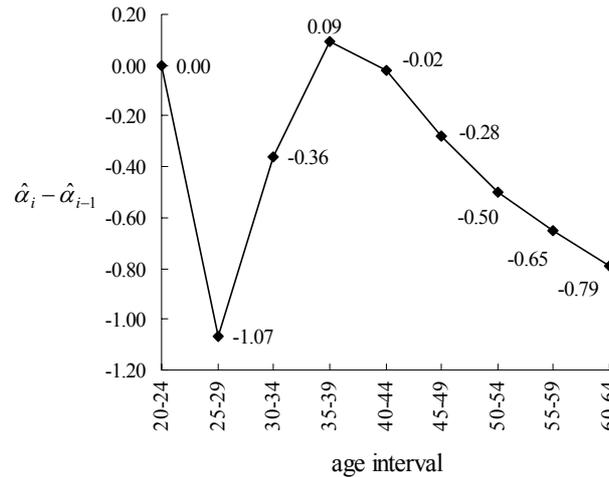


Figure 3.7.3-b： $\mu = 0$ ， $\alpha_2 - \alpha_1 = 0$ の制約下での加齢効果（年齢効果の1階差分）の推定値（女子労働力人口）．

3.8. 結び

3.8.1. 結論

年齢時代区別に観測された乳癌死亡率データを解析するためにAEモデルを提案し，このモデルが，計画行列の列空間がAPCモデルの計画行列の列空間に含まれるという意味で，APCモデルの副モデル（submodel）の1つであることを示した．しかしながら，AEモデルは，AICの意味で，日本と北欧4ヶ国で得られたデータに，APCモデルより良く当てはまることを示した．AEモデルを用いた日本のデータの解析から，日清・日露戦争，第2次大戦，復興景気と高度経済成長期の経済発展に伴う公害に関係すると思われる高いレベルの環境効果が推定された．また，北欧4ヶ国のデータ解析からは，フィンランドにおけるヘルシンキ・オリンピック後に悪化した環境に関係すると思われる高いレベルの環境効果と，「白い戦争の奇跡」を含む時代区分と関係する低いレベルの環境効果が推定された．

女子労働力人口データへ当てはめた結果から「10歳になった年度以降の年齢時に体験した時代環境が女子の就労の意思に影響を及ぼす」と推測された．また，可能な範囲に最低影響年齢を設定した全てのAEモデルは，AICの意味で，ACモデルよりデータへ良く当てはまり，いくつ

かの最低影響年齢に関しては，AEモデルはAPモデルよりデータへ良く当てはまることを示した．さらに，環境効果の推定値から，「1980年代の女性の高学歴化」，「1980年代の女性の職場環境の改善」，「バブル経済崩壊に伴う不況による女子大生の就職難」，「第1次，2次オイルショックに伴う不況」そして「世界恐慌および第2次大戦後の経済混乱」が女子労働力率に与えた影響を読み取ることができた．

3.8.2. リマーク (Remark)

Remark 3.1:

Hanayama (2001) は，多段階癌成長モデル (multi-stage cancer model; Armitage, Doll 1954, Moolgavkar, 1978参照) を参照しながら，年齢・時代区別に観測された癌死亡率データを解析するための simple two stage model (STSモデル) を提案している．(STSモデルについては，本論文の第4章で紹介する．) STSモデルは，環境効果を暴露されてからの経過時間に依って変化する暴露経過時間効果 (time-since-exposure effect) で重みを付けて累積するモデルである．そこで，STSモデルを乳癌死亡率データへ当てはめることも考えられるが，乳癌では年齢に関するリスクが知られているが，STSモデルは年齢に関わるパラメータを含んでいない．この意味で，乳癌死亡率データの解析には，年齢効果を持つAEモデルの方がSTSモデルより適していると考えられる．さらに，STSモデルは，非線形ポアソン回帰モデルのクラスに属するため，適切な初期値を与えないと最尤推定値を見つけることができない．実際，本論文で解析した日本の乳癌死亡率データにSTSモデルを当てはめても，最尤解を見つけることはできなかった．

Remark 3.2:

(3.2.2-a) や (3.4.1-a) で定義されているAEモデルでは，環境効果は，一律に重み1を課して累積されている．これは，時間的一定絶対リスク (UNSCEAR, 1994参照) の考えに基いている．しかしながら，この考えは，環境要因への暴露に関するリスクの趨勢を映し出すには，単純過ぎて不自然かもしれない．一方，次章で提案する Simple-two-stage model (STSモデル) では，暴露されてからの経過時間に依るパラメータを重みとして課して，環境効果を累積している．この意味で，次章のSTSモデルは，本章のAEモデルを発展させたものと考えられることができる．

Remark 3.3:

3.6節と3.7節で行った実データの解析では，Robertson, Boyle (1989) の提案している“出生区分の重複しない” APCモデル (副節2.3.2参照) も参照モデルに加えるべきかもしれない．(W_{ij}^0 と W_{ij}^1 へのデータの振り分けは，死亡者数を単純に2で割る方法を取ることによって，このモデルを我々のデータへ当てはめることができる．) しかし，異なるタイプのAPCモデルを参照モデルに

用いることで発生する混乱を避けるため、本論文では、彼等のAPCモデルは参照モデルとして採用しなかった。

Remark 3.4:

Holford *et al.* (1994) は、多段階癌成長モデル (multi-stage cancer model) を参考に、特殊なタイプのAPCモデルを提案している。そこでは、年齢効果について上に凸な曲線を仮定している。彼等の年齢効果に対する仮定は、北欧4ヶ国についてAEモデルを当てはめた結果とは整合するが、日本のデータへ当てはめた結果とは整合しない。

4. Simple-two-stage model (STSモデル)

この章ではSimple-two-stage model (STSモデル) を提案する。STSモデルは非線形ポアソン回帰モデルのクラスに属し、一般に性質が明らかになっている一般線形模型や一般加法模型のクラスに属さない。そこで、尤度関数と尤度方程式を示し、最尤推定量の存在性と一致性を示す。また、モデルの識別問題を議論し推定可能関数を示す。そして、日本人に関する全癌死亡データへSTSモデルを当てはめた結果を示し、残差尤度比とAIC両方の意味で、APCモデルよりもデータへ良く当てはまることを示す。さらに、日米のデータに対してSTSモデルを当てはめた結果を示し、国際比較を行う。

4.1. 背景と概要

4.1.1. 背景

最近の癌研究では、癌腫瘍は発癌性物質によって癌化の第1段階へ進められたたった1つの細胞が分裂して発現したもの、つまりクローンであるという考えが容認されている。しかし、人間の体が約 10^{14} 個の細胞からできていると考えると、DNAが損傷するのは極めて稀な出来事ではなければならない。(そうでなければ、ほとんどの人が、年少の間に癌になってなければならない。)この理由の1つとして、発癌物質によって癌化の第1段階へ進められた1つの細胞が成長して癌腫瘍になるまでには、潜伏期間中に、極めて生起確率の低い独立な変異を何回か起こさなくてはならないことが上げられる。つまり、発癌物質に暴露されると、沢山の細胞が癌化の第1段階へ進められるが、何十年か経って、その内の1つの細胞が最終段階の変異を起こした後、癌腫瘍が形成されると考えられるのである。(Voet, Voet 1995, pp. 1183-1184参照。)

実際、塩化ビニルの生産工場の労働者が20年後に塩化ビニルが原因の肝悪性血管内被腫にかかったことや、第2次大戦中に小島で毒ガス兵器の生産に従事させられていた人々が、20年後になって肺癌に襲われたことが知られている(Kuroki *et al.*, 1984, pp. 19-42参照)。つまり、ある時点で癌で死亡したり癌と診断されても、その原因は何年も前の生活環境や体質に起因しているのである。これらの事実を鑑みると、癌罹病率や死亡率データの解析から、人々が生きてきた過去の環境中の癌危険因子の総量、および、発癌物質によって癌化の第1段階へ進められた細胞が末期癌にまで成長してしまう危険強度を推定することは、今後の癌罹病や癌死亡の趨勢を予測するために重要である。

4.1.2. 概要

環境中の癌危険因子量の趨勢を映し出すためのモデルを提案するため、Moolgavkar, Venzon (1979)によるTwo-stage model (TSモデル)を参考にする。このモデルでは、発癌物質によ

て癌化の第1段階へ進まされた細胞の数が計数過程にしたがい、その細胞から発現した悪性腫瘍の大きさを出生死滅過程にしたがうと仮定するものである。TSモデルは毒物学や疫学のデータへ適用され、成果を上げている。(TSモデルの最近の研究結果については、例えば、Dewanji et al., 1999を参照。)

TSモデルは、発癌物質によって癌化の第1段階へ進まされた細胞の数と、その細胞から発現した悪性腫瘍の大きさを説明するものであるが、一方、我々が解析を試みるのは年齢・時代区分別(区間長は5歳と5年)に与えられた癌死亡者数と対応する人口である。このため、TSモデルを我々のデータへ直接適用することはできない。そこで、次の第4.2節では、年齢・時代区分データの解析へ適用可能なSimple-two-stage(STSモデル)を提案する。また、第4.3節では、STSモデルに含まれるパラメータの識別問題について議論し、さらに、パラメータについての最尤推定量の存在性と一致性を示す。

STSモデルは、分子生物学の分野で明らかにされている癌の発生・成長過程を参考に構築されたモデルであるが、転移については細かい記述を行っていない。つまり、原発性の癌であることを前提としている。したがって、全部位を対象とした癌死亡(全癌死亡)データへの当てはめが最も自然である。そこで、第4.4節では、APCモデルを参考モデルとしながら、日本で得られた全癌死亡データへSTSを当てはめた解析結果を示す。解析結果から、STSモデルはAPCモデルよりパラメータ数が少ないにも関わらず、残差尤度比の意味でAPCモデルより良い当てはまりを示すことが分かる。また、パラメータ推定値について、調査対象者が生きてきた時代の日本の歴史と関連付けながら考察を与える。

STSモデルは、癌死亡率を「環境中の発癌因子によって個体中の細胞が癌化の第1段階へ進まされてしまうハザード(環境危険強度)」と「その細胞が成長して末期癌に達するハザード(成長危険強度)」によって説明するモデルである。生活環境は地域によって異なるため、異なる地域で得られたデータにSTSモデルを当てはめ、環境危険強度の推定値から、地域による環境の違いを見ることは興味深いであろう。また、成長危険強度は免疫機能に関係するため(次節参照)、この値の地域差を見るのも興味深い。そこで第4.5節では、日米の全癌死亡データに対してSTSモデルを当てはめた結果を示し、国際比較を行う。

4.2. STSモデルの導出とその推定可能関数

4.2.1. 連続死亡強度に対するSTSモデル

まず、年齢 $u \in [0, a)$ と時点(年) $v \in [0, b)$ の全癌死亡強度関数を $\mu(u, v)$ と表し、以下を仮定する。

(i) 時刻 $t \in (-a, b)$ で生存している個人の体内の細胞が、微小時間区分 $[t, t + dt)$ に発癌物質によって癌化の第1段階へ進まされてしまう危険強度(環境危険強度と呼ぶ)は、時刻 t だけ

に依存する。(この仮定は、暗に、各個人が全て同じ数の正常細胞を持ち、それが年齢や時刻によって変化しないことを含んでいることに注意.)

(ii) 発癌物質によって時刻 $t \in (-a, b)$ に癌化の第1段階へ進まされた細胞が、微小時間区分 $[s, s+ds)$; $s \geq t$; $s \in [0, b)$; $s-t \in [0, a)$ に末期癌にまで成長してしまう危険強度(成長危険強度)は $s-t$ だけに依存する.

これらの仮定の下で、環境危険強度を $\xi(t)$ で、成長危険強度を $\psi(s-t)$ で表すと、全癌死亡強度 $\mu(u, v)$ は

$$\mu(u, v) = \int_{v-u}^v \xi(t) \psi(v-t) dt. \quad (4.2.1-a)$$

と表される. ただし、上のモデルでは、問題を簡単にするため、 $\psi(s-t)ds$ は $s-t$ 年前に癌化の第1段階へ進んだ細胞から、クローンとして派生した腫瘍が末期癌に成長してしまう確率を含んでいるものとする. また、時刻 t に源を発する悪性腫瘍は、他の時刻に源を発する悪性腫瘍と独立に成長すると仮定している.(Figure 4.2.1-a は、モデル(4.2.1-a)をレクシス・ダイアグラム上で表したものである.)

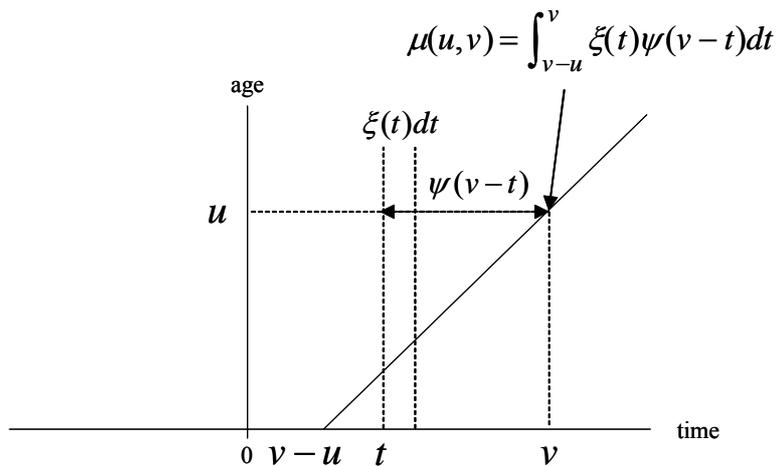


Figure 4.2.1-a: 連続死亡強度についてのSTSモデルの考え方.

発癌物質による第1段階への進展とその後の成長だけを考えれば、モデル(4.2.1-a)は癌による死亡強度を表すのに妥当なモデルである. また、 $\xi(t)$ と $\psi(s-t)$ は共に危険強度と定義されているため、正の値をとるはずである. しかしながら、最近では、癌は免疫性、つまり、悪性腫瘍にたいして免疫系が監視役を務め、悪性腫瘍はその成長過程で免疫系によって除外されることが知られている(Kripke, 1986を参照). 癌に対する免疫機能は、発癌物質によって細胞が癌化の第1段階へ進まされた時点で強化されると考えられるため、免疫機能の強さの期待値は、癌化の第1段階へ進まされた細胞の数に比例すると仮定できるであろう. つまり、時刻 v で u 歳

であるような個人の体内の癌に対する免疫力は $\int_{v-u}^v \xi(t)dt$ に比例すると仮定できるのである。

このような事情を鑑みると、モデル(4.2.1-a)は

$$\int_{v-u}^v \xi(t)\psi(v-t)dt = \int_{v-u}^v \xi(t)\psi^*(v-t)dt - \phi \int_{v-u}^v \xi(t)dt, \quad (4.2.1-b)$$

と解釈することができる。ただし、 $\phi > 0$ は比例定数を表し、 $\psi^*(v-t)$ は免疫機能を考えない場合の純粋な成長危険強度である。この解釈から、 $(u, v) \in [0, a) \times [0, b)$ 上の全ての点において $\mu(u, v) > 0$ が保たれる限り、 $\psi(v-t) = \psi^*(v-t) - \phi \in (-\infty, \infty)$; $v-t \in [0, a)$ 、つまり、 $\psi(s-t)$ は負の値を取りうると仮定する。

4.2.2. PCIモデルに対するSTSモデル

モデル(4.2.1-a)は、STSモデルの基本的な考え方を表しているが、我々が解析しようとするデータは年齢・時代区分別(区間長は5歳と5年)に与えられた癌死亡者数と対応する人口であるため、(4.2.1-a)を直接適用することはできない。

そこでここでは、年齢・時代区分データの解析にSTSモデルを適用できるように、PCIモデル(副節2.1.2参照)としてのSTSモデルを次のように定義する。

$$\mu_{ij} = \mu_{ij}(\boldsymbol{\theta}) \equiv \sum_{k=j-i+1}^j \xi_k \psi_{j-k+1}, \quad (4.2.2-a)$$

ただし、 $\boldsymbol{\theta} \equiv (\psi_1, \dots, \psi_I, \xi_{2-I}, \dots, \xi_J)' \in \Theta \equiv \{\boldsymbol{\theta} \mid \psi_1 > 0, \psi_i \in (-\infty, \infty); i = 2, \dots, I, \xi_k > 0; k = 2-I, \dots, J, \mu_{ij}(\boldsymbol{\theta}) > 0 \text{ for } i=1, \dots, I; j=1, \dots, J\}$ である。また、 $\mu_{1j}(\boldsymbol{\theta}) > 0$ と $\xi_j > 0$ を仮定するため、 $\psi_1 > 0$ である。

STSモデルとレクシス・ダイアグラム上のPCIモデルとの関係を理解するためにFigure 4.2.2-aを示す。また、モデル(4.2.2-a)は、 $I=3, J=4$ のとき、環境危険強度を回帰係数とした回帰モデル、または、環境危険強度を要因効果とした実験計画で用いられる分散分析型モデル(ANOVAモデル)の形で表すと、

$$\begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{14} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \\ \mu_{24} \\ \mu_{31} \\ \mu_{32} \\ \mu_{33} \\ \mu_{34} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \psi_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \psi_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \psi_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \psi_1 \\ 0 & \psi_2 & \psi_1 & 0 & 0 & 0 \\ 0 & 0 & \psi_2 & \psi_1 & 0 & 0 \\ 0 & 0 & 0 & \psi_2 & \psi_1 & 0 \\ 0 & 0 & 0 & 0 & \psi_2 & \psi_1 \\ \psi_3 & \psi_2 & \psi_1 & 0 & 0 & 0 \\ 0 & \psi_3 & \psi_2 & \psi_1 & 0 & 0 \\ 0 & 0 & \psi_3 & \psi_2 & \psi_1 & 0 \\ 0 & 0 & 0 & \psi_3 & \psi_2 & \psi_1 \end{pmatrix} \begin{pmatrix} \xi_{-2} \\ \xi_{-1} \\ \xi_0 \\ \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} \quad (4.2.2-b)$$

と表される．この表現を用いると，STSモデルは，回帰モデル，または，ANOVAモデルの計画行列にパラメータが入った形をしていることが分かる．

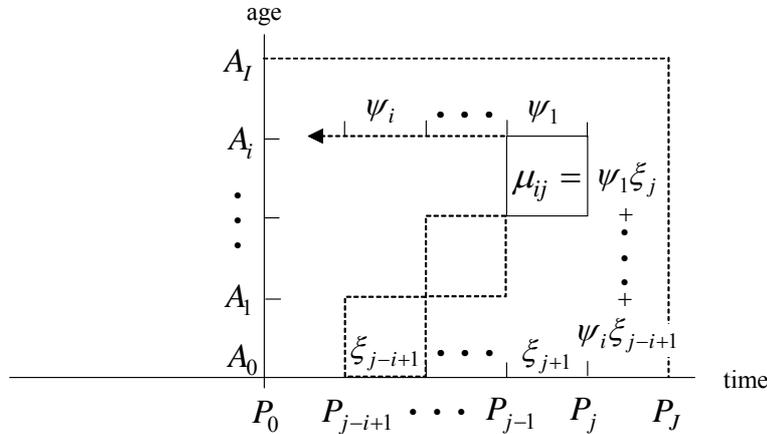


Figure 4.2.2-a : PCIモデルについてのSTSモデルの考え方 .

4.3. 尤度関数と識別問題

4.3.1. 尤度関数

副節2.2.1での議論から，PCIモデル版のSTSモデルの仮定の下で，レクシス・ダイアグラム上の断片 W_{ij} で生存する人々の生存時間の総和と癌死亡者数データ (x_{ij}^*, d_{ij}^*) に基づく部分尤度は

$$L^* \equiv \prod_{i=1}^I \prod_{j=1}^J \mu_{ij}^{d_{ij}^*}(\theta) \exp \left[- \{ \mu_{ij}(\theta) + \lambda_{ij} \} x_{ij}^* \right], \quad (4.3.1-a)$$

と表される ((2.2.1-b) 式参照) . ただし， $\mu_{ij}(\theta)$ は (4.2.2-a) 式のモデルを仮定した癌死亡強度で， λ_{ij} は癌以外の原因による死亡強度である . また，通常 x_{ij}^* は観測することができないため， W_{ij} に対応する人口年 (person-years at risk) によって近似される . (副節2.2.1のFigure 2.2.1-c 参照 .)

