

年齢時代区分データによる 超高齢者寿命分布の推測

渋谷 政昭¹・華山 宣胤²

(受付 2003 年 8 月 4 日; 改訂 2004 年 3 月 19 日)

要 旨

年齢時代区別に観測された超高齢者の生存者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する。使用したデータは厚生労働省(厚生省)より発表されている全国高齢者名簿および人口動態統計より得た。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た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する。また、既存の手法としてレキシス図上の Piecewise Constant Intensity Model も適用する。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。

キーワード: コホート分析, 最尤法, 人口動態統計, 全国高齢者名簿, Piecewise Constant Intensity Model, 平均余命関数。

1. まえがき

超高齢者の数が十分に多くなり、その年齢も信頼できるようになった。ここでは人口動態統計データ(死亡統計)および全国高齢者名簿(生存統計)を用いて、世代とともに高齢者の寿命分布がどのように変化しているかを解析する。日本では平均寿命が着実に増加しているが、高齢者寿命分布ではどのような変化が起きているか調べたい。特に極値理論を利用して分布限界が有限、無限を含むモデルを考え、寿命分布の上限を推定する。

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。

¹ 高千穂大学 経営学部: 〒168-8508 東京都杉並区大宮 2-19-1; sibuyam@takachiho.ac.jp

² 尚美学園大学 芸術情報学部: 〒350-1153 埼玉県川越市下松原 655; nob-hanayama@jcom.home.ne.jp

記録的な自然災害にたいする対策，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．

死亡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図 1 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．図

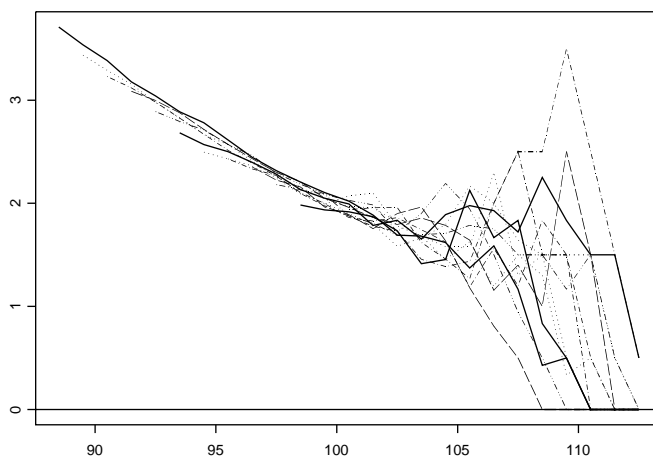
Mean residual life Ma 1880 - 1892

図 1. 死亡統計の世代ごと標本平均余命(年齢-年).

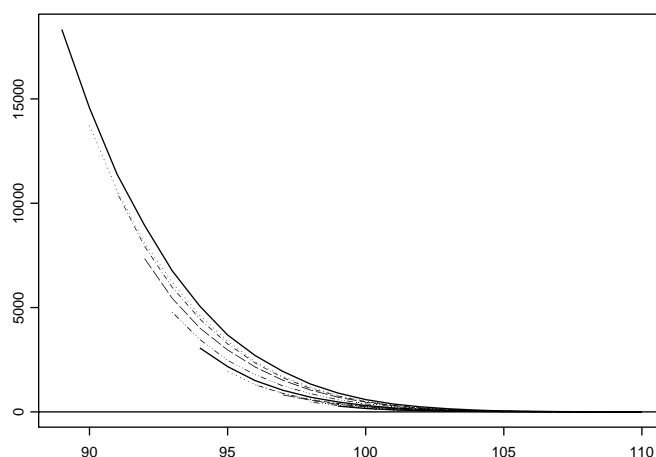
#deaths older than x-axis Ma 1880 - 1892

図 2. 死亡統計の世代ごと生存者数(年齢-年).

2 は同じ死亡統計で、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、図 1 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 節の始めて説明する。) 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(たとえば Kannisto(1999))では寿命分布に上限がない。

このように、高齢者の寿命を極値理論に基づいて議論することは有意義であるように思われるが、実際に我々が解析しようとしている高齢者寿命に関するデータは、年齢時代区分で作表されたデータ(いわゆる不完全データの種類)である。したがって、連続変量を前提として議論されている極値理論をこのデータの解析へ直接適用することはできない。そこで、本論文の要点は、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におけるコホート分析(Osmond and Gardener(1982))と結びつけるところにある。また、副次的に死亡統計と生存統計のコホートの違いを議論する。

2. 一般パレート分布

定義と諸性質

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

$$(2.1) \quad F(x) = F(x; \gamma, a) = \begin{cases} 1 - (1 + \gamma x/a)^{-1/\gamma}, & \gamma \neq 0, \\ 1 - \exp(-x/a), & \gamma = 0, \end{cases}$$

に従う。その分布範囲は

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

である。(2.1) 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。 $\gamma \geq 1$ では期待値が存在しない。災厄関数は

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

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

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

ただし、 $\gamma < 0$ のとき、 $y < a/(-\gamma) - x_0$ 。つまり $X - x_0 | X \geq x_0 \sim \text{GPrt}(\gamma, a + \gamma x_0)$ 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。この性質から条件付期待値(平均余命 mean residual life, 平均超過関数 mean excess function) $m(x_0)$ 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。

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

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

ただし、 $\{\}^{-}$ は、 $\{\}$ 内の関数の逆関数を表わす．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 (2.1) を導ける．

γ が負か零か

データ $\{x_j; j = 1, 2, \dots, n\}$ に (2.1) を当てはめたとき、 $\gamma < 0$ か $\gamma = 0$ か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．

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

$\gamma < 0$ の場合には、

$$(2.2) \quad F(x; \eta, \omega) = 1 - (1 - x/\omega)^\eta, \quad 0 \leq x < \omega; \quad \eta = -1/\gamma, \quad \omega = a/(-\gamma),$$

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

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

は生続率(perseverance parameter)と呼ばれている．災厄関数は

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

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“force of mortality”(ここでは災厄関数)が指数的に増加する、無限区間上に分布する 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 Gompertz 曲線は不自然である．後述の Aarssen and de Haan(1994)も、この点を議論している．

最尤推定

一般パレート分布 (2.1) からの大きさ n の確率標本の対数尤度は

$$(2.3) \quad l = -\frac{1 + \gamma}{\gamma} \sum_{j=1}^n \log(1 + \tau x_j) - n \log a, \quad \tau = \gamma/a,$$

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。 $-1/2 < \gamma$ ならば Fisher 情報行列

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

が有限で、その逆行列、つまり最尤推定量の漸近分散共分散行列は

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

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

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

となる(情報行列の計算は付録を参照)。 $\gamma = -1/2$ 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。 $-1/2 < \gamma < -1$ 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。 $\gamma \leq -1$ の場合に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表1にまとめておく。また、この節で述べたパラメータの意味と相互関係を付録でまとめておく。

3. 年齢時代区分データとそのモデル

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 $(i, j), i = 1, \dots, I; j = 1, \dots, J$ で表わすが、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、 $(i, j), i = 100, \dots, 120; j = 1970, \dots, 2000$ によってセルを表わす。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図(Lexis diagram, Keiding (1999))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(図3)。

表1. 一般パレート分布の特性.

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			

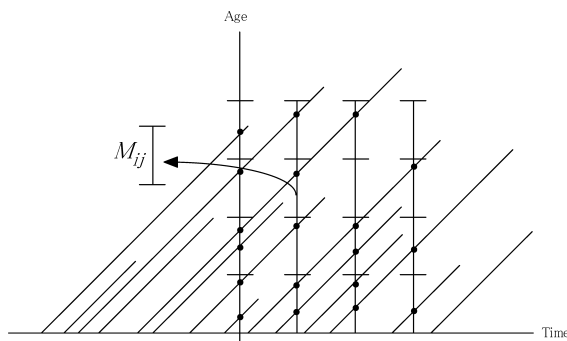


図 3. 生存統計のレキス図.

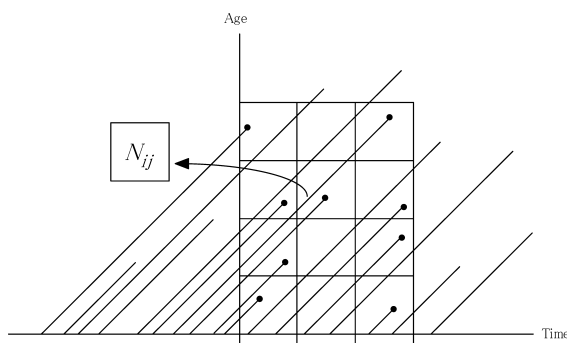


図 4. 死亡統計のレキス図.