パラメータベクトル θ についての最尤推定値は，尤度関数 $L(\theta)$ を θ について最大化することによって得られる . ところが，尤度関数 $L(\theta)$ は

$$\mu(\theta) \equiv (\mu_{11}(\theta), \dots, \mu_{1J}(\theta), \mu_{21}(\theta), \dots, \mu_{2J}(\theta), \dots, \mu_{I1}(\theta), \dots, \mu_{IJ}(\theta))'$$

を経由した関数であるから，各

$$\begin{aligned} \mu &\equiv (\mu_{11}, \dots, \mu_{1J}, \mu_{21}, \dots, \mu_{2J}, \dots, \mu_{I1}, \dots, \mu_{IJ})' \\ &\in M \equiv \{ \mu \mid \mu = \mu(\theta), \theta \in \Theta \} \end{aligned}$$

に対して，方程式の系 $\mu = \mu(\theta)$ において， θ の値が一意に定まるか否かを確認する必要がある .

4.3.2. 識別問題

まず，識別問題を議論するため， $\theta \in \Theta$ に対するマッピング

$$q(w_1, \theta) \equiv \left(\frac{\psi_1}{w_1}, \dots, \frac{\psi_I}{w_1}, w_1 \xi_{2-I}, \dots, w_1 \xi_J \right)'$$

を考えよう．ただし， w_1 は正の定数である．このとき，各 $\mu = \mu(\theta) \in M$ に対して，方程式の系 $\mu(\theta) = \mu(q(w_1, \theta))$ が，任意の $w_1 > 0$ について成り立つ．このように，STSに含まれるパラメータには，識別不能問題が存在することが分かる．さらに，識別不能問題は $\mu \in M$ が，ある k ($2-I \leq k \leq 0$) に対して

$$\mu_{i,k+i-1} - \mu_{i-1,k+i-1} (= \xi_k \psi_i) = 0 ; i = a_k, \dots, b_k - 1 \quad (4.3.2-a)$$

を満たすとき，さらに，複雑となる．ただし (4.3.2-a) 式で， $a_k = \max\{1, 2-k\}$ ， $b_k = \min\{I, J-k+1\}$ である．また， $2-I \leq k \leq 0$ の場合には $a_k \geq 2$ である．この場合 $\theta \in \Theta$ に対するマッピング

$$q_k(w_1, w_2, \theta) \equiv \left(\frac{\psi_1}{w_1}, \dots, \frac{\psi_{a_k-1}}{w_1}, \mathbf{0}_{b_k-a_k}, \frac{\psi_{b_k}}{w_2}, \dots, \frac{\psi_I}{w_2}, w_2 \xi_{2-I}, \dots, w_2 \xi_{k-1}, w_1 \xi_k, \dots, w_1 \xi_J \right)'$$

を考えると，各 $\mu = \mu(\theta) \in M$ に対して，方程式の系 $\mu(\theta) = \mu(q_k(w_1, w_2, \theta))$ が，任意の $w_1 > 0$ および $w_2 > 0$ について成り立つことが分かる．ただし， $\mathbf{0}_l$ は要素が全て0の l 次元行ベクトルである．

Figure 4.3.2-a はレクシス・ダイアグラム上での a_k と b_k の意味を表している．さらに， $k = -1$ ， $I = 4$ ， $J = 3$ について $\mu \in M$ が (4.3.2-a) を満たすとき，STSモデルは，行列とベクトルを用いて

$$\begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \\ \mu_{31} \\ \mu_{32} \\ \mu_{33} \\ \mu_{41} \\ \mu_{42} \\ \mu_{43} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & \psi_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \psi_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \psi_1 \\ 0 & 0 & 0 & \psi_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \psi_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \psi_1 \\ 0 & 0 & 0 & \psi_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \psi_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \psi_1 \\ \psi_4 & 0 & 0 & \psi_1 & 0 & 0 \\ 0 & \psi_4 & 0 & 0 & \psi_1 & 0 \\ 0 & 0 & \psi_4 & 0 & 0 & \psi_1 \end{pmatrix} \begin{pmatrix} \xi_{-2} \\ \xi_{-1} \\ \xi_0 \\ \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} \quad (4.3.2-b)$$

と表される．上式で， ξ_1 ， ξ_2 ， ξ_3 には ψ_1 だけが， ξ_{-2} ， ξ_{-1} ， ξ_0 には ψ_4 だけが乗じられることから，識別不能な要素が少なくとも2つあることが分かる．

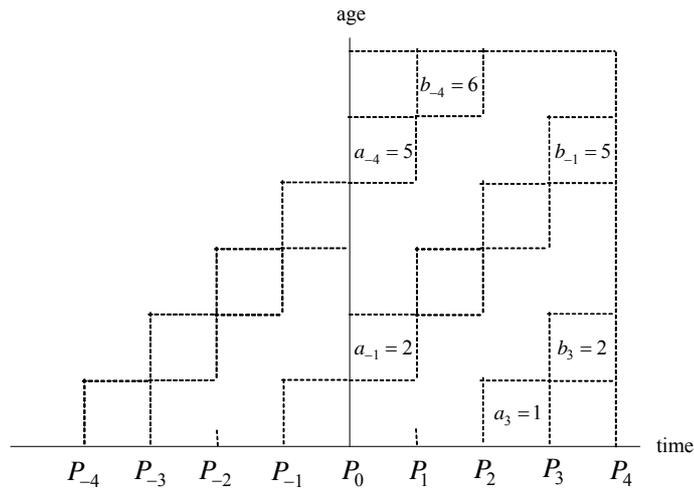


Figure 4.3.2-a : レクシス・ダイアグラム上で表した a_k と b_k .

複数の k に対して $\mu \in M$ が (4.3.2-a) を満たす場合も考えられるが, そのようなケースは極めて希である. 実際には, 1つの k について (4.3.2-a) が成り立つことも極めて希な出来事である. よって, 以後の議論では $\mu \in M$ がある k に対して (4.3.2-a) を満たす場合を除外し,

$$\mu \in M_0 \equiv \left\{ \mu \mid \begin{array}{l} \mu \in M, \text{ かつ, どの } k (= 2-I, \dots, J) \\ \text{ に対しても (4.3.2-a) が成り立たない. } \end{array} \right\}$$

の場合だけを考える. このとき, 各 $\mu = \mu(\theta) \in M_0$ について, 方程式の系 $\mu(\theta) = \mu(q(\psi_1, \theta))$ において $q(\psi_1, \theta)$ が一意に定まることを示すために, 次の命題を与える.

Proposition 4.3.2. $\theta_r \equiv (\psi_1^{(r)}, \dots, \psi_I^{(r)}, \xi_{2-I}^{(r)}, \dots, \xi_J^{(r)})' \in \Theta$; $r=1, 2$; $\mu(\theta_r) \in M_0$ と仮定する. このとき, $\mu(\theta_1) = \mu(\theta_2)$ であるための必要十分条件は $q(\psi_1^{(1)}, \theta_1) = q(\psi_1^{(2)}, \theta_2)$ である.

(証明は付録Cを参照.)

この命題から, $\mu \in M_0$ に対して, $\theta_* \equiv (\psi_2, \dots, \psi_I, \xi_{2-I}, \dots, \xi_J)' \in \Theta_* \equiv \{\theta_* \mid \psi_i \in (-\infty, \infty); i = 2, \dots, I, \xi_k > 0; k = 2-I, \dots, J, \mu_{ij}((1, \theta_*')) > 0 \text{ for } i = 1, \dots, I; j = 1, \dots, J\}$ が方程式の系 $\mu = \mu((1, \theta_*'))$ において一意に定まり, また, 任意の $\theta \in \Theta$ に対して, $\mu(\theta) = \mu((1, \theta_*'))$ を満たす $\theta_* \in \Theta_*$ が存在することが分かる.

4.4. 最尤推定量の存在性と一致性

4.4.1. 尤度方程式

副節4.3.2の議論から， θ_* についての尤度方程式を得る．

$$\frac{\partial \log L}{\partial \theta_*} \Big|_{\theta_* = \hat{\theta}_*} = \sum_{i=1}^I \sum_{j=1}^J \frac{\partial \mu_{ij}((1, \theta_*'))}{\partial \theta_*} \left(\frac{d_{ij}^*}{\mu_{i,j}((1, \theta_*'))} - x_{ij}^* \right) \Big|_{\theta_* = \hat{\theta}_*} = \mathbf{0}_{2I+J-2}, \quad (4.4.1-a)$$

ただし，

$$\frac{\partial \mu_{ij}((1, \theta_*'))}{\partial \theta_*} = \begin{cases} \left(\mathbf{0}_{I-1}', \mathbf{0}_{j-1+I-i}', \psi_i, \dots, \psi_2, 1, \mathbf{0}_{J-j}' \right)' & \text{if } i=1; j=1, \dots, J \\ \left(\xi_{j-1}, \dots, \xi_{j-i+1}, \mathbf{0}_{I-i-1}', \mathbf{0}_{j-1+I-i}', \psi_i, \dots, \psi_2, 1, \mathbf{0}_{J-j}' \right)' & \text{if } i=2, \dots, I; \\ & j=1, \dots, J \end{cases}$$

である．このように，尤度方程式 (4.4.1-a) の解 $\hat{\theta}_*$ は， $\theta_*^0 \equiv (\psi_2^0 / \psi_1^0, \dots, \psi_I^0 / \psi_1^0, \psi_1^0 \xi_{2-I}^0, \dots, \psi_1^0 \xi_J^0)'$ の一致推定量として得られる．ただし， $\xi_k^{(0)}$ ($k=2-I, \dots, J$) と $\psi_i^{(0)}$ ($i=1, \dots, I$) はパラメータの真値である．

尤度方程式 (4.4.1-a) は $D_{ij}^* (\equiv \sum_{n=1}^{N_{ij}} D_{ij}^n)$ が平均が $\mu_{ij}(\theta) x_{ij}^*$ であるようなポアソン分布に従うと仮定したものと同一形であるため，STSモデルはポアソン回帰モデルのクラスに属すると考えることができる．しかし，STSモデルでは， $\mu_{ij}(\theta)$ に一般線形模型 (McCullagh, Nelder, 1989参照) で使用されている結合関数 (link function) を用いて変換しても非線形であるため，非線形ポアソン回帰モデルの1つと考えられる．

4.4.2. 最尤推定量の存在性と一致性

STSモデルについて，最尤推定量の存在性と一致性を議論するために， $\psi_1 = 1$ と置いたSTSモデル，すなわち $\mu_{ij} = \mu_{ij}((1, \theta_*'))$ について，以下の性質を示そう．ただし， $Q_a(\theta_*^0)$ は十分に小さい半径 $a > 0$ の θ_*^0 の近傍を表す．

Property 4.4.2-a: すべての l, m, n に対し

$$\frac{\partial^3 \mu_{ij}((1, \theta_*'))}{\partial q_l \partial q_m \partial q_n} = 0$$

である．ただし，

$$q_l = \begin{cases} \psi_l & \text{if } 2 \leq l \leq I \\ \xi_{l-2I} & \text{if } I+1 \leq l \leq 2I+J-1. \end{cases}$$

Property 4.4.2-b: すべての l, m, n に対し, 次式を満たす定数 ρ が存在する.

$$\begin{aligned} & \sup_{\theta_* \in Q_a(\theta^0)} \left| \frac{\partial^3 \log \mu_{ij}((1, \theta_*'))}{\partial q_l \partial q_m \partial q_n} \right| \\ = & \sup_{\theta_* \in Q_a(\theta^0)} \left| 2 \frac{1}{\mu(1, \theta_*')^3} \frac{\partial \mu_{ij}((1, \theta_*'))}{\partial q_n} \frac{\partial \mu_{ij}((1, \theta_*'))}{\partial q_l} \frac{\partial \mu_{ij}((1, \theta_*'))}{\partial q_m} \right. \\ & - \frac{1}{\mu((1, \theta_*'))^2} \left\{ \frac{\partial^2 \mu_{ij}((1, \theta_*'))}{\partial q_m \partial q_l} \frac{\partial \mu_{ij}((1, \theta_*'))}{\partial q_n} + \frac{\partial^2 \mu_{ij}((1, \theta_*'))}{\partial q_n \partial q_l} \frac{\partial \mu_{ij}((1, \theta_*'))}{\partial q_m} \right. \\ & \left. \left. + \frac{\partial \mu_{ij}((1, \theta_*'))}{\partial q_l} \frac{\partial^2 \mu_{ij}((1, \theta_*'))}{\partial q_n \partial q_m} \right\} \right| < \rho. \end{aligned}$$

Property 4.4.2-c: 行列

$$\sum_{i=1}^I \sum_{j=1}^J \frac{\partial \mu_{ij}((1, \theta_*'))}{\partial \theta_*} \frac{\partial \mu_{ij}((1, \theta_*'))}{\partial \theta_*'} \Big|_{\theta_* = \theta_*^0}$$

が正値定符号である.

(Property 4.4.2-a についてはSTSモデルの形から直ちに確かめられる. Property 4.4.2-b と Property 4.4.2-c については, それぞれ, 付録Dと付録Eを参照.)

以上から, $\mu_{ij}((1, \theta_*'))$ は Proposition 2.2.1 の Condition 2.2.1-a および Condition 2.2.1-b を満たしていることが分かる. しがたって, STSモデルについて, 条件 Condition 2.2.1-c および Condition 2.2.1-d の下で, $\min\{M_{1-I}, \dots, M_{J-1}\} \rightarrow \infty$ のとき, 1 に近づく確率で, θ_*^0 へ一致する最尤推定量 $\hat{\theta}_*$ が存在することが分かる.

4.5. 全癌死亡データの解析

4.5.1. 解析結果の概要

Tables 4.5.1-a と Tables 4.5.1-b に, 日本の男性と女性について, 100,000人当たりの癌死亡者数と対応する人口年 (person-years at risk) を示す. ただし, i は $5(i-1)$ 歳から $5(i-1)+4$ 歳までの年齢区分に j は $1950+5(j-1)$ 年から $1950+5(j-1)+4$ 年までの時代区分に対応する. 尤度方程式 (4.4.1-a) を解いて最尤推定値を得るために x_{ij}^* は人口年を近似値として用いた.

APCモデルを参照モデルとした解析結果の概要を尤度比検定量 (likelihood ratio chi square statistics : G^2 statistics) を用いて Tables 4.5.1-c に示す . 表から , STSモデルは , APCモデルよりパラメータ数が少ないが , APCモデルより良い当てはまりを示している .

Table 4.5.1-a : 100,000人あたりの癌死亡者数と人口年 (person-years at risk)(男性 , 日本) .

<i>i</i>	Age group	<i>j</i>							
		1	2	3	4	5	6	7	8
		Time period							
		1950-54	1955-59	1960-64	1965-69	1970-74	1975-79	1980-84	1985-89
Cancer deaths rates per 100,000 for male									
1	0 - 4	5.84	7.41	8.44	8.38	7.56	7.02	5.44	4.18
2	5 - 9	3.19	4.34	4.92	5.57	5.94	5.55	5.07	4.21
3	10 - 14	3.23	4.21	4.84	4.91	4.76	4.68	4.38	3.98
4	15 - 19	4.36	5.46	6.84	6.94	6.66	6.64	5.88	5.45
5	20 - 24	5.23	6.40	7.84	8.63	7.68	7.59	6.69	5.83
6	25 - 29	8.52	10.04	11.28	12.16	12.33	11.23	9.81	8.71
7	30 - 34	15.78	18.14	19.19	19.65	19.11	18.62	17.30	14.55
8	35 - 39	31.58	33.77	35.12	35.65	34.57	32.22	30.06	27.84
9	40 - 44	62.83	64.91	64.45	61.89	64.02	59.54	53.39	52.16
10	45 - 49	119.43	124.53	123.39	115.85	111.03	116.55	111.62	96.52
11	50 - 54	215.90	224.33	222.79	213.74	199.11	196.40	216.55	202.16
12	55 - 59	360.61	377.18	384.48	370.04	353.50	341.28	349.11	372.33
13	60 - 64	540.91	582.72	594.97	605.28	572.02	550.67	548.62	561.82
14	65 - 69	745.54	812.65	861.01	887.14	886.08	843.72	851.58	840.48
15	70 - 74	849.09	1029.69	1109.78	1179.26	1203.24	1225.43	1231.38	1223.06
Person-years at risk for male									
1	0 - 4	27862490	21743330	20098563	21148581	24249505	24247653	20427838	18168200
2	5 - 9	24557426	27546491	21595331	19919011	21224644	24375267	24259227	20402800
3	10 - 14	23037387	24518800	27398477	21400170	20035006	21470013	24484462	24264500
4	15 - 19	21905567	22874369	24135763	26988341	21254341	20212716	21490921	24486700
5	20 - 24	20080815	21334415	21876266	23084297	26098558	21071815	19897017	21303800
6	25 - 29	16298898	19674382	20497656	21506028	22831569	26428687	20849252	19643500
7	30 - 34	12170240	16178239	19520898	20780254	21648837	23382513	26432380	20919300
8	35 - 39	11750105	12032498	15989208	19535509	20779995	21981146	23309728	26377600
9	40 - 44	11350955	11500750	11839344	15826666	19173406	20794047	21750879	23092600
10	45 - 49	10166848	11035515	11182804	11515594	15453868	19115962	20389696	21404900
11	50 - 54	9110275	9716249	10572674	10760903	11193891	15212119	18549231	19867200
12	55 - 59	7345661	8523703	9096182	9956469	10142700	10759581	14581018	17825200
13	60 - 64	5791567	6584793	7700574	8260089	9133039	9629318	10150902	13794700
14	65 - 69	4135919	4803056	5533993	6531867	7150260	8257671	8732457	9340900
15	70 - 74	2790291	3095776	3669566	4281994	5210330	5930548	6993106	7574100

Table 4.5.1-b : 100,000人あたりの癌死亡者数と人口年 (person-years at risk)(女性 , 日本).

i	Age group	j							
		1	2	3	4	5	6	7	8
		Time period							
		1950-54	1955-59	1960-64	1965-69	1970-74	1975-79	1980-84	1985-89
Cancer deaths rates per 100,000 for female									
1	0 - 4	4.38	5.57	6.52	6.69	6.48	5.56	4.50	3.58
2	5 - 9	2.25	3.16	3.72	4.09	4.42	4.18	3.51	2.96
3	10 - 14	2.54	3.32	4.05	4.24	4.13	3.85	3.52	3.29
4	15 - 19	3.57	4.27	5.40	5.42	5.16	4.51	4.38	3.70
5	20 - 24	5.15	6.25	7.35	7.66	7.21	6.28	5.69	4.57
6	25 - 29	11.57	12.59	13.88	13.64	13.51	12.38	10.53	8.87
7	30 - 34	26.22	26.65	27.00	26.29	24.55	22.98	21.03	17.15
8	35 - 39	55.27	49.95	47.31	44.45	41.33	37.71	34.82	32.64
9	40 - 44	96.49	89.68	80.97	73.21	68.46	61.41	55.14	53.05
10	45 - 49	147.70	137.45	130.98	116.95	107.64	98.04	87.82	80.92
11	50 - 54	209.19	199.71	189.97	179.34	159.11	147.62	136.57	125.24
12	55 - 59	280.19	274.15	267.05	253.62	233.74	213.45	197.08	182.64
13	60 - 64	369.97	370.00	364.89	358.24	332.59	303.79	283.95	261.97
14	65 - 69	478.67	496.82	493.53	490.93	467.36	435.49	408.73	375.19
15	70 - 74	531.08	611.11	641.23	654.10	636.94	613.24	588.24	548.54
Person-years at risk for female									
1	0 - 4	26727967	20758411	19206870	20235902	22996584	22993092	19386242	17298200
2	5 - 9	23784239	26427101	20698304	19191281	20305359	23173739	23061560	19443700
3	10 - 14	22500530	23746017	26370061	20666237	19242278	20505381	23248903	23123500
4	15 - 19	21482101	22430150	23482775	26328547	20644054	19381280	20511499	23289900
5	20 - 24	20139727	21301828	22032184	23176392	26141367	20659633	19327795	20459600
6	25 - 29	17891222	19849946	20768704	21885801	23167549	26197534	20572705	19169700
7	30 - 34	15062997	17659693	19503907	20788076	21912854	23294193	26139249	20594900
8	35 - 39	13517968	14877628	17436822	19417030	20705908	21986646	23235423	26057100
9	40 - 44	12282025	13257580	14632786	17209736	19150414	20750190	21862160	23125100
10	45 - 49	10177701	12007658	12972755	14343217	16957849	19173066	20546109	21693400
11	50 - 54	8931393	9867241	11645716	12626095	14203084	16858042	18941106	20274000
12	55 - 59	7273368	8537811	9448025	11135540	12171004	13870606	16520312	18608900
13	60 - 64	6128328	6814800	8072043	8888370	10678880	11941041	13489649	16123500
14	65 - 69	4926796	5400963	6135409	7256444	8232432	10207821	11338149	12955400
15	70 - 74	3804317	4048886	4529238	5185567	6408421	7466025	9314015	10555200

Table 4.5.1-c : STS , APCモデルを日本の癌死亡データへ当てはめた結果の概要 .