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

$$(3.1) \quad M_{ij} - M_{i+1, j+1} =: D_{ij},$$

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 W_{ij} の人数 N_{ij} が記録されている．レキス図で見ると， 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(図 4)．

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．まえがきで述べた図 1, 図 2 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, 生存統計が判断に役立つ．

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

$$(3.2) \quad q(k, i) := F_k(i+1) - F_k(i) = \bar{F}_k(i) - \bar{F}_k(i+1), \quad k = j - i,$$

の多項分布を当てはめる． $F_k(t)$ は $\text{GPrt}(\gamma_k, a_k)$ の左打ち切り分布, $\bar{F}_k(t)$ 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, (2.2) 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

$$\begin{aligned} \frac{\partial l}{\partial \omega} &= \sum_{i=1}^k \frac{m_i}{q(i)} \frac{\partial}{\partial \omega} (\bar{F}(i) - \bar{F}(i+1)), & \frac{\partial l}{\partial \eta} &= \sum_{i=1}^k \frac{m_i}{q(i)} \frac{\partial}{\partial \eta} (\bar{F}(i) - \bar{F}(i+1)), \\ \frac{\partial}{\partial \omega} \bar{F}(t) &= \frac{\eta t}{\omega^2} \left(1 - \frac{t}{\omega}\right)^{\eta-1}, & \frac{\partial}{\partial \eta} \bar{F}(t) &= \bar{F}(t) \log \left(1 - \frac{t}{\omega}\right), \quad 0 < t < \omega, \end{aligned}$$

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．

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

$$k + v, \quad j + w,$$

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

$$(3.3) \quad y = j + w - (k + v) = (j - k) + (w - v) = j - k + u,$$

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．レキシス図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 Piecewise Constant Intensity Model (PCI モデル) を適用する (Keiding (1990)). 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

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

で定義する. 次にレキシス図上の上記断片を

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

により定義する. この断片は $k(r - s = 0, 3, \dots)$ 世代 $k + 1$ 世代を完全に含み, $k - 1, k + 2$ 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 世代の災厄関数(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

$$(3.4) \quad \lambda_{rs} = h(r + 1, s - r), \quad r, s = 0, 3, 6, \dots$$

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, person-years と呼ばれる各年度各歳の生存人口数 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わされる.

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

ただし M_{ij} は \mathcal{V}_{ij} で僅かでも生存した人の数を表わし, (3.1) と少し違う. これをパラメータに関して最大化する (Keiding (1975)).

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. (3.3) 式より, 事象 W_{ij} の出現数 N_{ij} の個数の三角分布乱数 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, $u < 0$ ならば $k - 1$ 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 (3.3) 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

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

これと $[y] = i$ を合わせると, y についての条件

$$(3.5) \quad \text{死亡年度} = j \Leftrightarrow i \leq y < i+1-v, \quad \rho_0(k, i) := \int_0^1 F_k(i+1-v)dv - F_k(i),$$

が得られる. $F_k(t)$ は (3.2) の下で定義した. $\rho_0(k, i)$ は, 条件式の確率である. 同様に

(3.6)

$$\text{死亡年度} = j+1 \Leftrightarrow i+1-v \leq y < i+1, \quad \rho_1(k, i) := F_k(i+1) - \int_0^1 F_k(i+1-v)dv,$$

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 (3.2) 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}, R_k は $k-1$ 世代 k 世代の閾値における生存比である.

(3.5), (3.6) の積分は陽に表わせる: 指数分布 ($\gamma = 0$) の場合,

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

ベータ分布 ($\gamma < 0$) の場合,

$$(3.7) \quad \int_0^1 F(t-v)dv = 1 - \frac{\omega}{1+\eta} \left(\left(\frac{\omega+1-t}{\omega} \right)^{1+\eta} - \left(\frac{\omega-t}{\omega} \right)^{1+\eta} \right) \\ = F(t) + O(\omega^{-2}), \quad \omega \rightarrow \infty.$$

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 $(\gamma_{k-1}, a_{k-1}), (\gamma_k, a_k), (\gamma_{k+1}, a_{k+1})$ を推定する. この 2 世代の i 歳での死亡確率を 4 分割し, 対応する観測数を用いた対数尤度

$$\sum_i (N_{i,k+i} \log(R_k \rho_0(k, i)) + N_{i,k+i+1} (\log(R_k \rho_1(k+1, i)) + \log(R_{k+1} \rho_0(k+1, i))) \\ + N_{i,k+i+2} \log(R_{k+2} \rho_1(k+2, i))),$$

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.

4. 実データの解析

この論文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。女性のほうが寿命が長いために観測数が多く、そのために死滅を確認できる世代が限定されるという矛盾がある。以下の解析は男性データに限る。

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)などである。以下の解析で泉さんは、同世代の記録が少ないために含まれていない。(本稿査読中に中願寺さん, 本郷さんが亡くなられた。)

4.1 生存者数統計の解析

厚生労働省(厚生省)は 1964 年以來毎年 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。表 2 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。全死亡者を死亡統計と比較できる世代は 1880–1890 年に限られる。

死亡統計は歴年単位の集計である。生存統計は 7–8 月頃の調査である。そのために (3.2) 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, 図 5 では 1880 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, 完全ではなかったとも判断できる。

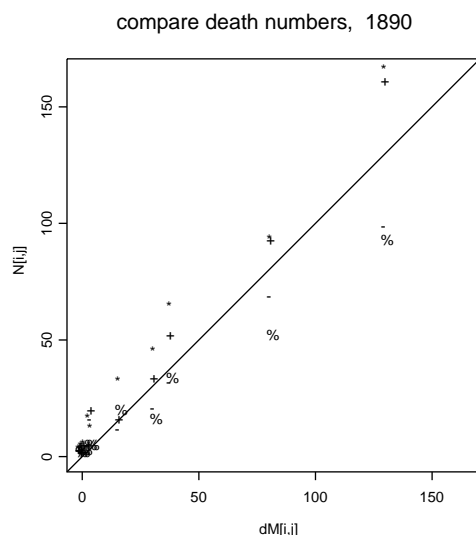


図 5. 生存統計と死亡統計の比較。+: $N(i,j)$, -: $N(i+1,j)$, *: $N(i,j+1)$, %: $N(i+1,j+1)$ 。

表 2. 年齢時代別 100 歳以上男性生存者数，全国高齢者名簿より。

生存者数 (男)	100 歳	101 歳	102 歳	103 歳	104 歳	105 歳	106 歳	107 歳	108 歳	109 歳	110 歳
1964 年	19	1	5	5	-	-	-	-	-	-	-
1965 年	24	7	1	1	2	-	-	-	-	-	-
1966 年	22	15	4	1	1	2	-	-	-	-	-
1967 年	26	10	12	2	1	-	-	-	-	-	-
1968 年	37	14	6	7	2	-	-	-	-	-	-
1969 年	26	25	8	6	3	2	-	-	-	-	-
1970 年	22	15	15	4	4	1	1	-	-	-	-
1971 年	31	11	10	10	4	2	1	1	-	-	-
1972 年	41	15	6	5	6	1	3	1	-	-	-
1973 年	56	27	8	4	2	5	-	-	1	-	-
1974 年	44	28	16	1	2	2	2	-	-	1	-
1975 年	43	23	18	11	1	2	1	2	-	-	1
1976 年	57	22	12	11	8	1	1	-	-	-	-
1977 年	56	31	13	6	8	5	1	1	-	-	-
1978 年	68	31	16	5	4	2	4	-	1	-	-
1979 年	101	45	21	4	2	3	-	3	-	-	-
1980 年	70	52	26	12	1	-	2	-	1	-	-
1981 年	98	47	32	16	4	1	-	2	-	1	-
1982 年	108	60	27	23	9	2	1	-	2	-	-
1983 年	132	68	33	18	14	2	1	-	-	-	-
1984 年	178	81	48	20	10	8	-	1	-	-	-
1985 年	157	102	53	27	8	6	4	-	1	-	-
1986 年	168	97	50	26	12	4	3	1	-	-	-
1987 年	225	108	65	32	20	6	2	3	1	-	-
1988 年	298	134	63	31	18	11	5	1	-	1	-
1989 年	302	171	80	40	13	11	7	4	1	-	1
1990 年	314	194	93	46	19	6	3	4	1	-	-
1991 年	373	183	100	47	27	14	4	-	1	-	-
1992 年	388	223	101	60	26	13	8	3	-	-	-
1993 年	430	243	143	62	35	15	8	6	1	-	-
1994 年	509	282	148	84	30	20	12	4	3	1	-
1995 年	640	302	158	78	47	13	8	4	2	2	1
1996 年	619	412	188	89	47	26	8	6	2	2	-
1997 年	722	385	265	96	52	25	18	4	2	-	1
1998 年	837	457	257	153	54	25	17	7	2	2	-
1999 年	885	495	292	157	86	30	12	9	4	1	2
2000 年	945	566	305	169	93	49	14	7	6	3	-
2001 年	1173	623	361	181	108	53	25	9	2	3	2
2002 年	1265	747	401	229	106	65	34	16	5	1	3