Model	G ² statistics	Degree of freedom
For male:		
STS	1518	84
APC	2502	78
For female:		
STS	977	84
APC	2203	78

4.5.2. パラメータ推定値についての考察

成長危険強度 ψ_i :

成長危険強度の推定結果をTable 4.5.2-a とFigure 4.5.2-a, Figure 4.5.2-b に示す。ただし, i ($= 1, \dots, 15$) は発癌物質へ暴露されてから $5(i-1)$ 年から $5(i-1) + 4$ 年経ったことに対応している。Figure 4.5.2-a から, $i=9$ の時点で i の増加に伴う傾斜が急激に変化していることが分かる。この事実から, 男性の体内では, 発癌物質によって癌化の第1段階に進まされた細胞が, 45年以上体内で生き延びると, その後急激に成長して末期癌に至る危険が高くなると推察される。一方, Figure 4.5.2-bから, 女性については, 男性の場合のような急激な変化は見られない。

Table 4.5.2-a : 成長危険強度の推定値 (日本)。

i	Years since a cell had been primed	Estimate	
		Male	Female
1	0 - 4	1.00	1.00
2	5 - 9	-0.27	-0.33
3	10 - 14	-0.05	0.02
4	15 - 19	0.21	0.14
5	20 - 24	0.12	0.27
6	25 - 29	0.51	0.93
7	30 - 34	1.03	1.76
8	35 - 39	1.98	2.50
9	40 - 44	3.48	3.50
10	45 - 49	6.86	4.68
11	50 - 54	12.19	5.93
12	55 - 59	19.37	7.18
13	60 - 64	26.68	9.34
14	65 - 69	36.26	12.45
15	70 - 74	43.69	16.27

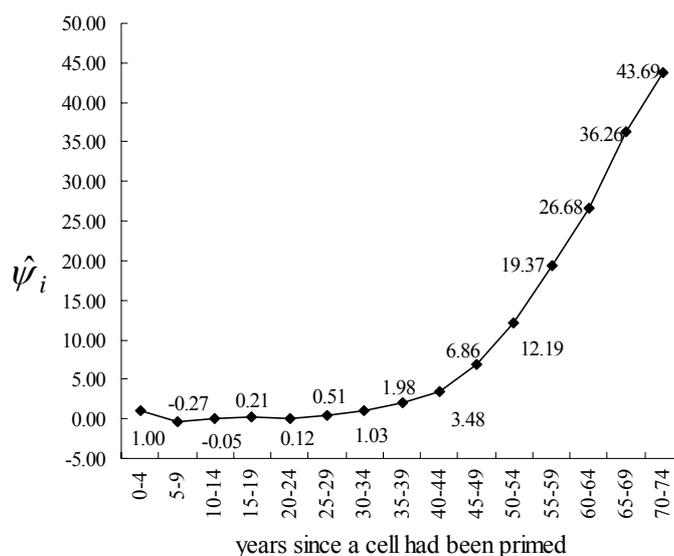


Figure 4.5.2-a : 成長危険強度の推定値 (男性, 日本)。

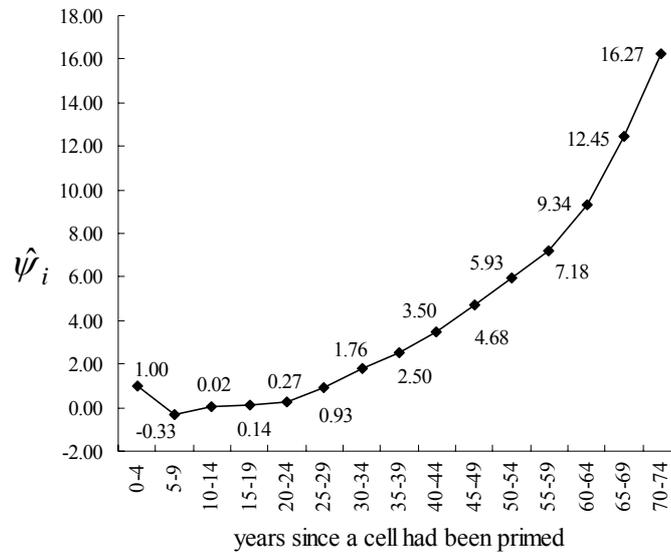


Figure 4.5.2-b : 成長危険強度の推定値 (女性, 日本) .

環境危険強度 ξ_k :

環境危険強度の推定結果をTable 4.5.2-b とFigure 4.5.2-c , Figure 4.5.2-d に示す . ただし , k ($= -13, \dots, 8$) は1950 + 5($k-1$)年から1950 + 5($k-1$) + 4の時代区分に対応している . Figure 4.5.2-c から , 男性の場合 , 環境危険強度には , 1905-09年の時代区分 , 1930-34年の時代区分 , そして , 1965-69年の時代区分に対応した3つの主なピークがあることが分かる .

これらは次のように解釈できる . 1つ目は , 1884年と1904年の日清・日露戦争に備えた , 明治政府の殖産興業政策の局面に伴う環境汚染に関連している . また , 2つ目は , 昭和初期の「世界恐慌」から「敗戦」に至る時代の貧困な食料事情が人々の免疫機能を弱ませたことと関係している . さらに , 3つ目は , 1945年から1960年に至る「戦後復興」と1960年から1975年の「高度経済成長」の時代の環境汚染に関係している . 一方 , 女性については , 男性の場合の2つ目に対応するピークが見当たらない . このことから , 男性だけが , 「世界恐慌」から「敗戦」に至る時代の貧困な食料事情に影響されたことが推察される .

Table 4.5.2-b : 環境危険強度の推定値 (日本).

k	period	Estimate	
		Male	Female
-13	1880 - 84	2.64	3.80
-12	1885 - 89	5.58	8.10
-11	1890 - 94	6.44	9.58
-10	1895 - 99	7.44	10.20
-9	1900 - 04	7.80	10.56
-8	1905 - 09	8.49	10.81
-7	1910 - 14	8.18	10.56
-6	1915 - 19	8.24	10.49
-5	1920 - 24	7.55	9.13
-4	1925 - 29	6.92	8.21
-3	1930 - 34	8.61	7.79
-2	1935 - 39	8.38	7.16
-1	1940 - 44	6.58	5.97
0	1945 - 49	7.78	6.82
1	1950 - 54	5.55	4.78
2	1955 - 59	6.23	4.86
3	1960 - 64	7.11	5.69
4	1965 - 69	7.97	6.39
5	1970 - 74	7.86	6.55
6	1975 - 79	7.54	6.07
7	1980 - 84	6.66	5.28
8	1985 - 89	5.57	4.24

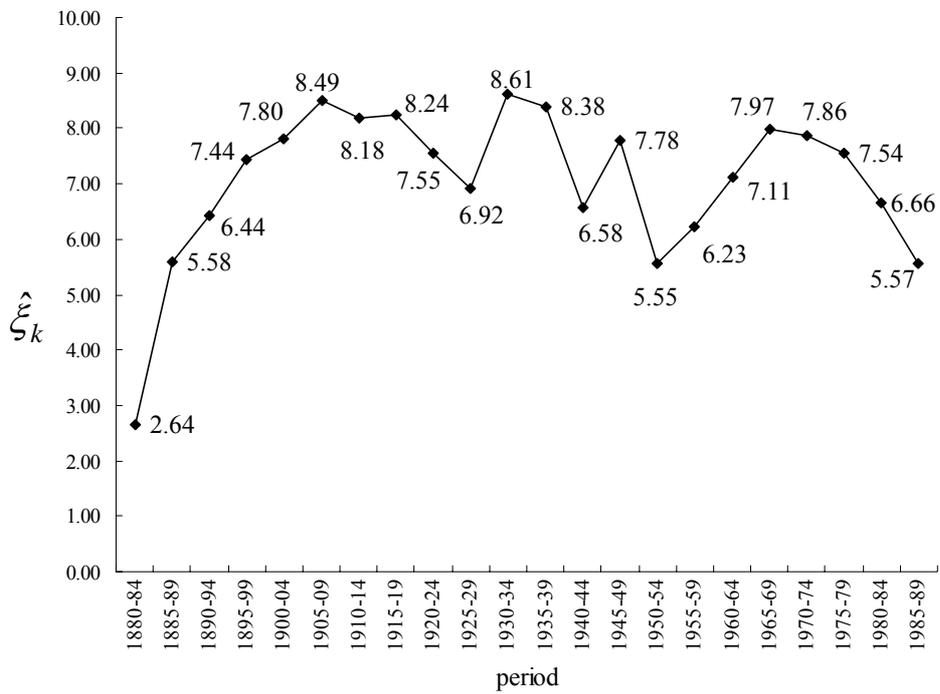


Figure 4.5.2-c : 環境危険強度の推定値 (男性, 日本).

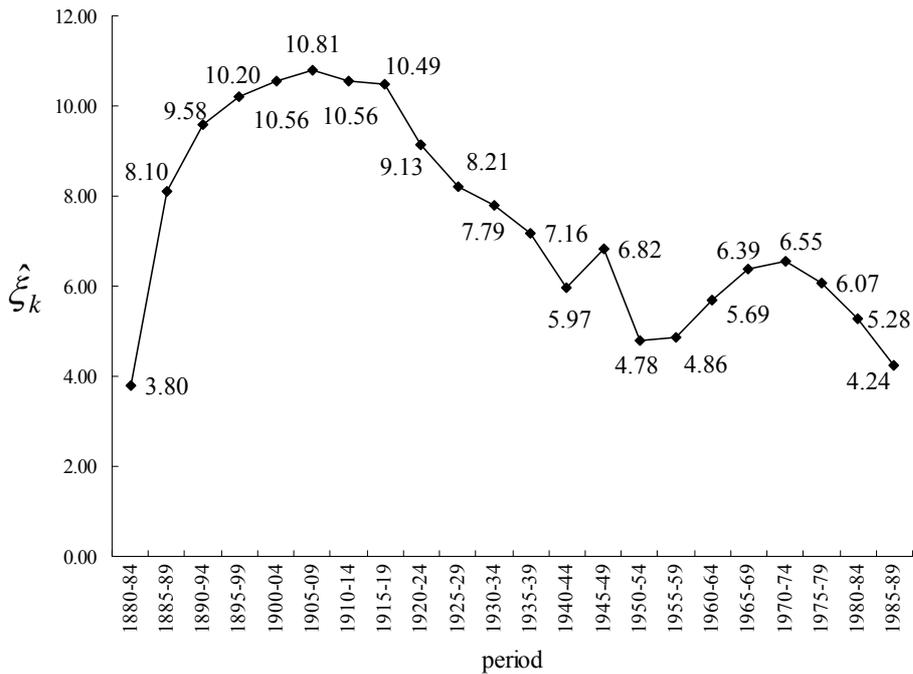


Figure 4.5.2-d : 環境危険強度の推定値 (女性, 日本).

4.6. 日米癌死亡データ解析結果の考察

4.6.1. 解析結果の概要

Table 4.6.1-a とTable 4.6.1-b に, 日本国および米国について, 100,000人当たりの癌死亡者数と千人単位の人口年 (person-years at risk) をそれぞれ示す. ただし, i ($= 1, \dots, 8$) は $10(i-1)$ 歳以上 $10(i-1)+9$ 歳以下の年齢階級に, j ($= 1, \dots, 5$) は $1950+10(j-1)$ 年に対応する.

STSモデルおよびAPCモデルを日米両国の癌死亡データに当てはめた結果の概要を, 残差尤度比 (G^2 統計量) を用いてTable 4.6.1-c に示す. (残差尤度比は, 表中に示された自由度を持つカイ2乗分布にしたがう. この値が小さいほどデータへの当てはまりが良い.) 表から分かるように, STSモデルは日本国の癌死亡データにはAPCモデルより良い当てはまりを示しているが, 米国の癌死亡データについては, 特に女性について, APCモデルより当てはまりが悪い. データによって当てはまりの優劣に差があるものの, STSモデルはAPCモデルと同様に十分有効なモデルであると考えられる.

Table 4.6.1-a : 癌死亡者数と対応する人口年 (person-years at risk)(米国).

<i>j</i>	year	<i>i</i>							
		1	2	3	4	5	6	7	8
		age interval							
		0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79
Cancer deaths rates per 100000 for male									
[Data are based on the National Vital Statistics System]									
1	1950	10.0	12.4	25.9	48.5	131.4	383.1	903.2	2003.6
2	1960	9.7	13.2	28.5	55.3	142.6	416.0	998.8	2152.8
3	1970	7.1	12.0	22.8	54.2	161.0	450.6	1100.8	2404.0
4	1980	4.9	8.7	16.3	39.4	149.2	469.5	1127.6	2500.2
5	1990	3.5	6.3	13.4	34.5	118.0	442.8	1170.6	2479.5
Person-years at risk for male in 1000									
[Data are based on decennial census updated by data from mutiple sources]									
1	1950	14424	11647	11258	11093	9622	7676	5361	2766
2	1960	19345	14968	11543	11467	10924	8815	6327	3571
3	1970	19126	19155	14884	11724	11215	9996	7115	3937
4	1980	17290	19637	19900	15476	11790	10581	8455	4812
5	1990	18585	18443	20240	20037	15371	11094	8931	5826
Cancer deaths rates per 100000 for female in U.S.A.									
[Data are based on decennial census updated by data from mutiple sources]									
1	1950	8.4	9.9	24.8	67.1	183.0	414.6	844.5	1713.3
2	1960	7.5	9.7	22.7	64.3	166.6	379.0	750.7	1508.1
3	1970	5.7	8.4	17.1	57.4	165.8	373.5	725.5	1447.1
4	1980	3.5	5.9	12.8	42.2	145.5	375.0	741.1	1399.7
5	1990	2.9	4.8	11.8	38.3	124.8	363.6	801.1	1466.3
Person-years at risk for female in 1000 in U.S.A.									
[Data are based on the National Vital Statistics System]									
1	1950	13900	11563	11672	11513	9776	7680	5494	3044
2	1960	18710	14776	11877	11983	11360	9215	6959	4245
3	1970	18401	18938	15290	12274	11939	10909	8398	5341
4	1980	16530	19078	19884	15883	12428	11671	10188	6844
5	1990	17721	17611	19847	20261	15875	11991	10649	8203

Table 4.6.1-b : 癌死亡者数と対応する人口年 (person-years at risk)(日本).

<i>j</i>	year	<i>i</i>							
		1	2	3	4	5	6	7	8
		age interval							
		0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79
Cancer deaths rates per 100000 for male									
1	1950	4.3	3.5	5.9	22.1	88.8	275.1	602.5	754.7
2	1960	6.6	5.4	8.4	25.0	94.5	303.8	703.0	1127.8
3	1970	7.1	6.1	10.6	28.3	84.6	279.3	722.6	1254.1
4	1980	5.8	5.6	8.8	24.4	86.7	263.7	696.8	1398.9
5	1990	3.5	4.6	7.1	19.9	71.6	274.6	690.3	1391.5
Person-years at risk for male in 1000									
1	1950	10544	8718	6658	4736	4218	3098	1905	808
2	1960	8715	10298	8220	6510	4531	3843	2465	1070
3	1970	8623	8514	9770	8262	6304	4169	3139	1489
4	1980	9446	8759	8445	9957	8155	6025	3667	2158
5	1990	7127	9466	8473	8392	9805	7773	5424	2753
Cancer deaths rates per 100000 for female									
1	1950	3.2	2.7	8.0	40.7	126.9	246.2	418.2	476.8
2	1960	5.0	4.5	10.0	37.1	110.3	226.8	423.6	667.5
3	1970	5.5	4.8	10.4	33.1	90.0	201.0	402.8	697.1
4	1980	4.3	4.3	9.1	28.2	73.3	170.2	355.2	695.1
5	1990	3.0	2.9	5.8	24.8	63.8	146.6	300.3	622.7
Person-years at risk for female in 1000									
1	1950	10184	8550	7253	5514	4270	3040	2169	1159
2	1960	8334	10028	8308	7046	5305	4000	2627	1448
3	1970	8223	8283	9862	8237	6841	5010	3544	1905
4	1980	8979	8356	8316	9903	8199	6727	4723	2885
5	1990	6779	8997	8225	8267	9802	8010	6403	4071

Table 4.6.1-c : 日米の癌死亡データへSTS , APCモデルを当てはめた結果の概要 .