	111 歳	112 歳	113 歳	114 歳	115 歳	116 歳	117 歳	118 歳	119 歳	120 歳
1964 年	-	1	-	-	-	-	-	-	-	-
1965 年	-	-	1	-	-	-	-	-	-	-
1966 年	-	-	-	1	-	-	-	-	-	-
1967 年	-	-	-	-	1	-	-	-	-	-
1968 年	-	-	-	-	-	1	-	-	-	-
1969 年	-	-	-	-	-	-	1	-	-	-
1970 年	-	-	-	-	-	-	-	1	-	-
1971 年	-	-	-	-	-	-	-	-	1	-
1972 年	-	-	-	-	-	-	-	-	-	1
1973 年	-	-	-	-	-	-	-	-	-	-
1974 年	-	-	-	-	-	-	-	-	-	-
1975 年	-	-	-	-	-	-	-	-	-	-
1976 年	1	-	-	-	-	-	-	-	-	-
1977 年	-	1	-	-	-	-	-	-	-	-
1978 年	-	-	1	-	-	-	-	-	-	-
1979 年	-	-	-	1	-	-	-	-	-	-
1980 年	-	-	-	-	1	-	-	-	-	-
1981 年	-	-	-	-	-	1	-	-	-	-
1982 年	-	-	-	-	-	-	1	-	-	-
1983 年	-	-	-	-	-	-	-	1	-	-
1984 年	-	-	-	-	-	-	-	-	1	-
1985 年	-	-	-	-	-	-	-	-	-	1
1986 年	-	-	-	-	-	-	-	-	-	-
1987 年	-	-	-	-	-	-	-	-	-	-
1988 年	-	-	-	-	-	-	-	-	-	-
1989 年	-	-	-	-	-	-	-	-	-	-
1990 年	-	-	-	-	-	-	-	-	-	-
1991 年	-	-	-	-	-	-	-	-	-	-
1992 年	-	-	-	-	-	-	-	-	-	-
1993 年	-	-	-	-	-	-	-	-	-	-
1994 年	-	-	-	-	-	-	-	-	-	-
1995 年	-	-	-	-	-	-	-	-	-	-
1996 年	1	-	-	-	-	-	-	-	-	-
1997 年	-	-	-	-	-	-	-	-	-	-
1998 年	1	-	-	-	-	-	-	-	-	-
1999 年	-	-	-	-	-	-	-	-	-	-
2000 年	1	-	-	-	-	-	-	-	-	-
2001 年	-	1	-	-	-	-	-	-	-	-
2002 年	2	-	1	-	-	-	-	-	-	-

表 3. 生存者統計を用いたパラメータ推定結果.

	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

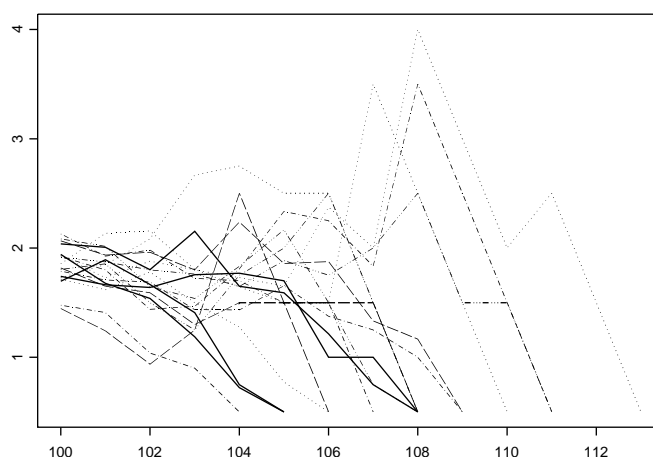


図 6. 生存統計の標本平均余命(年齢-年).

表 3 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, 生存者統計にモデル (3.2) 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 1885-1887 年の世代は, 死亡統計に 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, γ 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 1890-1892 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 3 つの世代である 1864-1866 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. 図 6 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.

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. 表 4 は 1980-2001 年の 84-117 歳の死亡者数である. 図 1 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.

表 4. 年齢時代別男性死亡者数 .

全死亡 (男)	84 歳	85 歳	86 歳	87 歳	88 歳	89 歳	90 歳	91 歳	92 歳	93 歳	94 歳	95 歳
1980 年	8208	7469	6206	5165	4241	3281	2903	2256	1570	1050	718	511
1981 年	8587	7638	6588	5330	4348	3555	2804	2235	1692	1145	836	516
1982 年	8999	7518	6618	5536	4586	3710	2827	2083	1743	1287	808	577
1983 年	9403	8603	7127	6239	5206	4132	3316	2550	1811	1363	1004	648
1984 年	9948	8614	7770	6342	5474	4436	3455	2593	1963	1414	1031	800
1985 年	10501	9259	7879	7066	5773	4772	3847	2899	2210	1623	1139	846
1986 年	10770	9684	8309	7095	6193	4954	4039	3181	2370	1755	1161	811
1987 年	10888	10034	8820	7594	6273	5487	4275	3428	2638	1821	1281	910
1988 年	11358	10958	9914	8666	7224	5981	5056	3765	2962	2137	1581	1078
1989 年	10983	10564	10113	8755	7398	6235	4859	3988	3046	2297	1744	1217
1990 年	11317	11041	10114	9413	8218	6837	5564	4344	3488	2555	1963	1384
1991 年	12777	10780	10131	9310	8474	7301	5757	4679	3558	2738	1987	1445
1992 年	13560	12156	10443	9443	8567	7617	6362	5122	3933	2828	2256	1543
1993 年	13648	12973	11565	9332	8757	7604	6583	5459	4315	3155	2382	1789
1994 年	13911	13114	12162	10581	8465	7426	6597	5638	4576	3447	2521	1804
1995 年	14839	13820	13000	11687	10117	7871	6992	6013	5125	4016	2815	2060
1996 年	14122	13450	12490	11372	9877	8447	6567	5692	4765	3848	2926	2144
1997 年	14586	13805	12808	11837	10594	9192	7771	5769	4788	4014	3191	2326
1998 年	14949	14118	13049	12079	10922	9600	8168	6689	4963	4142	3284	2557
1999 年	15694	15178	14026	13123	11803	10273	9139	7621	5860	4347	3511	2692
2000 年	15052	14266	14016	12670	11561	10217	8731	7695	6205	4848	3351	2556
2001 年	14574	14398	13811	13299	11807	10437	9229	8059	6471	5201	3908	2674

つづき	96 歳	97 歳	98 歳	99 歳	100 歳	101 歳	102 歳	103 歳	104 歳	105 歳	106 歳
1980 年	312	208	134	81	47	26	16	8	2	0	0
1981 年	339	257	130	93	43	27	15	9	1	0	0
1982 年	351	218	153	81	61	19	13	6	5	1	0
1983 年	425	235	159	123	64	28	18	13	5	7	1
1984 年	460	279	178	116	78	35	14	12	7	4	1
1985 年	577	360	222	122	70	56	31	17	8	4	5
1986 年	578	456	245	163	69	57	25	16	8	6	2
1987 年	616	429	293	164	88	44	30	11	6	2	0
1988 年	724	474	325	217	128	77	35	26	8	4	1
1989 年	844	561	319	222	135	89	33	19	11	8	4
1990 年	938	604	397	224	159	97	50	33	11	5	5
1991 年	977	705	431	264	164	91	67	30	15	3	2
1992 年	1053	782	488	290	181	91	50	30	18	7	2
1993 年	1174	850	551	339	186	114	62	32	19	9	3
1994 年	1292	828	557	408	207	135	94	43	14	10	3
1995 年	1434	1052	676	433	291	158	79	43	30	18	14
1996 年	1452	989	746	421	279	190	98	41	27	10	3
1997 年	1679	1114	776	487	303	178	131	56	24	18	14
1998 年	1814	1242	793	491	318	192	111	66	38	14	5
1999 年	2100	1431	904	626	375	257	134	94	43	20	14
2000 年	2011	1409	1025	641	410	243	159	79	42	29	6
2001 年	2067	1522	1062	721	481	265	169	84	52	27	19

つづき	107 歳	108 歳	109 歳	110 歳	111 歳	112 歳	113 歳	114 歳	115 歳	116 歳	117 歳
1980 年	-	1	-	-	-	-	-	-	-	-	-
1981 年	-	-	-	-	-	-	-	-	-	-	-
1982 年	-	2	1	-	-	-	-	-	-	-	-
1983 年	1	-	-	-	-	-	-	-	-	-	-
1984 年	-	-	-	-	-	-	-	-	-	-	-
1985 年	-	1	-	-	-	-	-	-	-	-	-
1986 年	-	-	-	-	-	-	-	-	-	-	-
1987 年	-	-	-	-	1	-	-	-	-	-	-
1988 年	2	1	-	-	-	-	-	-	-	-	-
1989 年	-	-	-	-	-	-	-	-	-	-	-
1990 年	4	1	1	1	-	-	-	-	-	-	-
1991 年	1	3	1	-	-	-	-	-	-	-	-
1992 年	1	-	-	-	-	-	-	-	-	-	-
1993 年	2	1	-	-	-	-	-	-	-	-	-
1994 年	3	2	-	-	-	-	-	-	-	-	-
1995 年	5	-	1	-	-	-	-	-	-	-	-
1996 年	3	1	1	-	-	-	-	-	-	-	-
1997 年	3	4	1	1	-	1	-	-	-	-	-
1998 年	3	1	-	-	-	-	-	-	-	-	-
1999 年	6	2	-	-	-	1	-	-	-	-	-
2000 年	1	3	1	1	1	-	-	-	-	-	-
2001 年	6	1	2	1	-	-	-	-	-	-	-

表 5. Piecewise Constant Intensity Model 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	-

表 6. 擬似確率標本による GPrt の推定 .