Model	Data	G ² statistics	Degree of freedom
For male:			
STS	U.S.A.	208	21
APC	U.S.A.	116	17
STS	Japan	413	21
APC	Japan	561	17
For female:			
STS	U.S.A.	1886	21
APC	U.S.A.	117	17
STS	Japan	303	21
APC	Japan	501	17

4.6.2. パラメータ推定値についての考察

環境危険強度 ξ_k の推定値

Table 4.6.2-a (米国) および Table 4.6.2-b (日本) に環境危険強度 ξ_k の推定値を、そして Figure 4.6.2-a (米, 男性), Figure 4.6.2-b (米, 女性), Figure 4.6.2-c (日本, 男性), Figure 4.6.2-d (日本, 女性) にそのグラフを示す。ただし, k ($= -6, \dots, 5$) は $1941 + 10(k - 1)$ 年から $1941 + 10(k - 1) + 9$ 年までの時代区分に対応する。

図から分かるように, 米国の男性および女性, 日本国の男性に共通して, 1931年から1940年の環境効果の値が相対的に大きい。(ただし, 米国の女性については1921年から1930年の環境効果の方が僅かに大きい。) さらに, 日本の男性と女性に共通して, 1961年から1970年を中心としたピークが存在する。

1931年から1940年の環境効果については, 1930年代の「世界恐慌」による劣悪な栄養事情および社会不安と関連するのではないかと考えることができる。一方, 日本の女性についてはこれに対応するピークは存在しない。このことから, 日本女性は, 日本男性, 米国男性および女性に比べて「劣悪な栄養事情および社会不安に影響されにくい」と言えるのかもしれない。

また, 日本の男性と女性に共通した1961年から1970年を中心としたピークについては, いわゆる「高度経済成長期」における環境汚染と関連するのではないかと考えられる。一方, 米国の男性および女性についてはこのようなピークは存在しない。このことから, 日本で1970年前後の環境に暴露された人々の癌死亡が, 米国のそれと比べてどのように推移するか注目すべきと考える。

Table 4.6.2-a: 環境危険強度の推定値 (米国)。

k	period	Estimate (US)	
		Male	Female
-6	1871 - 80	0.062	0.096
-5	1881 - 90	0.065	0.088
-4	1891 - 0	0.078	0.087
-3	1901 - 10	0.087	0.083
-2	1911 - 20	0.094	0.090
-1	1921 - 30	0.108	0.105
0	1931 - 40	0.110	0.103
1	1941 - 50	0.084	0.080
2	1951 - 60	0.087	0.070
3	1961 - 70	0.079	0.056
4	1971 - 80	0.055	0.037
5	1981 - 90	0.041	0.036

Table 4.6.2-b : 環境危険強度の推定値 (日本).

k	period	Estimate (JP)	
		Male	Female
-6	1871 - 80	0.156	0.187
-5	1881 - 90	0.449	0.678
-4	1891 - 0	0.546	0.773
-3	1901 - 10	0.628	0.830
-2	1911 - 20	0.620	0.771
-1	1921 - 30	0.568	0.625
0	1931 - 40	0.636	0.544
1	1941 - 50	0.468	0.396
2	1951 - 60	0.558	0.459
3	1961 - 70	0.714	0.530
4	1971 - 80	0.626	0.481
5	1981 - 90	0.411	0.277

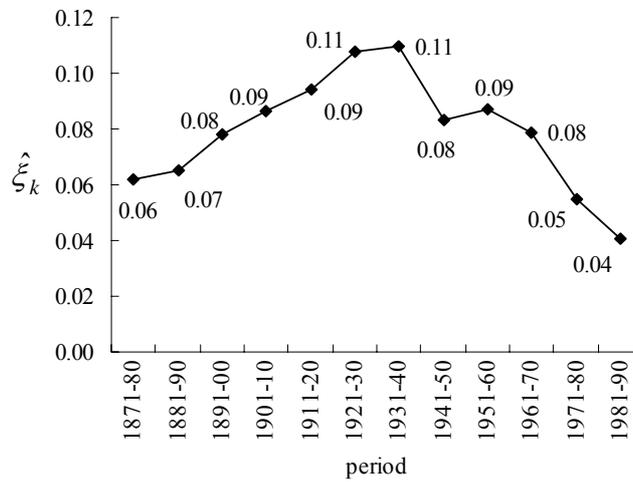


Figure 4.6.2-a : 環境危険強度の推定値 (男性, 米国).

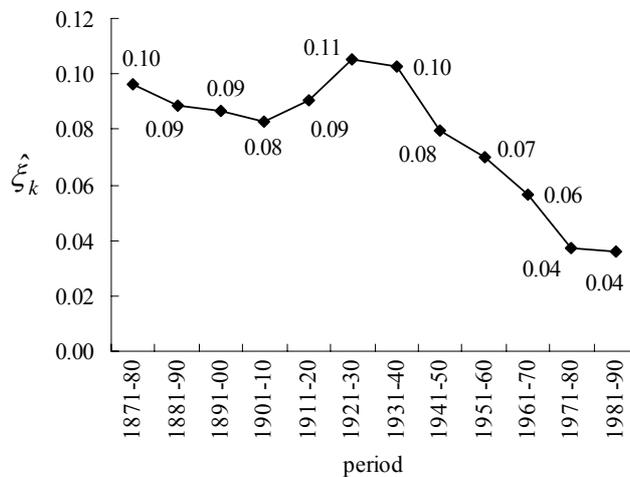


Figure 4.6.2-b : 環境危険強度の推定値 (女性, 米国).

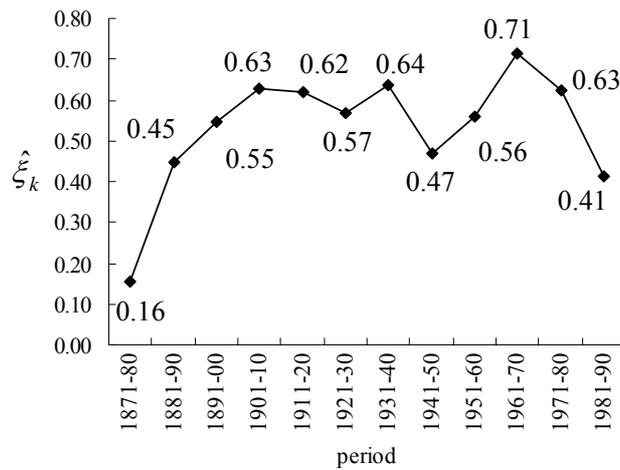


Figure 4.6.2-c : 環境危険強度の推定値 (男性, 日本).

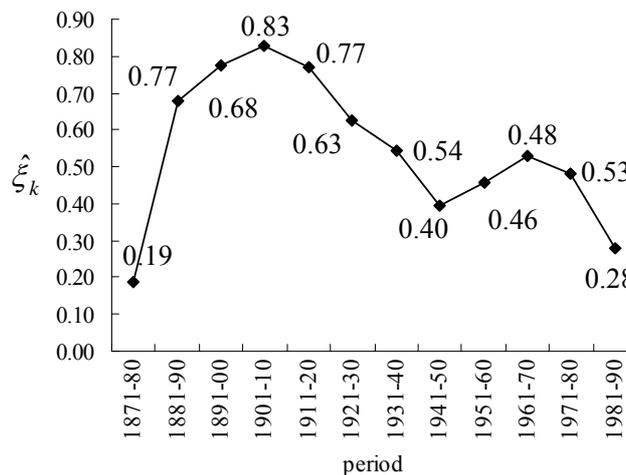


Figure 4.6.2-d : 環境危険強度の推定値 (女性, 日本).

成長危険強度 ψ_{j-k+1} の推定値

Table 4.6.2-c (米国) および Table 4.6.2-d (日本) に成長危険強度 ψ_{j-k+1} の推定値をそして Figure 4.6.2-e (米国男性), Figure 4.6.2-f (米国女性), Figure 4.6.2-g (日本国男性), Figure 4.6.2-h (日本国女性) にそのグラフを示す. ただし, $i (= 1, \dots, 8)$ は, 人々が時代環境に暴露されてからの経過時間 (10($i-1$) ~ 10 i 年間) に対応する.

図から分かるように日本の男性と米の男性, 日本の女性と米の女性のグラフの形状はよく似ている. つまり, 日米共に男性については, 40~49歳を境に成長危険強度 ψ_{j-k+1} の傾向に急激な変化が見られるが, 女性については日米共に男性に見られるような急激な変化は見られな

い．このことは「日米を問わず，男性については，体内で免疫システムを逃れて40年以上生き延びた異常細胞は，その後急激に成長することができる」ということを示唆しているのではないかと考えられる．

Table 4.6.2-c：成長危険強度の推定値（米国）．

i	Years since a cell had been primed	Estimate (US)	
		Male	Female
1	0 - 9	0.01	0.01
2	10 - 19	0.00	0.00
3	20 - 29	0.01	0.01
4	30 - 39	0.02	0.03
5	40 - 49	0.08	0.08
6	50 - 59	0.22	0.15
7	60 - 69	0.43	0.23
8	70 - 79	0.62	0.29

Table 4.6.2-d：成長危険強度の推定値（日本）．

i	Years since a cell had been primed	Estimate (JP)	
		Male	Female
1	0 - 9	0.01	0.01
2	10 - 19	0.00	0.00
3	20 - 29	0.00	0.01
4	30 - 39	0.03	0.04
5	40 - 49	0.10	0.09
6	50 - 59	0.32	0.15
7	60 - 69	0.71	0.24
8	70 - 79	1.08	0.40

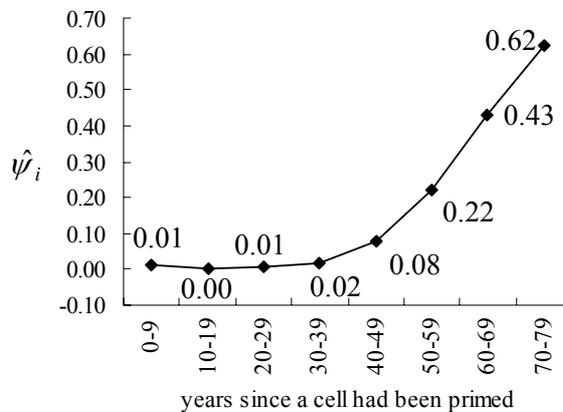


Figure 4.6.2-e：成長危険強度の推定値（男性，米国）．

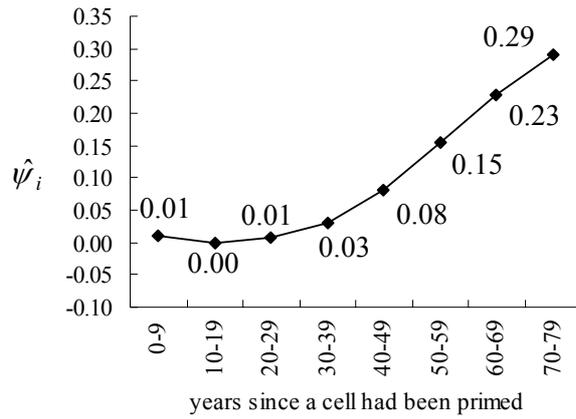


Figure 4.6.2-f: 成長危険強度の推定値 (女性, 米国).

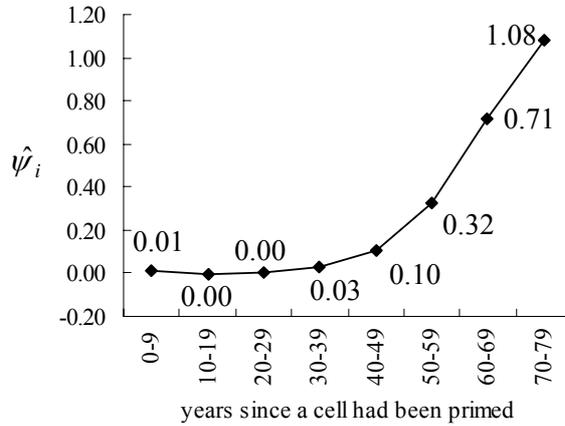


Figure 4.6.2-g: 成長危険強度の推定値 (男性, 日本).

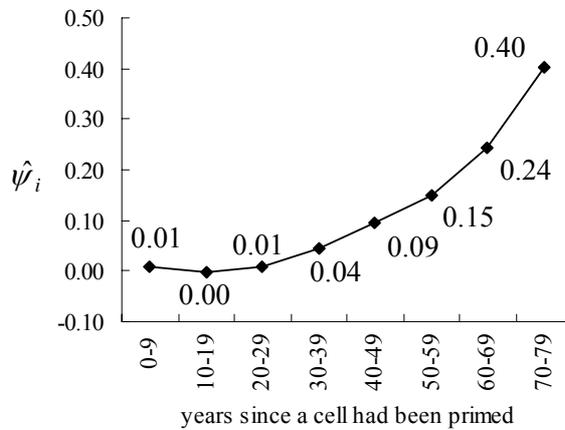


Figure 4.6.2-h: 成長危険強度の推定値 (女性, 日本).

4.7. 結び

4.7.1. 結論

この章では、年齢・時代区別に観測された癌死亡データの解析から環境中の危険因子の量を投影するためにSTSモデルを提案した。STSモデルは、分子生物学の分野で研究されている2段階癌成長モデルの考えを、年齢・時代区分データの解析に利用することによって導出された。Holford, Zhang, McKay (1994) もまた、多段階癌成長モデルを参考にして導出した新しいタイプのAPCモデルを年齢・時代区分データに適用している。しかし、年齢効果に与える制約を導き出すために癌成長モデルを参考に行っているのに対し、STSモデルではモデル全体の構築に癌成長モデルを参考としている点で、Holford等とは考え方が異なる。

また、STSモデルを日本国および米国の癌死亡データに当てはめ、両国のデータの解析結果を比較した。そして、環境ハザードの推定値から、米国の男性および女性そして日本国の男性に共通して1929年後の世界恐慌による食料事情と社会不安の影響を、また、日本国の男性および女性について「高度経済成長期」の環境汚染の影響を読み取ることができた。

4.7.2. リマーク (Remark)

Remark 4.1:

STSモデルおよびAPCモデルをデータへ当てはめた際、 G^2 統計量は水準0.05で有意となってしまう。これは、死亡者数が多数であるため、モデルからのわずかなズレですら検出されることが理由の1つである。これまで、多数の研究者がこの問題に直面している。(例えば、Holford, 1983, Robertson, Boyle, 1986, Tango, 1987.) この問題に対処する1つの方法は、特殊ポアソン変動 (extra-Poisson variation) を仮定する2種類のF-検定 (Wedderburn, 1974, Breslow, 1984 参照) を利用する方法である。しかし、Robertson, Boyle (1998) は、彼等の提案したAPCモデルをデータへ当てはめた場合のモデルからの残差に規則性があることを示し、これらの方法を使用することが妥当ではないと主張している。我々の場合も、モデルからの残差に規則性が見られた。このため、特殊ポアソンを用いた方法よりも、より複雑なモデルを考えることの方が妥当な対処方法であると判断した。(ここでは、残差の値を示すことは省略する。)

Remark 4.2:

STSモデルに含まれる環境危険強度 ξ_k は、環境中 (社会環境を含む) に存在する癌に関する危険因子の総量を表すものである。そこで、 ξ_k を推定し、その趨勢を検討することにより、癌に関して環境が悪化しているのか、それとも改善しているのかを知ることができる。また、環境中の複数の種類の汚染物質量を経年的に測定し、その趨勢を ξ_k の推定値と比較検討することにより、危険因子が何であるかを特定するために、STSモデルを利用することができる。

Remark 4.3:

STSモデルでは、年齢によって変化する発癌または癌抑制に関する身体的要因や、進展する医療技術による処置に関する要因は考慮していない。よってSTSモデルは、幅広い年齢と時代が観測される癌死亡データを解析するには単純過ぎると思われるかもしれない。そこで、例えば、体内から分泌される発癌物質（乳癌の要因と考えられるエストロゲンなど）によって細胞が癌化の第1段階へ進まされる危険強度が年齢に依存すると仮定して $\delta(u)$ と表し、免疫によって癌化した細胞が除外される強度が年齢に依存すると仮定し $\alpha^*(u)$ と表し、定期健診などによって癌腫瘍が発見され、医療処置によって腫瘍が切除される強度が時代に依存すると仮定し $\beta(v)$ と表すと、次のようなモデルを考えることができる。

$$\mu(u, v) = \int_{v-u}^v \{ \delta(t - v + u) + \xi(t) \} \psi(v - t) dt - \beta(v) - \alpha^*(u) \quad (4.7.2-a)$$

しかし、このようなモデルは大変複雑で、実際のデータ解析へ用いるには、さらなる研究が必要である。一方、STSモデルは、これまで同様のデータの解析で頻繁に用いられてきたAPCモデルより、データへ良く当てはまった。この意味で、STSモデルは、単純ではあるが、十分に有用性を持つと言える。

Remark 4.4:

分子生物学研究の分野では、Knudson (1971) が提唱した、「2ヒット説」が受け入れられている。これは、小児癌のひとつである網膜芽細胞腫の臨床遺伝学的研究から、2回の突然変異が遺伝子に起こることで癌が発症するという考えである。この意味で、STSモデルは「1ヒット説」を採用していることになる。「2ヒット説」を採用し、時点 t に1つの細胞が1回目のヒットを受ける危険強度を $\xi^{(1)}(t)$ 、その後の時点 $s > t$ にその細胞が2回目のヒットを受ける危険強度を $\xi^{(2)}(s)$ とすると、

$$\mu(u, v) = \int_{v-u}^v \xi^{(1)}(t) \left\{ \int_t^v \xi^{(2)}(s) ds \right\} \psi(v - t) dt. \quad (4.7.2-b)$$

と書くことができるであろう。さらに、「遺伝性の癌の場合は、生まれたときに生殖細胞がすでに突然変異を受けており、あと1回、眼の細胞に突然変異が加わったときに癌が発症する」という考えを採用すれば、出生時点以前の環境危険強度も累積する必要があり、モデルはより複雑となる。このような複雑なモデルを、疫学調査や人口動態調査データの解析に用いることは、相当な困難を伴うと考えられる。繰り返しになるが、この意味で、STSモデルは、単純ではあるが、十分な有用性を持つと言える。

5. 超高齢者モデル

この章では、年齢・時代区別に観測された超高齢者の生存者数統計および死亡者数統計の解析を行う。まず、極値理論で用いられる一般パレート分布に関する一般的性質を示し、超高齢者データへ一般パレート分布を仮定することの妥当性を議論する。そして、連続変量に対して議論されている極値理論を年齢・時代区分データへ適用するためのモデルを提案する。さらに、実際の生存者数統計および死亡者数統計のデータ解析から、寿命分布の限界を議論する。

5.1. 背景と概要

5.1.1. 目的

近年の老化に関する研究分野では、2つの説が主流となっている。一つは分子障害説（Harman, 1956）と呼ばれ、人の細胞は時間とともにフリーラジカルと呼ばれる破壊分子により酸化され、次第に老化していくというものである。2つ目はプログラム説（Hayflick, 1961）と呼ばれ、人の細胞には生物時計（バイオリジカル・クロック）がプログラムされており、細胞が再生できる寿命が設定されているというものである。前者は、「抗酸化物質」を摂取し、活性酸素の働きを抑えることによって寿命を延ばせるという考えの根拠となっている。一方後者は、どんなに健康を保っても、寿命には限界があるという考えの根拠となっている。本研究の目的は、このような老化に関する研究事情を鑑みながら、年齢・時代区別に観測された超高齢者の生存者数統計および死亡者数統計を解析し、寿命分布の限界を議論するものである。

5.1.2. 背景

超高齢者（the oldest-old）の寿命については、種々の側面から興味もたれている。生物（動物）学では加齢、老化現象を生理学的に、あるいは遺伝子の機能まで視野に入れて研究が行われている。このような研究ではヒトを対象とする必要はない。たとえばCarey *et al.* (1992) は、生活環境を変えることにより地中海八工の寿命が延びた実験結果を示し、「生物の寿命が有限であることを立証しようとしたすべての研究は失敗している」と断言している。人間社会では、医療、医療制度への影響が直接的である。さらに高齢者の介護、健康、生活が問題であり、関連した労働、社会、ビジネスが考えられている。これら社会経済問題の場合には、高齢者の数、年齢構成、その将来予測が主要な関心であるが、ここでは人口学的研究には立ち入らず、むしろ好奇心だけから寿命分布の上限を調べる。国勢調査などの作表では超高齢者を年齢区分の最高区分としているが、その限度を何歳にするかは関連した問題である。記録的な自然災害への対策、製品の寿命または信頼性の保証のために極値理論が利用されている。確率標本の、ある閾値以上の条件付分布は、適当に標準化し閾値を大きくした極限で一般パレート分布に従うと

いう、理論的結果がある．人口を世代（いわゆる出生時コホート）ごとに区分すれば，高齢まで生存した条件のもとで寿命が一般パレート分布に従うと仮定できるのではないか，というのが動機である．

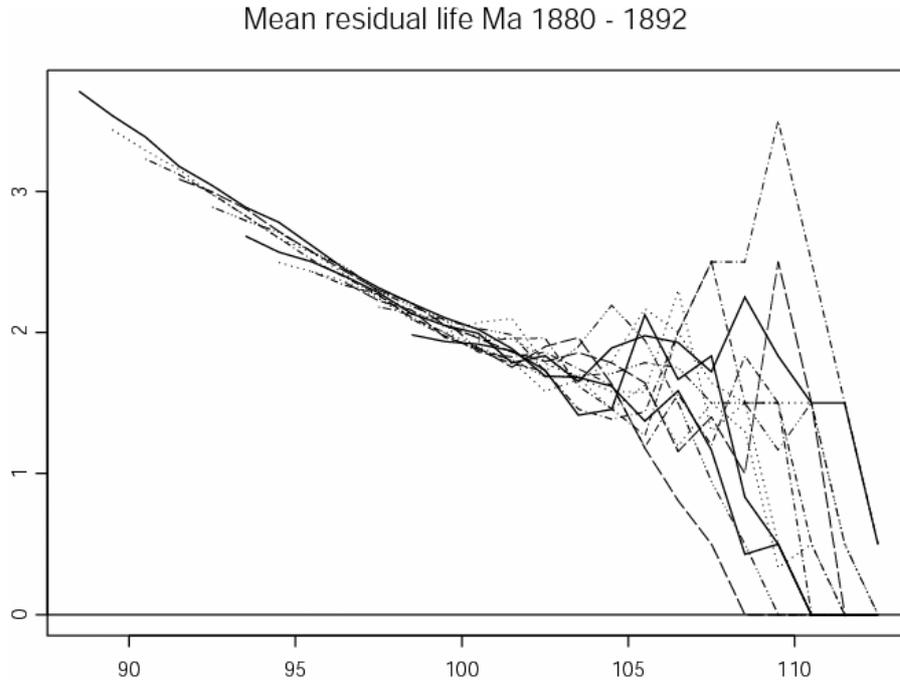


Figure 5.1.2-a : 死亡者数統計から算出した世代別平均余命の推移 .