generation	size	thrsh	$\hat{\gamma}$	SE($\hat{\gamma}$)	$\hat{\omega}$	SE($\hat{\omega}$)	$\hat{\alpha}$	$\hat{\eta}$	p-val
1880	85	100.0	-0.2067	0.0860	112.6	13.1	0.171	4.84	0.008
1881	123	100.0	-0.1507	0.0766	117.9	23.0	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	-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	-0.1016	0.0765	121.3	37.0	0.092	9.84	0.092
1883	463	97.5	-0.0568	0.0438	135.5	67.7	0.054	17.60	0.097
1884	195	100.0	-0.0234	0.0699	178.9	510.3	0.023	42.67	0.369
1884	873	96.5	-0.1246	0.0296	117.0	12.1	0.111	8.03	0
1885	213	100.0	-0.1813	0.0561	112.9	9.5	0.153	5.52	0.001
1885	1228	95.5	-0.1223	0.0250	118.3	11.9	0.109	8.18	0
1886	188	100.0	-0.0205	0.0714	198.6	772.6	0.020	48.88	0.387
1886	1966	94.5	-0.0885	0.0206	124.5	17.5	0.081	11.30	0
1887	234	100.0	-0.1730	0.0541	113.0	9.6	0.147	5.78	0.001
1887	3020	93.5	-0.1202	0.0160	117.9	8.5	0.107	8.32	0
1888	306	100.0	-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	-0.0756	0.0479	129.0	42.6	0.070	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	-0.1649	0.0429	113.0	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: number of observations exceeding the threshold, thrsh. p-val: p-value in testing $\gamma = 0$.

In the generation 1880 dataset all exceed 100.

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。

表 5 に、レキシス図上の断片 \mathcal{V}_{rs} ; $r = 95, 98, \dots, 113$, $s = 1980, 1983, \dots, 1998$ に PCI モデル (3.4) 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、1885-1887 年度世代である。 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 125-127 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、その標準偏差が

ら $\gamma = 0$ の仮説は棄却された(表 5 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一般の認識と一致する.

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, 図 1 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.

表 6 に, Coles(2001) の関数 $\text{gpd.fit}()$ 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.

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 1889 年度世代は, 中願寺さんが生存中であるため対象外であることを, あえて存命を除外して解析した. 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. 終わりに

Aarssen and 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を解析した. コホートを作るため, 1881(, 80, ..., 77) 年生まれで 1971 年 1 月 1 日に 90(, 91, ..., 94) 歳に達していない人を除いた. また 1 年内の高齢者死亡率に季節性があり, 冬に高く夏に低いので, 一様分布となるように調整した. 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.

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. Aarssen and de Haan(1994) の結論は本稿と同じである. しかし著者たちは彼らの結論をさらに裏付けたと確信しているわけではない.

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. 全死者中の癌死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(competing risks)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ことになる. 医療技術の発達により癌死亡(の年齢分布)が変化するだろうし, このような意味で, 一般パレート分布(ベータ分布)が当てはまるという結論も一時期の判断となる可能性がある.

この論文で使用した元のデータおよび世代別に分け直したデータを, 要求があれば配布することができる.

付 録

一般パレート分布の情報量
次のモーメントを計算しておく。

$$\begin{aligned} E((1 + \tau X)^r) &= (\tau/\gamma) \int_0^\infty (1 + \tau x)^{r-1-1/\gamma} I[1 + \tau x \geq 0] dx \\ &= 1/(1 - \gamma r), \quad r < 1/\gamma, \quad \tau = \gamma/a. \end{aligned}$$

特に

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

対数尤度 (2.3) より, $n = 1$ のとき,

$$\begin{aligned} \frac{\partial l}{\partial \gamma} &= \frac{1}{\gamma^2} \log\left(1 + \frac{\gamma x}{a}\right) - \frac{1 + \gamma}{a\gamma} \frac{x}{1 + \gamma x/a}, \quad \frac{\partial l}{\partial a} = \frac{1 + \gamma}{a^2} \frac{x}{1 + \gamma x/a} - \frac{1}{a}, \\ \frac{\partial^2 l}{\partial \gamma^2} &= -\frac{2}{\gamma^3} \log\left(1 + \frac{\gamma x}{a}\right) + \frac{2}{a\gamma^2} \frac{x}{1 + \gamma x/a} + \frac{1 + \gamma}{a^2 \gamma} \left(\frac{x}{1 + \gamma x/a}\right)^2, \\ \frac{\partial^2 l}{\partial \gamma \partial a} &= \frac{1}{a^2} \frac{x}{1 + \gamma x/a} - \frac{1 + \gamma}{a^3} \left(\frac{x}{1 + \gamma x/a}\right)^2, \\ \frac{\partial^2 l}{\partial a^2} &= -\frac{2(1 + \gamma)}{a^3} \frac{x}{1 + \gamma x/a} + \frac{\gamma(1 + \gamma)}{a^4} \left(\frac{x}{1 + \gamma x/a}\right)^2 + \frac{1}{a^2}, \end{aligned}$$

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。

一般パレート分布のパラメータ

一般パレート分布 (2.1) のパラメータ (γ, a) は種々の別表現 (parametrization) があり, 第2節でも6つのパラメータについて述べている。ここでそれらの意味と相互関係をまとめておく。

まず γ は極値理論で裾の指数 (tail index), 極値指数 (extreme index) と呼ばれる重要なパラメータである。分布関数を表わすには η の方が自然であるが, 指数分布を $\gamma = 0$ の場合として含むことと, γ の大きい方が裾が重いというわかりやすさで γ が標準的となった。 a はこの表現での尺度パラメータである。 $\gamma < 0$ の場合に限ると分布範囲が有限区間となるので, 上限

表 7. 3 パラメータ間の関係。

$\overline{\gamma}$	γ	η	α
γ	-	$-1/\gamma$	$-\gamma/(1 - \gamma)$
η	$-1/\eta$	-	$1/(1 + \eta)$
α	$-\alpha/(1 - \alpha)$	$(1 - \alpha)/\alpha$	-

表 8. 4 パラメータ間の関係。

	$(\overline{\gamma}, a)$
τ	$\tau = \gamma/a$
ω	$\omega = -a/\gamma$

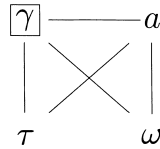


図 7. 4 パラメータ間の関係 .

ω と、確率変数を ω で割って得られるベータ分布のべき乗パラメータ η が分かりやすい . 平均超過関数(平均余命)の直線性が一般パレート分布の特徴なので, その勾配(の負数) α は応用上分かりやすいパラメータであり, 推定方法によっては役立つであろう . 最尤方程式の計算では, 和の中のパラメータを一つにする τ を導入すると計算しやすくなる .

表 7 は 3 つのパラメータ γ, η, α の間の関係で, その 1 つで他の 2 が定まる . これらの代表を γ とする . 表, 図では $\boxed{\gamma}$ によって, 代表であることを強調している . (γ, a) のうちの一方を表 8 にしたがって ω , または τ に変えることができる . 言い換えると, 4 つのパラメータ γ, a, ω, τ から 2 つを択ぶ 6 通りの組合せのうち, 図 7 の示すように (ω, τ) を除く 5 通りの組合せにより分布は定まる .

参 考 文 献

Aarssen, K. and de Haan, L. (1994) On the maximal life span of humans, *Mathematical Population Studies*, 4, 259–281.

Carey, R. J., Liedo, P. Orozco, D. and Vaupel, W. J. (1992) Slowing of mortality rates at older ages in large medfly cohorts, *Science*, 258, 457–461.

Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, London.

Kannisto, V. (1999) Trends in the mortality of the oldest-old, *Statistics, Registries and Science: Experiences from Finland* (ed. J. Alho), 177–194, Statistics Finland.

Kaufmann, E. (2001) About the longevity of humans, Section 14.2, *Statistical Analysis of Extreme Values* (written by R. D. Reiss and M. Thomas), 2nd ed., Birkhaueser, Basel.

Keiding, N. (1975) Maximum likelihood estimation in the birth-and-death process, *Annals of Statistics*, 3, 363–372.

Keiding, N. (1990) Statistical inference in the Lexis diagram, *Philosophical Transactions of the Royal Society of London, Series A*, 332, 487–509.

Keiding, N. (1999) Lexis diagram, *Encyclopedia of Statistical Sciences* (eds. S. Kotz, C. B. Read and D. L. Banks), Update Volume 3, 411–413, Wiley, New York.

Osmond, C. and Gardener, M. J. (1982) Age, period and cohort models applied to cancer mortality, *Statistics in Medicine*, 1, 245–259.

Smith, R. L. (1985) Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, 72, 67–90.

Smith, R. L. (1987) Estimating tails of probability distributions, *Annals of Statistics*, 15, 1174–1207.

Estimation of Human Longevity Distribution Based on Tabulated Statistics

Masaaki Sibuya

(Faculty of Business Management, Takachiho University)

Nobutane Hanayama

(Faculty of Art and Information, Shobi-gakuen University)

(Age, period)-specific data for the oldest-old survivors and deaths are analyzed using the extreme value theory and the limit of longevity distribution is discussed. The data were obtained from National Oldest-old Survivors List and Population Movement Statistics by the Ministry of Health and Labor in Japan. In applying the theory of extreme value statistics of continuous variables to the analysis of (age, period)-specific tabulated data, we propose a procedure for applying the multinomial distribution model based on probabilities calculated from the generalized Pareto distribution, and another procedure using a continuous pseudo random sample generated by adding random numbers. Furthermore, the piecewise constant intensity model on the Lexis diagram, widely used for such data, is also applied. ML estimates show the finite upper limit of longevity distribution.