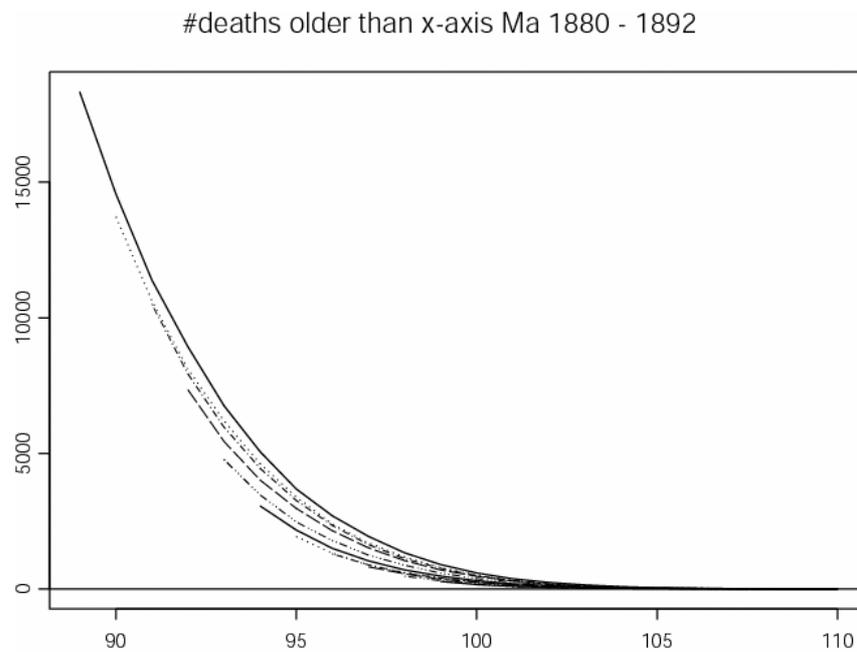


Figure 5.1.2-b : 世代別死亡者数の推移 .

死亡者数統計の粗い近似により，世代ごとの標本平均余命を計算するとFigure 5.1.2-a のようにほぼ直線的に減少する．各線が1世代の標本平均余命であり，左端の点の横軸が1980年（もっとも古いデータ）における，その世代の年齢である．すべての世代について言えるのではないし，100歳を越えると不安定になるが，直線性が特徴付ける一般パレート分布を検討する根拠となる．Figure 5.1.2-b は同じ死亡者数統計で，各世代の1980年以降の生存数である．この図のように超高齢者数は着実に増えているが，Figure 5.1.2-a を見るとその平均余命はほとんど変わっていない．（これらの図については第3節の始めで説明する．）極値理論を用いた解析にはKaufmann（2001）などがあるが，結論が明確でなかった．人口学で利用しているモデルでは（たとえばKannisto，1999）寿命分布に上限がない．

5.2. 一般パレート分布

5.2.1. 定義と性質

一般極値分布の吸引領域に属する確率分布の，ある閾値以上の条件付分布は，適当に標準化し，閾値を大きくした極限で一般パレート分布

$$F(x; \gamma, a) = \begin{cases} 1 - (1 + \gamma x/a)^{-1/\gamma} & \gamma \neq 0 \\ 1 - \exp(-x/a) & \gamma = 0 \end{cases} \quad (5.2.1-a)$$

に従う．その分布範囲は

$$\begin{cases} 0 < x < \infty & \gamma \geq 0 \\ 0 < x < a/(-\gamma) & \gamma < 0 \end{cases} \quad (5.2.1-b)$$

である．(5.2.1-a) の分布を $\text{GPrt}(\gamma, a)$ で表わす． $\gamma = 0$ の場合は指数分布である．一般パレート分布については，たとえばColes（2001）参照．

確率変数 $X \sim \text{GPrt}(\gamma, a)$ の期待値は $E(X) = a/(1 - \gamma)$ ， $\gamma < 1$ ，である． $\gamma \geq 1$ では期待値は存在しない．ハザード関数は

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{1}{a + \gamma x} \quad \forall \gamma \in R$$

である． $X \geq x_0$ の条件の下で， $\text{GPrt}(\gamma, a + \gamma x_0)$ の生存関数は，

$$P\{X - x_0 \geq y \mid X \geq x_0\} = \frac{1 - F(x_0 + y)}{1 - F(x_0)} = \left(1 + \frac{\gamma y}{a + \gamma x_0}\right)^{-1/\gamma}, \quad y \geq 0; \gamma \neq 0,$$

つまり $X - x_0 \mid X \geq x_0 \sim \text{GPrt}(\gamma, a + \gamma x_0)$ であり，左裾を打ち切ったとき同じ型の分布に従う． $\gamma = 0$ の場合も同様である．

閾値を変えたときに分布の型が変わらないことは，一般極値分布の標本最大値の分布型が変

わらない最大値安定性に対応している．この性質から平均余命（mean residual life） $m(x_0)$ および条件付生存関数の逆関数 $y(p, x_0)$ （ p は生存確率）が次のように閾値 x_0 の1次関数となる．

$$m(x_0) := E(X - x_0 | X \geq x_0) = \frac{a + \gamma x_0}{1 - \gamma}, \quad \gamma < 1$$

$$y(p, x_0) := \left\{ \frac{1 - F(x_0 + y)}{1 - F(x_0)} \right\}^{\leftarrow} = (p^{-\gamma} - 1)(a + \gamma x_0) / \gamma, \quad \forall \gamma \neq 0,$$

ただし， $\{ \}^{\leftarrow}$ は， $\{ \}$ 内の関数の逆関数を表す．逆に条件付期待値 $m(x_0)$ が閾値 x_0 の1次関数であれば，積分方程式を解いて，分布関数（5.2.1-a）を導ける．

5.2.2. γ が負か零か

データ $\{x_j; j=1, 2, \dots, n\}$ に一般パレート分布（5.2.1-a）を当てはめたとき， $\gamma < 0$ か $\gamma = 0$ が興味の焦点の一つである．それを見る簡便な方法は次の標本平均余命 \hat{m} のグラフである．これは閾値を決定するときの判断にも利用される．

$$\hat{m} := \frac{\sum_{j=1}^n (x_j - x_0) I[x_j \geq x_0]}{\sum_{j=1}^n I[x_j \geq x_0]}, \quad (I[\text{true}] = 1, I[\text{false}] = 0).$$

$\gamma < 0$ の場合には， $X \sim \text{GPrt}(\gamma, a)$ の分布関数は

$$F(x; \eta, \omega) = 1 - (1 - x/\omega)^\eta, \quad 0 \leq x \leq \omega; \quad \eta = -1/\gamma, \quad \omega = a/(-\gamma) \quad (5.2.2-a)$$

となる．つまり X/ω がベータ分布 $\text{Be}(1, \eta)$ に従う．期待値は $E(X) = \omega/(1+\eta)$ ，条件付期待値（平均余命）は

$$\begin{aligned} m(x_0) &= E(X - x_0 | X \geq x_0) = \frac{\gamma}{1 - \gamma} \left(x_0 + \frac{a}{\gamma} \right) \\ &= \frac{-\gamma}{1 - \gamma} (\omega - x_0) = \frac{\omega - x_0}{1 + \eta} \end{aligned}$$

である．負の勾配

$$\alpha := \frac{-\gamma}{1 - \gamma} = \frac{1}{1 + \eta} = E \left(\frac{X - x_0}{\omega - x_0} \middle| X \geq x_0 \right)$$

は生続率（perseverance parameter）と呼ばれている．ハザード関数は

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{\eta}{\omega} \left/ \left(1 - \frac{x}{\omega} \right) \right.$$

である．

寿命分布に上限 $\omega < \infty$ があれば、極限分布で $\gamma < 0$ となるとは限らない。有界の上限でハザード関数 $h(x)$ が必ず無限大となるが、その速度が遅いときだけ $\gamma < 0$ となる。人口学ではハザード関数 (“force of mortality” と呼ばれている) が指数的に増加する、無限区間上に分布するゴンペルツ (Gompertz) 曲線が用いられている。これは負のゲンベル分布で、最小値極限分布のひとつで $\gamma = 0$ である。高齢者の寿命分布に限ると極値理論の観点からゴンペルツ曲線は不自然である。後述の Aarssen, de Haan (1994) も、この点を議論している。

5.2.3. 最尤推定

一般パレート分布 (5.2.1-a) からの大きさ n の確率標本の対数尤度は

$$l = -\frac{1+\gamma}{\gamma} \sum_{j=1}^n \log(1+\tau x_j) - n \log a, \quad \tau = \frac{\gamma}{a} \quad (5.2.3-a)$$

である。これから最尤方程式は

$$\begin{cases} \frac{1}{n} \sum_{j=1}^n \log(1+\tau x_j) = \gamma \\ \frac{1+\gamma}{\gamma} \frac{1}{n} \sum_{j=1}^n \frac{x_j}{1+\tau x_j} = \frac{1}{\tau} \end{cases}$$

となる。第1式を第2式に代入すれば、 τ だけが未知の方程式となる。 $\gamma < 0$ のとき分布範囲がパラメータに依存して、最尤推定量の正則条件が満たされないが、このような場合を Smith (1985) が議論している。

対数尤度 (5.2.3-a) からフィッシャー (Fisher) 情報行列を導出するために次のモーメントを与える。

$$E[(1+\tau X)^r] = \frac{\tau}{\gamma} \int_0^{\infty} (1+\tau x)^{r-1-1/\gamma} I[1+\tau x \geq 0] dx = 1/(1-\gamma r)$$

ただし、 $r < 1/\gamma$ である。また、特に

$$\begin{aligned} E[(1+\tau X)^{-1}] &= 1 - E[\log(1+\tau X)]/(\gamma+1) = 1/(\gamma+1) \\ E[\{X/(1+\tau X)\}^2] &= E[\{1-1/(1+\tau X)\}^2/\tau^2] = 2\gamma^2/\{\tau^2(2\gamma+1)(\gamma+1)\}. \end{aligned}$$

さらに、 $n=1$ のとき

$$\begin{aligned} \frac{\partial l}{\partial \gamma} &= \frac{1}{\gamma^2} \log\left(1+\frac{\gamma}{a}x\right) - \frac{1+\gamma}{a\gamma} \frac{x}{1+\gamma x/a}, \quad \frac{\partial l}{\partial a} = \frac{1+\gamma}{a^2} \frac{x}{1+\gamma x/a} - \frac{1}{a}, \\ \frac{\partial^2 l}{\partial \gamma^2} &= -\frac{1}{\gamma^3} \log\left(1+\frac{\gamma}{a}x\right) + \frac{2}{a\gamma^2} \frac{x}{1+\gamma x/a} + \frac{1+\gamma}{a^2\gamma} \left(\frac{x}{1+\gamma x/a}\right)^2, \end{aligned}$$

$$\frac{\partial^2 l}{\partial \gamma \partial a} = \frac{1}{a^2} \frac{x}{1+\gamma x/a} - \frac{1+\gamma}{a^3} \left(\frac{x}{1+\gamma x/a} \right)^2,$$

$$\frac{\partial^2 l}{\partial a^2} = -\frac{2(1+\gamma)}{a^3} \frac{x}{1+\gamma x/a} + \frac{\gamma(1+\gamma)}{a^4} \left(\frac{x}{1+\gamma x/a} \right)^2 + \frac{1}{a^2}$$

である．これらの x を X に変え，上記のモーメントを用いて期待値を計算するとフィッシャー情報行列

$$I(\gamma, a) = \frac{n}{(2\gamma+1)(\gamma+1)} \begin{pmatrix} 2 & 1/a \\ 1/a & (\gamma+1)/a^2 \end{pmatrix}, \gamma > -\frac{1}{2} \text{ または } \eta > 2,$$

を得る．フィッシャー情報行列は $\gamma > -1/2$ ならば有限で，その逆行列，つまり最尤推定量の漸近分散共分散行列は

$$\text{Var}((\hat{\gamma}, \hat{a})) = \frac{\gamma+1}{n} \begin{pmatrix} \gamma+1 & -a \\ -a & 2a^2 \end{pmatrix}, \gamma > -\frac{1}{2} \text{ または } \eta > 2,$$

となり，正則条件と同じ議論が成り立つ．これから $-1/2 < \gamma < 0$ のときの分布上限 $\omega = a/(-\gamma)$ の（漸近）推定誤差は

$$\text{Var}(\hat{\omega}) = \frac{a^2(\gamma+1)^2(2\gamma+1)}{n\gamma^4} \leq \frac{a^2}{n\gamma^4}, \frac{1}{\omega} SD(\hat{\omega}) \sim \frac{1}{\sqrt{n}|\gamma|} = \frac{\eta}{\sqrt{n}}, \gamma \rightarrow 0,$$

となる． $\gamma = -1/2$ の場合も最尤推定量の漸近有効性，漸近正規性が成り立つが，収束速度が速くなり $(n \log n)^{-1/2}$ の位となる． $-1 < \gamma \leq -1/2$ の場合，最尤推定量が存在するが漸近正規性は成り立たない．漸近有効性も疑わしいが未解決である． $\gamma \leq -1$ の場合は一般に最尤推定量は存在しない．なお，Smith (1987) は，一般パレート分布の吸引領域に属する分布からの確率標本について，最尤推定量の精度を議論している．閾値を決定するには，標本平均余命関数のグラフ，および閾値を変えた時の推定値・推定誤差のグラフを描いて判断する．パラメータ γ の値に依存する一般パレート分布の特徴を Table 5.2.3-a にまとめておく．

Table 5.2.3-a : パラメータの値に依存する一般パレート分布の特徴．

γ の範囲	$(-\infty, -1]$	$(-1, -1/2]$	$(-1/2, 0)$	$\{0\}$	$(0, 1)$	$[1, \infty)$
分布の別名	beta分布			指数分布	Pareto分布	
分布の範囲	有限分布範囲			上限が存在しない		
期待値	有限期待値					期待値が存在しない
最尤推定量	存在しない	無効? (未解決)	有効			

5.3. 年齢・時代区分データとそのモデル

5.3.1. 年齢・時代区分データ

一般に年齢時代区分によって与えられた統計表については，項目（セル）の番号を (i, j) ， $i=1, 2, \dots, I$ ； $j=1, 2, \dots, J$ で表すが，この節では添字 i により年齢を， j により年度を表わす．つまり，例えば1970-2000年の100-120歳について与えられた統計表では， (i, j) ， $i=1970, \dots, 2000$ ； $j=100, \dots, 120$ によってセルを表す．利用可能なデータの詳細についても後で述べる．性別，出生年度別のコホートを一様な集団とみなすことを，再び強調しておく．

生存者数統計

j 年度の一定の調査日に i 歳で生存する人々を考え，その人数を M_{ij} により表わすことにする．（時間，年齢）平面の上で傾斜45度の線分で1人の生涯を表わすレキシス・ダイアグラムを用いると， M_{ij} は調査時点を表わす縦線と，斜線の交点の数である．（Figure 2.2.2-a 参照．）このグループの人が1年以内に死亡する事象を V_{ij} とする．その人数は

$$M_{ij} - M_{i+1, j+1} =: D_{ij} \quad (5.3.1-a)$$

である． $k = j - i$ で出生コホート年度を表わす． M_{ij} に数えられる人々の年齢は $[i, i+1)$ に分布している，つまり， j 年度の調査時点までの生存時間が同一ではない．このため，厳密には「出生年度別コホートを一様な集団とみなす」という仮定だけの下では D_{ij} に数えられる人々の死亡が同一分布に従うと考えられない．しかしここでは， D_{ij} を， $k - j - i$ 年度に生まれ満 i 歳以上 $i+1$ 歳未満で死亡した人数の近似値と考える．高齢生存者数を毎年記録しているので $M_{ij} = 0$ を満たす j を確認できれば，コホート全体の死亡データが確定する．

死亡者数統計

人口動態統計では，第 j 年度に i 歳以上 $i+1$ 歳未満で死亡した，という事象 \bar{W}_{ij} の人数 N_{ij} が記録されている．レキシス・ダイアグラムで見ると， N_{ij} は個人の死亡を表わす斜線の端点が正方形の領域に落ちる数である．（Figure 2.2.1-a 参照．）このときには出生コホートが不明になる．もっとも簡単な解決法は N_{ij} を2等分して，あい続くコホートに振り分けることである．生存者数統計の，下記の確率 $q(k, i)$ の事象の数を $(N_{ij} + N_{i, j+1})/2$ により近似する．副節5.1.2で述べた Figure 5.1.2-a，Figure 5.1.2-b はこのような近似計算で作成した．他の方法については，推測と関連して述べる．コホート全体が死亡したかどうかは，データから判断するが，生存者数統計が判断に役立つ．

5.3.2. 生存者数統計のモデル

j 年度に i 歳であるコホート（この集団を $k = j - i$ で区別する）に関するデータ M_{ij} にたいして、 $P(V_{ij})$ の近似である確率

$$q(k, i) := F_k(i+1) - F_k(i) = \bar{F}_k(i) - \bar{F}_k(i+1), \quad k = j - i \quad (5.3.2-a)$$

の多項分布を当てはめる． $F_k(t)$ は $\text{GPrt}(\gamma_k, a_k)$ の左打切り分布， $\bar{F}_k(t)$ はその生存関数である．簡単のために添え字 k を省略し， $\gamma < 0$ の場合に限って，(5.2.2-a) のパラメータを用いると

$$q(i) = \bar{F}(i) - \bar{F}(i+1), \quad \bar{F}(t) = \bar{F}(t; \omega, \eta) = \begin{cases} 1 & t < 0, \\ (1 - t/\omega)^\eta & 0 \leq t < \omega, \\ 0 & \omega \leq t, \end{cases}$$

対数尤度は

$$l = \sum_{i=1}^k m_i \log q(i), \quad m_i = M_{ij} - M_{i+1, j+1},$$

尤度方程式は

$$\frac{\partial l}{\partial \omega} = \sum_{i=1}^k \frac{m_i}{q(i)} \frac{\partial}{\partial \omega} (\bar{F}(i) - \bar{F}(i+1)) = 0, \quad \frac{\partial l}{\partial \eta} = \sum_{i=1}^k \frac{m_i}{q(i)} \frac{\partial}{\partial \eta} (\bar{F}(i) - \bar{F}(i+1)) = 0$$

$$\frac{\partial}{\partial \omega} \bar{F}(t) = \frac{\eta t}{\omega^2} \left(1 - \frac{t}{\omega}\right)^{\eta-1} = 0, \quad \frac{\partial}{\partial \eta} \bar{F}(t) = \bar{F}(t) \log \left(1 - \frac{t}{\omega}\right) = 0, \quad 0 < t < \omega,$$

となる．最尤法が有効であるためには， ω が 1 に比べて十分大きいことが必要である．

5.3.3. 死亡者数統計のモデル

k 年度に誕生した人が， j 年度に死亡したとし，その誕生日(時点)，死亡日(時点)を，それぞれ $k + v$ ， $j + w$ ，ただし v ， w は，独立な $(0, 1)$ 一様確率変数と仮定する．連続化した死亡年齢 y は

$$y = j + w - (k + v) = (j - k) + (w - v) = j - k + u \quad (5.3.3-a)$$

となる． $u = w - v$ は $(-1, 1)$ 上の対称三角分布である．したがってこの人の死亡年齢は $j - k$ または $j - k - 1$ である．つまり j 年度に i 歳で死んだという事象 V_{ij} には， $j - i$ コホートの人と $j - i + 1$ コホートの人が混じっている．このため，コホート毎に *i.i.d.* を仮定するだけでは，一般パレート分布のパラメータを推定することはできない．以下では，以上の問題に対処する方法として，ポアソン点過程モデルを考える方法，擬似確率標本により推定する方法および表から直接推定する方法を提案する．

ポアソン点過程モデル

死亡点から出生点が自動的に定まる．死亡点をポアソン過程とみなし，その強度の変化が緩やかであると想定する．レキシス・ダイアグラムの上で死亡強度が断片（piece）ごとに一定であることを仮定したPCIモデルを適用する．ただし断片の一点での強度が一般パレート分布のハザード関数から定まるとする．これも近似法であるが，競合リスクを分離して推定できる点が重要である．年齢・時代区分データによる寿命分布解析におけるPCIモデルを構成するため，1年度1歳単位の正方形領域9個からなる断片達 \mathcal{V}_{rs} を次のように定義する．まず2重添字の集合 $\mathcal{N}^2 = \{(i, j); i, j \in \mathcal{N}\}$ の中の部分集合を

$$\mathcal{M}_{rs} = \bigcup_{\epsilon_1, \epsilon_2 \in \{0, 1, 2\}} (r + \epsilon_1, s + \epsilon_1 + \epsilon_2); r, s = 0, 3, 6, \dots$$

で定義する．次にレキシス・ダイアグラム上の上記断片を

$$\mathcal{V}_{rs} = \bigcup_{(i, j) \in \mathcal{M}_{rs}} \tilde{W}_{ij}; r, s = 0, 3, 6, \dots$$

により定義する．この断片は $k+1$ ($k = s - r = 0, 3, 6, \dots$) 世代を完全に含み， k ， $k+2$ 世代の半分を含む．断片 \mathcal{V}_{rs} 上の強度が一様であるとし， λ_{rs} で表わす．一方コホートごとの寿命分布が一定で，閾値以上の条件付分布が一般パレート分布に従うならば，その k 世代のハザード関数（強度）は前述の通り， $h(x; k) = 1/(a_k + \gamma_k x)$ である．その \mathcal{V}_{rs} の中心での値を

$$\lambda_{rs} = h(r + 1.5, s - r + 1); r, s = 0, 3, 6, \dots \quad (5.3.3-b)$$

とする． \mathcal{V}_{rs} で僅かでも生存した全ての人々の生存時間の累和 x_{rs+} を，人口年（person-years） M_{ij} の累和で近似すると，点過程の尤度は以下のように表される．

$$\begin{aligned} L &\propto \prod_{r, s} \lambda_{rs}^{d_{rs}} \exp(-x_{rs+} \lambda_{rs}) \\ &\approx \prod_{r, s} \lambda_{rs}^{d_{rs}} \exp(-\lambda_{rs} \sum_{(i, j) \in \mathcal{M}_{rs}} M_{ij}), \quad d_{rs} = \sum_{(i, j) \in \mathcal{M}_{rs}} N_{ij} \end{aligned}$$

これをパラメータに関して最大化する．

擬似確率標本による推定

一般パレート分布からの確率標本に基づく推定法（たとえば最尤法）を適用する簡便法として，区分計数データから復元シミュレーションにより，擬似確率標本を生成する．(5.3.3-a)式より，事象 \tilde{W}_{ij} の出現数である N_{ij} 個の三角分布乱数 u を独立に生成し，寿命 $y = j - i + u$ を $u > 0$ ならば $k = j - i$ コホート， $u \leq 0$ ならば $k = j - i - 1$ コホートに属するデータとする．死亡日が既知であっても季節変動があるならば，それを調整するために擬似化したデータが必要である．

表からの直接推定

事象の確率をより詳しく書き，その組合せで推定方程式を構成できる．ただしその解法は容易ではなく検討中である．式(5.3.3-a)に戻って，出生年度 k ，死亡年齢(日計算) y の人(死亡年齢(年計算)は，整数部分 $[y]=i$)の死亡年度がいつになるかを考えると死亡日が $k+v+y$ であるから，死亡年度が j となる条件は

$$j \leq k+v+y < j+1 \Leftrightarrow i \leq y+v < i+1 .$$

これと $[y]=i$ を合わせると， y についての条件

$$\begin{aligned} & \text{死亡年度} = j \\ & \Downarrow \\ & i \leq y < i+1-v, \rho_0(k, i) := \int_0^1 F_k(i+1-v)dv - F_k(i), \end{aligned} \quad (5.3.3-b)$$

が得られる． $F_k(t)$ は(5.3.2-a)の下で定義した． $\rho_0(k, i)$ は，条件式の確率である．同様に

$$\begin{aligned} & \text{死亡年度} = j+1 \\ & \Downarrow \\ & i+1-v \leq y < i+1, \rho_1(k, i) := F_k(i+1) - \int_0^1 F_k(i+1-v)dv, \end{aligned} \quad (5.3.3-c)$$

となる． $\rho_1(k, i)$ は，条件式の確率である．出生年度 $k-1$ の人が i 歳で j 年度に死亡する確率は，この議論で k を $k-1$ に変えるだけでよく $\rho_1(k-1, i)$ である．(注意：この記号を用いると $P(V_{ij})$ は(5.3.2-a)の $q(k, i)$ ではなく

$$P(V_{ij}) = R_k \rho_1(k, i) + R_{k-1} \rho_0(k-1, i+1)$$

である．) R_{k-1} および R_k は，それぞれ $k-1$ 世代， k 世代の閾値における生存比である．

式(5.3.3-b)と(5.3.3-c)の積分は陽に表わせる：

指数分布($\gamma=0$)の場合，

$$\int_0^1 F(t-v)dv = 1 - \int_0^1 \exp\{-(t-v)/a\}dv = 1 - a \{ \exp(1/a) - 1 \} \exp(-t/a) ,$$

ベータ分布($\gamma < 0$)の場合，

$$\begin{aligned} \int_0^1 F(t-v)dv &= 1 - \frac{\omega}{1+\eta} \left\{ \left(\frac{\omega+1-t}{\omega} \right)^{1+\eta} - \left(\frac{\omega-t}{\omega} \right)^{1+\eta} \right\} \\ &= F(t) + O(\omega^{-2}), \quad \omega \rightarrow \infty. \end{aligned} \quad (5.3.3-d)$$

尤度の計算には，閾値における各世代の生存比が必要であるが，観測比率をそのまま既知パラメータとする．

簡単のため $k-1$ ， k の2世代を含む3時代区分データ($N_{ij}, N_{i,j+1}, N_{i,j+2}$)， $k=j-i$ ，

$i = 0, 1, \dots$ から3組のパラメータ (γ_{k-1}, a_{k-1}) , (γ_k, a_k) , (γ_{k+1}, a_{k+1}) を推定する．この2世代の i 歳での死亡率を4分割し，対応する観測数を用いた対数尤度

$$\sum_i \left[N_{i,k+i} \log(R_k \rho_0(k, i)) + N_{i,k+i+1} \{ \log(R_k \rho_1(k+1, i)) + \log(R_{k+1} \rho_0(k+1, i)) \} \right. \\ \left. + N_{i,k+i+2} \log(R_{k+2} \rho_1(k+2, i)) \right],$$

から推定する．ただし R_k , R_{k+1} , R_{k+2} は k , $k+1$, $k+2$ 世代の閾値における生存比であり，

$$\rho_0(k, i) = \frac{1}{1 + \eta_k} \left\{ (i+1 + \eta_k - \omega_k) \left(1 - \frac{i}{\omega_k} \right)^{\eta_k} - (i+1 - \omega_k) \left(1 - \frac{i+1}{\omega_k} \right)^{\eta_k} \right\}$$

$$\rho_1(k, i) = \frac{\omega_k - i}{1 + \eta_k} \left(1 - \frac{i}{\omega_k} \right)^{\eta_k} - \frac{\omega_k - i + \eta_k}{1 + \eta_k} \left(1 - \frac{i+1}{\omega_k} \right)^{\eta_k},$$

である．世代数を増やすと，同時に解かねばならないパラメータ数が増え数値解法が困難になる．これはコホートモデルで絶えず生じる困難である．

5.4. 実データの解析

ギネス認定の，最近の世界最高齢者は泉重千代さん（1865-1986，120歳），ジャンヌ・カルマンさん（1875-1997，122歳），本郷かまとさん（1887-，115歳女性最高），中願寺雄吉さん（1889-，114歳男性最高）などで，男性と女性両方が名を連ねている．しかし，本章では，データの理解と解析法の検討に重点を置き，男女の比較には至っていない．そこで，以下では男性データだけを解析する．

5.4.1. 生存者数統計の解析

厚生労働省(厚生省)は1964年以来毎年9月に，100歳以上の全国高齢者名簿，いわゆる長寿番付，を発表している．その付表で100歳以上1歳刻みで男女別高齢者数を数えている．Table 5.4.1-aは1964-2002年の男性生存者数である．ただし，これは名簿（氏名，住所，誕生日）で，厚生労働省の要請にたいする市町村の自由な回答の収集である．各市町村がどのように調べているか，統一されておらず不明である．東京都の市では，住民基本台帳から自動的に選出して報告しているようであり，死亡者数統計と比較すると，最近ではこのような処理が支配的であると予想される．世代ごとに分けると，約50人以上のデータが得られる世代の誕生年度は1873-1902年である．全死亡者を死亡者数統計と比較できる世代は1880-1890年に限られる．

Table 5.4.1-a : 100歳以上生存者数 (男性).

年齢	年度																			
	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982
100歳	2	19	24	22	26	37	26	22	31	41	56	44	43	57	56	68	101	70	98	108
101歳	12	1	7	15	10	14	25	15	11	15	27	28	23	22	31	31	45	52	47	60
102歳	5	5	1	4	12	6	8	15	10	6	8	16	18	12	13	16	21	26	32	27
103歳	-	5	1	1	2	7	6	4	10	5	4	1	11	11	6	5	4	12	16	23
104歳	-	-	2	1	1	2	3	4	4	6	2	2	1	8	8	4	2	1	4	9
105歳	-	-	-	2	-	-	2	1	2	1	5	2	2	1	5	2	3	-	1	2
106歳	-	-	-	-	-	-	-	1	1	3	-	2	1	1	1	4	-	2	-	1
107歳	-	-	-	-	-	-	-	-	1	1	-	-	2	-	1	-	3	-	2	-
108歳	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	1	-	1	-	2
109歳	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	1
110歳	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
111歳	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
112歳	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
113歳	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
114歳	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
115歳	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
116歳	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-
117歳	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
118歳	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
119歳	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
120歳	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
合計	20	31	36	46	52	67	70	62	70	78	103	96	102	113	122	132	180	165	202	233

年齢	年度																			
	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
100歳	132	178	157	168	225	298	302	314	373	388	430	509	640	619	722	837	885	945	1173	1265
101歳	68	81	102	97	108	134	171	194	183	223	243	282	302	412	385	457	495	566	623	747
102歳	33	48	53	50	65	63	80	93	100	101	143	148	158	188	265	257	292	305	361	401
103歳	18	20	27	26	32	31	40	46	47	60	62	84	78	89	96	153	157	169	181	229
104歳	14	10	8	12	20	18	13	19	27	26	35	30	47	47	52	54	86	93	108	106
105歳	2	8	6	4	6	11	11	6	14	13	15	20	13	26	25	25	30	49	53	65
106歳	1	-	4	3	2	5	7	3	4	8	8	12	8	8	18	17	12	14	25	34
107歳	-	1	-	1	3	1	4	4	-	3	6	4	4	6	4	7	9	7	9	16
108歳	-	-	1	-	1	-	1	1	1	-	1	3	2	2	2	2	4	6	2	5
109歳	-	-	-	-	-	1	-	-	-	-	1	2	2	2	-	2	1	3	3	1
110歳	-	-	-	-	-	-	1	-	-	-	-	-	1	-	1	-	2	-	3	2
111歳	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	1	-	2
112歳	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
113歳	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
114歳	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
115歳	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
116歳	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
117歳	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
118歳	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
119歳	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
120歳	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
合計	269	347	359	361	462	562	630	680	749	822	943	1093	1255	1400	1570	1812	1973	2158	2541	2875

死亡者数統計は歴年単位の集計である．生存者数統計は7-8月頃の調査である．そのために (5.3.2-a) 式の D_{ij} ; $j = k + i$ は N_{ij} , $N_{i+1,j}$, $N_{i,j+1}$, $N_{i+1,j+1}$ のそれぞれの一部である．これらの間の違いを見るために , Figure 5.4.1-a では1980年誕生世代男性について D_{ij} ; $j = k + i$, $i = 100, 101, \dots$ を横軸に , N_{ij} , $N_{i+1,j}$, $N_{i,j+1}$, $N_{i+1,j+1}$ を縦軸にプロットした．それぞれの点を + , - , * , % で表わす．調査時点が年後半であるため , D_{ij} が N_{ij} より $N_{i+1,j+1}$ に近いのは当然かもしれない．少なくともこの時期では地方自治体の報告が正確ではなかったとも判断できる．

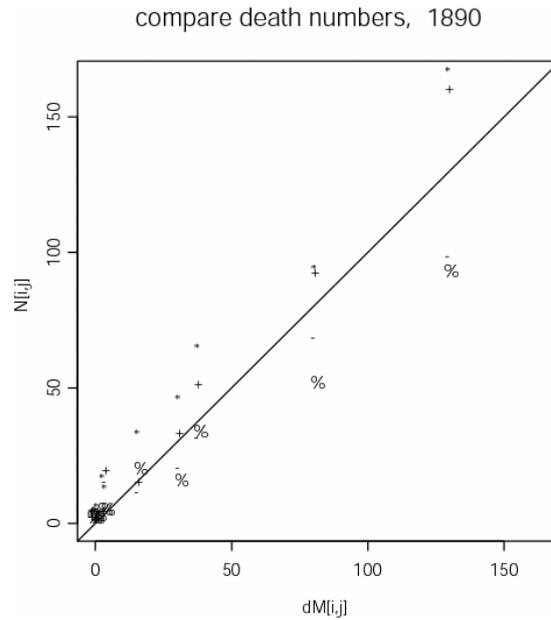


Figure 5.4.1-a : 生存者数統計から得た死亡者数と死亡者数統計の比較 .

Table 5.4.1-b に , 100歳以上の各人の寿命が同一のパレート分布に従うと仮定し , 生存者数統計にモデル (5.3.2-a) を当てはめた結果を示す . ただし , 110歳以上の死亡者数 , 死亡確率を統合し推定した . 解析に用いた1885-87年の世代は , 死亡者数統計にPCIモデルを当てはめた場合と対応している . この世代について , の推定値およびその標準偏差 (フィッシャー情報行列から算出) から $\gamma = 0$ の仮説は棄却される . また , $-a/\gamma$ の推定値から寿命分布の限界は134歳と推定された . 先に述べた実際の長寿記録よりかなり大きな値であるが , 利用できる最も新しい世代である1890-92年の世代を用いた推定もほとんど同様の結果となる . 一方 , 最も古い13つの世代である1864-66年の世代を用いた推定では , γ の推定値が正となり , 標準偏差から $\gamma = 0$ の仮説は棄却されている . Figure 5.4.1-b の標本平均余命でも , 減少傾向があるが水平に近く , これらの解析結果と一致している .

Table 5.4.1-b : 一般パレート分布に基づく多項モデルによるパラメータの推定結果 .

	1885-87年世代	1890-92年世代	1864-66年世代
γ の推定値	-0.06	-0.06	0.23
標準偏差	0.02	0.01	0.09
a の推定値	2.04	2.01	1.49
標準偏差	0.02	0.01	0.08
$-a/\gamma$ の推定値	34.00	33.50	-6.48
$-a/\gamma$ を含む年齢区分	134歳	133歳	-

Mean residual life, Male, 1880 - 1890

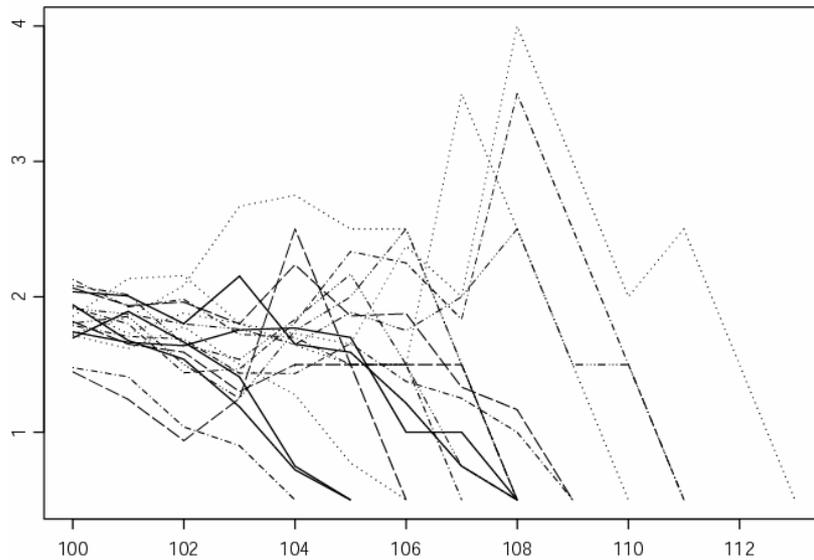


Figure 5.4.1-b : 生存者数統計から算出した世代別平均余命の推移 .

5.4.2. 死亡者数統計の解析

厚生労働省（厚生省）が毎年発表している人口動態統計では，85歳以上（年度によって90歳以上）の原因別死亡者数が1つの年齢区分にまとめられている．しかし，同省収蔵の保管第3表を閲覧すれば原因別各歳死亡者数が得られる．Table 5.4.2-a は1980-2001年の84-117歳の死亡者数である．Figure 5.1.2-a を見ると，83歳以下についても，平均余命関数が直線を示すことが予想されるため，83歳以下のデータも加えるべきだろう．

世代に分けると，約50人以上のデータが得られる，死亡確定できる世代の誕生年度は，男性が1880-1892年，女性が1878-1896年である．

Table 5.4.2-b に，レキス図上の断片 \mathcal{V}_{rs} ; $r = 95, 98, \dots, 113$, $s = 1980, 1983, \dots, 1998$ に PCIモデル (5.3.3-b) を当てはめた結果を示す．この断片領域は全員が死亡している3年連続した世代のなかで最も新しいもので，1885-87年世代である． γ の推定値およびその標準偏差から $\gamma = 0$ の仮説は棄却され，分布の上限 $\omega = -a/\gamma$ の推定値から寿命分布の限界は125-27歳と推定された．この結果は生存者数統計から推定されたものより小さいものである．

先に述べたように，PCIモデルの適用は，競合リスクを分離して推定できる利点がある．そこで，癌死亡について同様のモデルの適用を試みた結果， $\gamma > 0$ と推定され，その標準偏差から $\gamma = 0$ の仮説は棄却された(Table 5.4.2-b を参照)．つまり，超高齢者に関しては，癌による死亡危険率は年齢の増加とともに減少すると推定された．これは，ある高年齢を超えると，癌での死亡率は減るという一般の認識と一致する．

Table 5.4.2-a : 84-117歳の死亡者数 (男性).

年度	年齢																
	84歳	85歳	86歳	87歳	88歳	89歳	90歳	91歳	92歳	93歳	94歳	95歳	96歳	97歳	98歳	99歳	100歳
1980年	8208	7469	6206	5165	4241	3281	2903	2256	1570	1050	718	511	312	208	134	81	47
1981年	8587	7638	6588	5330	4348	3555	2804	2235	1692	1145	836	516	339	257	130	93	43
1982年	8999	7518	6618	5536	4586	3710	2827	2083	1743	1287	808	577	351	218	153	81	61
1983年	9403	8603	7127	6239	5206	4132	3316	2550	1811	1363	1004	648	425	235	159	123	64
1984年	9948	8614	7770	6342	5474	4436	3455	2593	1963	1414	1031	800	460	279	178	116	78
1985年	10501	9259	7879	7066	5773	4772	3847	2899	2210	1623	1139	846	577	360	222	122	70
1986年	10770	9684	8309	7095	6193	4954	4039	3181	2370	1755	1161	811	578	456	245	163	69
1987年	10888	10034	8820	7594	6273	5487	4275	3428	2638	1821	1281	910	616	429	293	164	88
1988年	11358	10958	9914	8666	7224	5981	5056	3765	2962	2137	1581	1078	724	474	325	217	128
1989年	10983	10564	10113	8755	7398	6235	4859	3988	3046	2297	1744	1217	844	561	319	222	135
1990年	11317	11041	10114	9413	8218	6837	5564	4344	3488	2555	1963	1384	938	604	397	224	159
1991年	12777	10780	10131	9310	8474	7301	5757	4679	3558	2738	1987	1445	977	705	431	264	164
1992年	13560	12156	10443	9443	8567	7617	6362	5122	3933	2828	2256	1543	1053	782	488	290	181
1993年	13648	12973	11565	9332	8757	7604	6583	5459	4315	3155	2382	1789	1174	850	551	339	186
1994年	13911	13114	12162	10581	8465	7426	6597	5638	4576	3447	2521	1804	1292	828	557	408	207
1995年	14839	13820	13000	11687	10117	7871	6992	6013	5125	4016	2815	2060	1434	1052	676	433	291
1996年	14122	13450	12490	11372	9877	8447	6567	5692	4765	3848	2926	2144	1452	989	746	421	279
1997年	14586	13805	12808	11837	10594	9192	7771	5769	4788	4014	3191	2326	1679	1114	776	487	303
1998年	14949	14118	13049	12079	10922	9600	8168	6689	4963	4142	3284	2557	1814	1242	793	491	318
1999年	15694	15178	14026	13123	11803	10273	9139	7621	5860	4347	3511	2692	2100	1431	904	626	375
2000年	15052	14266	14016	12670	11561	10217	8731	7695	6205	4848	3351	2556	2011	1409	1025	641	410
2001年	14574	14398	13811	13299	11807	10437	9229	8059	6471	5201	3908	2674	2067	1522	1062	721	481

年度	年齢																
	101歳	102歳	103歳	104歳	105歳	106歳	107歳	108歳	109歳	110歳	111歳	112歳	113歳	114歳	115歳	116歳	117歳
1980年	26	16	8	2	0	0	0	1	0	0	0	0	0	0	0	0	0
1981年	27	15	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1982年	19	13	6	5	1	0	0	2	1	0	0	0	0	0	0	0	0
1983年	28	18	13	5	7	1	1	0	0	0	0	0	0	0	0	0	0
1984年	35	14	12	7	4	1	0	0	0	0	0	0	0	0	0	0	0
1985年	56	31	17	8	4	5	0	1	0	0	0	0	0	0	0	0	0
1986年	57	25	16	8	6	2	0	0	0	0	0	0	0	0	0	0	0
1987年	44	30	11	6	2	0	0	0	0	0	1	0	0	0	0	0	0
1988年	77	35	26	8	4	1	2	1	0	0	0	0	0	0	0	0	0
1989年	89	33	19	11	8	4	0	0	0	0	0	0	0	0	0	0	0
1990年	97	50	33	11	5	5	4	1	1	1	0	0	0	0	0	0	0
1991年	91	67	30	15	3	2	1	3	1	0	0	0	0	0	0	0	0
1992年	91	50	30	18	7	2	1	0	0	0	0	0	0	0	0	0	0
1993年	114	62	32	19	9	3	2	1	0	0	0	0	0	0	0	0	0
1994年	135	94	43	14	10	3	3	2	0	0	0	0	0	0	0	0	0
1995年	158	79	43	30	18	14	5	0	1	0	0	0	0	0	0	0	0
1996年	190	98	41	27	10	3	3	1	1	0	0	0	0	0	0	0	0
1997年	178	131	56	24	18	14	3	4	1	1	0	1	0	0	0	0	0
1998年	192	111	66	38	14	5	3	1	0	0	0	0	0	0	0	0	0
1999年	257	134	94	43	20	14	6	2	0	0	0	1	0	0	0	0	0
2000年	243	159	79	42	29	6	1	3	1	1	1	0	0	0	0	0	0
2001年	265	169	84	52	27	19	6	1	2	1	0	0	0	0	0	0	0

Table 5.4.2-b：一般パレート分布に基づくPCIモデルによるパラメータ推定結果。

	全死亡	癌死亡
γ の推定値	-0.11	4.57
標準偏差	0.03	1.77
a の推定値	3.30	76.71
標準偏差	0.09	7.32
$-a/\gamma$ の推定値	30.00	-
$-a/\gamma$ を含む年齢区分	125-27歳	

擬似確率標本による解析結果

前述のように N_{ij} 個の3角乱数により $j-i$ 年度, $j-i-1$ 年度のデータを生成する。1880-1892 年度世代より1880-1891年度世代データを得る。このデータの標本平均余命も, Figure 5.1.2-a と同様に100歳頃までは常に直線的であるが, それ以上になると不安定であって, 水平に近くなる世代もある。

Table 5.4.2-c に, Coles (2001) の関数 “gpd.fit()” による推定の結果をまとめる。閾値は, 利用可能なすべてのデータを用いる場合と, 高齢者名簿とも対応する100歳を選んだ場合を表に含めた。同じ世代の2列の, 上段と下段である。推定値の標準偏差は漸近分散に推定値を入れて求めた。p値の計算も同様である。

Table 5.4.2-c：Coles (2001) の関数“gpd.fit()” による推定の結果。

generation	size	thrsh	$\hat{\gamma}$	SE($\hat{\gamma}$)	$\hat{\omega}$	SE($\hat{\omega}$)	$\hat{\alpha}$	$\hat{\eta}$	p-val
1880	85	100	-0.2067	0.086	112.6	13.1	0.171	4.84	0.008
1881	123	100	-0.1507	0.0766	117.9	23	0.131	6.63	0.025
1881	156	99.5	-0.0967	0.0723	125.4	47.4	0.088	10.34	0.091
1882	123	100	-0.0966	0.0815	120.2	37.8	0.088	10.36	0.118
1882	263	98.5	-0.0818	0.0566	123.3	38.6	0.076	12.22	0.074
1883	138	100	-0.1016	0.0765	121.3	37	0.092	9.84	0.092
1883	463	97.5	-0.0568	0.0438	135.5	67.7	0.054	17.6	0.097
1884	195	100	-0.0234	0.0699	178.9	510.3	0.023	42.67	0.369
1884	873	96.5	-0.1246	0.0296	117	12.1	0.111	8.03	0
1885	213	100	-0.1813	0.0561	112.9	9.5	0.153	5.52	0.001
1885	1228	95.5	-0.1223	0.025	118.3	11.9	0.109	8.18	0
1886	188	100	-0.0205	0.0714	198.6	772.6	0.02	48.88	0.387
1886	1966	94.5	-0.0885	0.0206	124.5	17.5	0.081	11.3	0
1887	234	100	-0.173	0.0541	113	9.6	0.147	5.78	0.001
1887	3020	93.5	-0.1202	0.016	117.9	8.5	0.107	8.32	0
1888	306	100	-0.0738	0.0529	129.1	47.9	0.069	13.54	0.082
1888	4814	92.5	-0.1208	0.0127	119.2	7.6	0.108	8.28	0
1889	373	100	-0.0756	0.0479	129	42.6	0.07	13.24	0.057
1889	7286	91.5	-0.1272	0.0102	118.6	6.1	0.113	7.86	0
1890	379	100	-0.1649	0.0429	113	7.8	0.142	6.06	0
1890	10587	90.5	-0.1676	0.0081	112.8	3.1	0.144	5.97	0

(size: 閾値を超えた観測数, thrsh: 閾値, p-val: 仮説 $\gamma = 0$ の下での $\hat{\gamma}$ のp値.)

p値がかなり大きい場合もあるが、 γ の推定値はすべて負である。 ω の推定値がかなり大きいものと小さいものがあるが、閾値を100歳に制限しない場合はだいたい118-135に納まっている。古い世代はデータが少ないこともあり、推定区間の幅が広い。なお、このデータを解析した時期には、中願寺さんが生存中であつたが、あえて中願寺を除外し、1889年世代も対象として解析した。また推定誤差には、擬似確率標本作成の影響を考慮していない。

5.5. 結び

5.5.1. 結論

全国高齢者名簿および人口動態統計より得た、年齢時代区分別の超高齢者の生存者数統計および死亡者統計を極値理論を用いて解析し、寿命分布の限界を議論した。連続変数に対して議論している極値理論を、年齢時代区分別に作表したデータへ適用することが必要である。そのため、一般パレート分布関数から計算した死亡確率を用いて多項分布モデルを当てはめる方法、乱数を加えて得られる連続量擬似確率標本を用いる方法を提案し、レキシス・ダイアグラム上のPCIモデルを適用した。推定の結果、日本人男性寿命分布の有限な上限値が得られた。

5.5.2. リマーク (Remark)

Remark 5.1:

Aaarssen, de Haan (1994) は1877-1881年に生まれ、1971年1月1日に生存しており、オランダ住民として死亡した合計約1万人の性別、日単位享年のデータについて、順序統計量の対数に基づくモーメント推定量を用いて、 $\gamma = 0$ の仮説を棄却し、 ω の推定量113-124、 α の推定量0.07-0.14を導いている。また、Kaufmann (2001) は1993年西ドイツ女性90歳以上の年齢別死亡数を解析して、 $\gamma = 0$ の仮説を否定できないと主張している。

このように極値理論を適用した研究結果でも $\gamma < 0$ 、 $\gamma = 0$ の両論がある。ただしKaufmann (2001) は年度のデータを用い、世代の寿命を調べていない。Aaarssen, de Haan (1994) の結論は本稿と同じである。しかし本論文の筆者は、彼らの結論をさらに裏付けたと確信しているわけでもない。

Remark 5.2:

現在、高齢者の死因では癌と循環器病が大きな割合を占めており、その割合が死亡年齢により異なる。全死亡者中の癌死亡者の割合を見ると年齢とともに減少の傾向がある。しかも前述のように循環器病死亡者数の割合が多い100歳以上では平均余命が減少しない。このように対抗する死因が混合している場合、極値理論に立つならば、さらに閾値を大きくして、平均余命関数が直線となるように選ぶことになるが、現実にはデータが不足で確認できない。

付録A (3.3.3-aの証明)

$J - \phi \geq 3$ のとき，行列 \mathbf{F} の階数は明らかに列ベクトルの数に等しいので， $\mathbf{X}\mathbf{g} = \mathbf{0}$ の必要十分条件は， $\mathbf{F}\mathbf{g} \in R(\mathbf{Z}')^\perp$ である．よって

$$R(\mathbf{X}')^\perp = \{\mathbf{g} \mid \mathbf{F}\mathbf{g} \in R(\mathbf{Z}')^\perp\}, \quad (\text{A.1})$$

である．ただし， $R(\cdot)^\perp$ は $R(\cdot)$ の直交補空間である．ここで，下のように定義されるベクトルを考えよう．

$$\begin{aligned} \mathbf{m}_1 &= (-1, \mathbf{1}_I, \mathbf{0}_J, \mathbf{0}_{I+J-1})', \\ \mathbf{m}_2 &= (-1, \mathbf{0}_I, \mathbf{1}_J, \mathbf{0}_{I+J-1})', \\ \mathbf{m}_3 &= (-1, \mathbf{0}_I, \mathbf{0}_J, \mathbf{1}_{I+J-1})', \\ \mathbf{m}_4 &= (\phi, 1, 2, \dots, I, J-1, J-2, \dots, 0, 1-I-J-\phi, 2-I-J-\phi, \dots, -1-\phi)'. \end{aligned}$$

すると， $R(\mathbf{Z}')^\perp = R((\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \mathbf{m}_4))$ であることと，

$$R(\mathbf{F}) \cap R(\mathbf{Z}')^\perp = R((\mathbf{m}_1, \mathbf{m}_4)) \quad (\text{A.2})$$

であることから， $\mathbf{F}\mathbf{g} \in R(\mathbf{Z}')^\perp$ の必要十分条件は $\mathbf{F}\mathbf{g} \in R((\mathbf{m}_1, \mathbf{m}_4))$ である．したがって， $J - \phi \geq 3$ の場合には， \mathbf{X} の階数は， \mathbf{X} の列ベクトルの数から2を引いた数，すなわち， $2I + J + \phi - 2$ であり， $R(\mathbf{X}) \subset R(\mathbf{Z})$ であることが分かる．

$J - \phi \leq 2$ の場合には， \mathbf{X} に含まれる $\xi_{J-\phi}, \dots, \xi_1$ に対応した列ベクトルの全てが $\mathbf{1}_{I \times J}'$ ベクトルで，このことから，これらのパラメータ（効果）は総平均 μ_c と識別できないことが分かる．これは，観測された人々全員が時代区分 $[P_{J-\phi-1}, P_1]$ における環境リスクに暴露されていることに起因すると解釈できる．そこで， \mathbf{X} から $\xi_{J-\phi}, \dots, \xi_1$ に対応する列ベクトルを取り除いた行列を \mathbf{X}^* で表し，方程式の系 $\boldsymbol{\eta} = \mathbf{X}^* \boldsymbol{\theta}^*$ を考えよう．ただし， $\boldsymbol{\theta}^* \equiv (\mu, \alpha_1, \dots, \alpha_I, \xi_{2-I-\phi}, \dots, \xi_{J-\phi-1}, \xi_2, \dots, \xi_J)'$ である．すると，明らかに， $\boldsymbol{\eta} = \mathbf{X}^* \boldsymbol{\theta}^*$ は，APCモデルにおいてパラメータを累積表示したもので， $R(\mathbf{X}) = R(\mathbf{X}^*)$ である．よって， $R(\mathbf{X}) = R(\mathbf{Z})$ で， \mathbf{X} の階数は $2I + 2J - 4$ である．

付録B (3.4.2-aの証明)

次のように定義されるベクトルを考えよう．

$$\begin{aligned}\mathbf{h}_1 &= (-1, \mathbf{1}_I, \mathbf{0}_{I+J-1+\phi-\rho})', \\ \mathbf{h}_2 &= (\phi-\rho, 1, 2, \dots, I, -\mathbf{1}_{I+J-1+\phi-\rho})' .\end{aligned}$$

すると， $\mathbf{m}_1 = \mathbf{Fh}_1$ ， $\mathbf{m}_4 = \mathbf{Fh}_2$ と書け，このことおよび付録Aの(A.1)と(A.2)から $R(\mathbf{X}')^\perp = R(\mathbf{h}_1, \mathbf{h}_2)$ であることが分かる．したがって θ の線形結合($\mathbf{c}'\theta$ と書く)を考えるとき，方程式の系 $\eta = \mathbf{X}\theta$ を満たす任意の θ に対し $\mathbf{c}'\theta$ が一意に定まるための必要十分条件は $\mathbf{c}'(\mathbf{h}_1, \mathbf{h}_2) = (0, 0)$ である．よって， $R(\mathbf{h}_1, \mathbf{h}_2)^\perp$ の基底ベクトルを考えることによって，AEモデルの推定可能関数は(3.4.2-a)とその線形結合によって表されることが分かる．

付録C (Proposition 4.3.2 の証明)

$\mu(\mathbf{q}(\psi_1^{(1)}, \boldsymbol{\theta}_1)) = \mu(\mathbf{q}(\psi_1^{(2)}, \boldsymbol{\theta}_2))$ を直接解くことによって示すことができる．便利のため，以下の記法を使用する．

$$\mathbf{q}(\psi_1^{(r)}, \boldsymbol{\theta}_r) = \mathbf{q}_r \equiv (p_1^{(r)}, \dots, p_I^{(r)}, q_{2-I}^{(r)}, \dots, q_J^{(r)})' ; r = 1, 2 ,$$

$$\Delta\mu_{ij}(\mathbf{q}_r) \equiv \mu_{ij}(\mathbf{q}_r) - \mu_{i-1,j}(\mathbf{q}_r) \quad (i \geq 2) ; \Delta\mu_{1j}(\boldsymbol{\theta}) \equiv \mu_{1j}(\boldsymbol{\theta}) .$$

すると，命題は

$$\Delta\mu_{ij}(\mathbf{q}_1) = \Delta\mu_{ij}(\mathbf{q}_2) \quad \text{for } i = 1, \dots, I; j = 1, \dots, J \quad (\text{C.1.1})$$

であるための必要十分条件が $\mathbf{q}_1 = \mathbf{q}_2$ であることを示すことにより証明される．(以下では， $\Delta\mu_{ij}(\mathbf{q}_r) = q_{j-i+1}^{(r)} p_i^{(r)}$; $r = 1, 2$ であることに注意．)

まず， \mathbf{q}_r ; $r = 1, 2$ の定義から， $p_1^{(1)} = p_1^{(2)} = 1$ が直ぐに分かる．そして，方程式

$$q_k^{(1)} p_1^{(1)} = \Delta\mu_{1k}(\mathbf{q}_1) = \Delta\mu_{1k}(\mathbf{q}_2) = q_k^{(2)} p_1^{(2)} \quad \text{for } k = 1, \dots, J$$

から，

$$q_k^{(1)} = q_k^{(2)} \quad \text{for } k = 1, \dots, J \quad (\text{C.2.1})$$

を得る．さらに，(C.2.1) と

$$q_1^{(1)} p_l^{(1)} = \Delta\mu_{1l}(\mathbf{q}_1) = \Delta\mu_{1l}(\mathbf{q}_2) = q_1^{(2)} p_l^{(2)}; \quad l = a_1, \dots, b_1$$

から，

$$p_l^{(1)} = p_l^{(2)} \quad \text{for } l = 1 (= a_1), \dots, b_1 \quad (\text{C.2.2})$$

を得る．

ところで， $\mu(\mathbf{q}_r) \in M_0$ と仮定しているため，

$$a_1 + 1 = a_0 \leq l_* \leq b_0 - 1 \leq b_1 \quad \text{and } p_{l_*}^r \neq 0 \quad \text{for } r = 1, 2$$

であるような l_* が少なくとも1つ存在する．したがって，方程式

$$q_0^{(1)} p_{l_*}^{(1)} = \Delta\mu_{l_*, l_*-1}(\mathbf{q}_1) = \Delta\mu_{l_*, l_*-1}(\mathbf{q}_2) = q_0^{(2)} p_{l_*}^{(2)}$$

から

$$q_0^{(1)} = q_0^{(2)} \quad (\text{C.3.1})$$

を得る．さらに，方程式

$$q_0^{(1)} p_l^{(1)} = \Delta\mu_{l, l-1}(\mathbf{q}_1) = \Delta\mu_{l, l-1}(\mathbf{q}_2) = q_0^{(2)} p_l^{(2)} \quad \text{for } l = a_0, \dots, b_0$$

から

$$p_l^{(1)} = p_l^{(2)} \text{ for } l = 1, \dots, b_0 \quad (\text{C.3.2})$$

を得る .

ここで , この命題を帰納的に示すために ,

$$q_k^{(1)} = q_k^{(2)} \text{ for } k = J, J-1, \dots, 0, \dots, 3-i \quad (\text{C.4.1})$$

と仮定しよう . この仮定と , 方程式

$$q_{3-i}^{(1)} p_l^{(1)} = \Delta \mu_{l, 2-i+l}(\mathbf{q}_1) = \Delta \mu_{l, 2-i+l}(\mathbf{q}_2) = q_{3-i}^{(2)} p_l^{(2)} \text{ for } l = a_{3-i}, \dots, b_{3-i}$$

から直ぐに

$$p_l^{(1)} = p_l^{(2)} \text{ for } l = 1, \dots, b_{3-i} \quad (\text{D.4.2})$$

が得られる .

このとき , (C.3.1) を導出した場合と同様に , $\mu(\mathbf{q}_r) \in M_0$ の仮定から ,

$$a_{3-i} + 1 = a_{2-i} \leq l_* \leq b_{2-i} - 1 \leq b_{3-i} \text{ and } p_{l_*}^{(r)} \neq 0 \text{ for } r = 1, 2$$

であるような l_* が少なくとも1つ存在する . したがって , 方程式

$$q_{2-i}^{(1)} p_{l_*}^{(1)} = \Delta \mu_{l_*, 1-i+l_*}(\mathbf{q}_1) = \Delta \mu_{l_*, 1-i+l_*}(\mathbf{q}_2) = q_{2-i}^{(2)} p_{l_*}^{(2)}$$

から

$$q_k^{(1)} = q_k^{(2)} \text{ for } k = J, J-1, \dots, 0, \dots, 2-i \quad (\text{D.5.1})$$

を得る . さらに , (C.5.1) と方程式

$$q_{2-i}^{(1)} p_l^{(1)} = \Delta \mu_{b_{2-i}, 1-i+l}(\mathbf{q}_1) = \Delta \mu_{b_{2-i}, 1-i+l}(\mathbf{q}_2) = q_{2-i}^{(2)} p_l^{(2)} \text{ for } l = a_{2-i}, \dots, b_{2-i}$$

から ,

$$p_l^{(1)} = p_l^{(2)} \text{ for } l = 1, \dots, b_{2-i} \quad (\text{C.5.2})$$

を得る .

このように , $\mathbf{q}_1 = \mathbf{q}_2$ が帰納的に示された .

付録D (Property 4.4.2-b の証明)

まず, STSモデルの形から直ぐに

$$\frac{\partial \mu_{ij}((1, \theta_*'))}{\partial q_l} = \begin{cases} \xi_{j-l} & \text{if } i \geq 2 \text{ and } 2 \leq l \leq i \\ \psi_{l-2I+1-j+i} & \text{if } i \geq 2 \text{ and } j-i+1 \leq l-2I+1 \leq j \\ 1 & \text{if } i=1 \text{ and } l-2I+1 = j \\ 0 & \text{otherwise,} \end{cases}$$

$$\frac{\partial^2 \mu_{ij}((1, \theta_*'))}{\partial q_l \partial q_m} = \begin{cases} 1 & \text{if } i \geq 2, 2 \leq l \leq i \text{ and } j-i+1 \leq m-2I+1 \leq j \text{ or} \\ & \text{if } i \geq 2, 2 \leq m \leq i \text{ and } j-i+1 \leq l-2I+1 \leq j \\ 0 & \text{otherwise.} \end{cases}$$

が分かる. このとき,

$$\frac{2 \left[\max \left\{ 1, \left| \frac{\psi_2^0}{\psi_1^0} \right|, \dots, \left| \frac{\psi_I^0}{\psi_1^0} \right|, \left| \psi_1^0 \xi_{1-I}^0 \right|, \dots, \left| \psi_1^0 \xi_{J-1}^0 \right| \right] \right]^3}{\mu_{ij}^3 \left(\left(1, \frac{\psi_2^0}{\psi_1^0}, \dots, \frac{\psi_I^0}{\psi_1^0}, \psi_1^0 \xi_{1-I}^0, \dots, \psi_1^0 \xi_{J-1}^0 \right)' \right)}$$

$$+ \frac{3 \left[\max \left\{ 1, \left| \frac{\psi_2^0}{\psi_1^0} \right|, \dots, \left| \frac{\psi_I^0}{\psi_1^0} \right|, \left| \psi_1^0 \cdot \xi_{1-I}^0 \right|, \dots, \left| \psi_1^0 \cdot \xi_{J-1}^0 \right| \right] \right]^2}{\mu_{ij}^2 \left(\left(1, \frac{\psi_2^0}{\psi_1^0}, \dots, \frac{\psi_I^0}{\psi_1^0}, \psi_1^0 \xi_{1-I}^0, \dots, \psi_1^0 \xi_{J-1}^0 \right)' \right)},$$

よりも十分大きく ρ を選ぶことによってProperty 4.4.2-b が確かめられる.

付録E (Property 4.4.2-c の証明)

Condition 4.4.2-c は, $\mathbf{q} = \mathbf{0}$ であるための必要十分条件が

$$\mathbf{q}' \frac{\partial \Delta \mu_{ij}((1, \boldsymbol{\theta}_*'))}{\partial \boldsymbol{\theta}_*} \Big|_{\boldsymbol{\theta}_* = \boldsymbol{\theta}^0} = 0 \text{ for all } i \text{ and } j \quad (\text{E.1})$$

であることを示すことによって確かめられる。ただし,

$$\mathbf{q} \equiv (p_2, \dots, p_I, q_{2-I}, \dots, q_J)' \in R^{2I+J-2};$$

$$\Delta \mu_{ij}(\boldsymbol{\theta}) \equiv \mu_{ij}(\boldsymbol{\theta}) - \mu_{i-1,j}(\boldsymbol{\theta}) \quad (i \geq 2), \quad \Delta \mu_{1j}(\boldsymbol{\theta}) \equiv \mu_{1j}(\boldsymbol{\theta})$$

である。まず最初に, 方程式

$$\mathbf{q}' \frac{\partial \Delta \mu_{1,j}((1, \boldsymbol{\theta}_*'))}{\partial \boldsymbol{\theta}_*} \Big|_{\boldsymbol{\theta}_* = \boldsymbol{\theta}^0} = q_j = 0 \text{ for } j = 1, \dots, J$$

から

$$q_j = 0 \text{ for } j = 1, \dots, J \quad (\text{E.2})$$

が分かる。次に, 方程式

$$\mathbf{q}' \frac{\partial \Delta \mu_{i,i}((1, \boldsymbol{\theta}_*'))}{\partial \boldsymbol{\theta}_*} \Big|_{\boldsymbol{\theta}_* = \boldsymbol{\theta}^0} = \frac{\psi_i^0}{\psi_1^0} q_1 + \xi_1^0 \psi_1^0 p_i = 0 \text{ for } i = 2, \dots, b_1,$$

を解いて,

$$p_i = 0 \text{ for } i = 2, \dots, b_1 \quad (\text{E.3})$$

が, 仮定 $\xi_1^0 > 0$ と (E.2) から得られる。

ここで, 仮定 $\boldsymbol{\mu}(\mathbf{q}_r) \in M_0$, すなわち $\boldsymbol{\mu}((1, \boldsymbol{\theta}_*')) \in M_0$ から,

$$a_1 + 1 = a_0 \leq i_* \leq b_0 - 1 \leq b_1, \quad \psi_{i_*}^0 \neq 0$$

であるような i_* が少なくとも1つ存在する。したがって,

$$q_0 = 0. \quad (\text{E.4})$$

が, 方程式

$$\mathbf{q}' \frac{\partial \Delta \mu_{i_*, i_*-1}((1, \boldsymbol{\theta}_*'))}{\partial \boldsymbol{\theta}_*} \Big|_{\boldsymbol{\theta}_* = \boldsymbol{\theta}^0} = \frac{\psi_{i_*}^0}{\psi_1^0} q_0 + \xi_0^0 \psi_1^0 p_{i_*} = 0$$

から得られる。さらに, 方程式

$$\mathbf{q}' \frac{\partial \Delta \mu_{b_0, b_0-1}((1, \boldsymbol{\theta}'_*))}{\partial \boldsymbol{\theta}_*} \Big|_{\boldsymbol{\theta}_* = \boldsymbol{\theta}_*^0} = \frac{\psi_{b_0}^0}{\psi_1^0} q_0 + \xi_0^0 \psi_1^0 p_{b_0} = 0$$

を解くことによって,

$$p_{b_0} = 0 \quad (\text{E.5})$$

が, 仮定 $\xi_0^0 > 0$ と (E.4) から得られる.

ここで, 命題

$$\mathbf{q} = \mathbf{0} \leftrightarrow \mathbf{q}' \frac{\partial \Delta \mu_{ij}((1, \boldsymbol{\theta}'_*))}{\partial \boldsymbol{\theta}_*} \Big|_{\boldsymbol{\theta}_* = \boldsymbol{\theta}_*^0} = 0$$

を, 帰納的に示すために,

$$q_{2-l} = 0 \quad \text{for } l = 2, \dots, i \quad (\text{E.6})$$

と仮定しよう. このとき, 方程式

$$\mathbf{q}' \frac{\partial \Delta \mu_{l1}((1, \boldsymbol{\theta}'_*))}{\partial \boldsymbol{\theta}_*} \Big|_{\boldsymbol{\theta}_* = \boldsymbol{\theta}_*^0} = \frac{\psi_l^0}{\psi_1^0} q_{2-l} + \xi_{2-l}^0 \psi_1^0 p_l = 0 \quad \text{for } l = 2, \dots, i$$

を解くことによって,

$$p_l = 0 \quad \text{for } l = 2, \dots, i \quad (\text{E.7})$$

が仮定 $\xi_{2-l}^0 > 0$ と (E.6) から得られる.

(E.4) を示した場合と同様に, 仮定 $\boldsymbol{\mu}(\mathbf{q}_r) \in M_0$, すなわち $\boldsymbol{\mu}((1, \boldsymbol{\theta}'_*)) \in M_0$ から,

$$a_{2-i} + 1 = a_{1-i} \leq i_* \leq b_{1-i} - 1 \leq b_{2-i} \quad \text{and } \psi_{i_*}^0 \neq 0$$

であるような i_* が少なくとも1つ存在するから, 方程式

$$\mathbf{q}' \frac{\partial \Delta \mu_{i, i_*-i}((1, \boldsymbol{\theta}'_*))}{\partial \boldsymbol{\theta}_*} \Big|_{\boldsymbol{\theta}_* = \boldsymbol{\theta}_*^0} = \frac{\psi_{i_*}^0}{\psi_1^0} q_{1-i} + \xi_{1-i}^0 \psi_1^0 p_{i_*} = 0$$

を解いて,

$$q_{1-i} = 0 \quad (\text{E.8})$$

が得られる. さらに, これらの結果と, 方程式

$$\mathbf{q}' \frac{\partial \Delta \mu_{b_{1-i}, b_{1-i}-1}((1, \boldsymbol{\theta}'_*))}{\partial \boldsymbol{\theta}_*} \Big|_{\boldsymbol{\theta}_* = \boldsymbol{\theta}_*^0} = \frac{\psi_{b_{1-i}}^0}{\psi_1^0} q_{1-i} + \xi_{1-i}^0 \psi_1^0 p_{b_{1-i}} = 0$$

から,

$$p_i = 0 \quad \text{for } i = 2, \dots, b_{1-i}. \quad (\text{E.9})$$

が得られる．

このように，

$$\mathbf{q} = \mathbf{0} \leftrightarrow \mathbf{q}' \frac{\partial \Delta \mu_{ij}((1, \boldsymbol{\theta}'_*))}{\partial \boldsymbol{\theta}_*} \Big|_{\boldsymbol{\theta}_* = \boldsymbol{\theta}_*^0} = 0$$

が示され，これから，Property 4.4.2-c が示された．

参考文献

- Andersen, P.K., Borgan, Ø., Gill, R.D., Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer-Verlag: New York.
- Armitage, P., Doll, R. (1954). The age distribution of cancer and multi-stage theory of carcinogenesis. *British Journal of Cancer*, Vol. 8, 1-12.
- Arssen, K., de Haan, L. (1994). On the maximal life span of humans. *Mathematical Population Studies*, Vol. 4, 259-281.
- Berzuini, C., Clayton, D. (1994). Bayesian analysis of survival on multiple time scale. *Statistics in Medicine*, Vol. 13, 823-838.
- Breslow, N.E. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics*, Vol. 33, 38-44.
- Carey, R. J., Liedo, P., Orozco, D., Vaupel, W.J. (1992). Slowing of mortality rates at older ages in large medfly cohorts, *Science*, Vol. 258, 457-461.
- Clayton, D., Schifflers E. (1987 a). Model for temporal variation in cancer rates I: Age-period and age-cohort models. *Statistics in Medicine*, Vol. 6, 449-467.
- Clayton, D., Schifflers, E. (1987 b). Model for temporal variation in cancer rates II: Age-period-cohort model. *Statistics in Medicine*, Vol. 6, 469-481.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, London.
- Davis, D.L., *et al.* (1993). Medical hypothesis, xenoestrogens as preventable causes of breast cancer. *Environmental Health Perspectives*, Vol. 101, 372-377.
- Harman, D. (1956). Aging: A Theory Based on Free Radical and Radiation Chemistry. *Journal of Gerontology*, Vol. 11, 298-300.
- Dewandji, A., Goddard, M.J., Moolgavkar, S.H. (1999). Two stage model for carcinogenesis: number and size distributions of premalignant clones in longitudinal studies. *Mathematical Biosciences*, Vol. 155, 1-12.
- Glenn, N.D., (1977). *Cohort analysis*. Sage Publications: London.
- Hanayama, N. (2001). A simple two-stage model for cancer risk in the environment. *Environmetrics*, Vol. 12, 757-773.
- Hanayama, N. (2003). Age-environment model for breast cancer. *Environmetrics*, Vol. 15, 219-232.
- Hayflick, L., Moorhead, P.S. (1961). The serial cultivation of human diploid cell strain. *Exp. Cell Res.*, Vol. 25, 585-621.

- Hirotsu, C. (1988). A class of estimable contrasts in an age-period-cohort model. *Ann. Inst. Statist. Math.*, Vol. 40, No. 3, 451-465.
- Holford, T.R. (1983). The estimation of age, period and cohort effects for vital rate. *Biometrics*, Vol. 39, 311-324.
- Holford, T.R., Zhang, Z., McKay, L.A. (1994). Estimating age, period and cohort effects using the multistage model for cancer. *Statistics in Medicine*, Vol. 13, 23-41.
- Kannisto, V. (1999). Trends in the mortality of the oldest-old, *Statistics, Registries, and Science: Experiences from Finland* (ed. Alho, J.), Statistics Finland, 177-194.
- Kaufmann, E. (2001). About the longevity of humans, Section 14.2, *Statistical Analysis of Extreme Values* (written by Reiss, R.D., Thomas, M.) 2nd ed., Birkhauser: Basel.
- Keiding, N. (1975). Maximum likelihood estimation in the birth-and-death process. *Annals of Statistics*, Vol. 3, 363-372.
- Keiding, N. (1990). Statistical inference in the Lexis diagram. *Philosophical Transactions of the Royal Society of London Series A*, Vol. 332, 487-509.
- Keiding, N. (1999). Lexis diagram, *Encyclopedia of Statistical Sciences* (ed. Kotz, S., Read, C.B., Banks, D.L.), Update Vol. 3, Wiley: New York, 411-413.
- Knudson, Jr, A.G. (1971). Mutation and cancer: statistical study of retinoblastoma, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 68, 820-823.
- Kripke, M.L. (1986). The University of Texas M.D. Anderson Symposium on Fundamental Cancer Research vol. 38, *Immunology and Cancer* (ed. Kripke, M.L., Frost, M.D.P.), University of Texas Press: Austin, 113-120.
- Kupper, L.L., *et al.* (1985). Statistical age-period-cohort analysis. *Journal of Chronic Diseases*, Vol. 38, 811-830.
- Kuroki, T., *et al.* (1984). *An Outline of Medical Science*, Vol. 5, Cancer, Nakayama-shoten: Tokyo. (in Japanese).
- Lehmann, E.L., Casella, G. (1998). *Theory of Point Estimation* 2nd ed. Springer-Verlag: New York, 461-465.
- McCullagh, P., Nelder, J. A. (1983). *Generalized Linear Model* 2nd ed. Chapman & Hall: London, 193-244.
- Moolgavkar, S.H. (1978). The multistage theory of carcinogenesis and the age distribution of cancer in man. *Journal of the National Cancer Institute*, Vol. 61, 49-52.
- Moolgavkar, S.H., Venzon, D.J. (1979). Two event models for carcinogenesis: Incidence curves for

- childhood and adult tumors. *Mathematical Biosciences*, Vol. 47, 55-77.
- Osmond, C., Gardener, M.J. (1982). Age, period and cohort models applied to cancer mortality. *Statistics in Medicine*, Vol. 1, 245-259.
- Osmond, C., Gardener, M.J. (1989). Age, period, and cohort models: Non-overlapping cohorts don't resolve the identification problem. *American Journal of Epidemiology*, Vol. 129, No. 1, 31-35.
- Robertson, C., Boyle, P. (1986). Age, period and cohort model : The use of individual record. *Statistics in Medicine*, Vol. 5, 527-538.
- Robertson, C., Boyle, P. (1998). Age-period-cohort analysis of chronic disease rates. I: Modeling Approach. *Statistics in Medicine*, Vol. 17, 1305-1323.
- Rostgaard, K., *et al.* (2001). Age-period-cohort modeling of breast cancer incidence in the Nordic countries. *Statistics in Medicine*, Vol. 20, 47-61.
- Scheffé, H. (1959). *The Analysis of Variance*, John Wiley & Sons: New York, 14-19.
- Smith, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, Vol. 72, 67-90.
- Smith, R.L. (1987). Estimating tails of probability distributions, *Annals of Statistics*, Vol. 15, 1174-1207.
- Tango, T., Kurashina, S. (1987). Age period and cohort analysis of trends in mortality from major disease in Japan, 1955 to 1979: Peculiarity of the cohort born in the early Showa era. *Statistics in Medicine*, Vol. 6, 709-726.
- Tango, T. (1988). Letter to the editor Re: "Statistical modeling of lung cancer and laryngeal cancer incidence in Scotland, 1960-1979". *American Journal of Epidemiology*, Vol. 127, No. 3, 677-678.
- UNSCEAR (United Nations Scientific Committee on the Effect of Atomic Radiation) (1994). Sources, effects and risks of ionizing radiation, United Nations: New York.
- Voet, D., Voet, J.G. (1995). *Biochemistry*, John Wiley & Sons: New York.
- Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models. *Biometrika*, Vol. 61, 439-447.
- Wedderburn, R.W.M. (1976). On the existence and uniqueness of the likelihood estimates for certain generalized linear models. *Biometrics*, Vol. 63, 27-32.
- Wolff M. S., Weston A. (1997). Breast cancer risk and environmental exposures. *Environmental Health Perspectives*, Vol. 105 (Suppl 4), 891-896.
- 赤井豊秋, 渋谷政昭 (1985). 年度, 年齢, 世代の効果をもつ線形模型. *応用統計学*, Vol. 14, No. 1, 1-13.

渋谷政昭, 華山宣胤 (2004). 年齢・時代区分データによる超高齢者寿命分布の推測. 統計数理, 第52巻, 第1号. (掲載予定.)

中村 隆 (1982). ベイズ型コウホート・モデル - 標準コウホート表への適用 - . 統計数理研究所彙報, 第29巻, 第2号, 77-96.

華山宣胤 (1999). 時代環境の影響を解析するための応答モデル. 応用統計学, 第28巻, 1-20.

華山宣胤 (2002). Simple Two Stage Model による日米癌死亡データの解析. コンピュータサイエンス, Vol. 7, No. 1, 49-58.

謝辞

本論文の完成に至る道程でいくつもの困難があったが、それを乗り越えるための知識と心構えを授けて下さった主査の篠崎信雄教授に心より感謝する。また、本論文に掲載した研究成果を上げられたのは、指導者であり共同研究者である渋谷政昭教授に依るところが大きく、教授のご尽力に深く感謝する。

さらに、副査をお願いした、柳本武美教授、森雅夫教授、および栗田治教授から、論文作成のための有益なご意見を頂いたことは、感謝の極みである。

最後に、統計学の研究者に成りたい、という私の希望を理解し、博士課程へ導いて下さった鷲尾泰俊名誉教授に、感謝と敬意を表する。