What is SEM?

～死亡率予測に対する新アプローチ～

清水　泰隆 *
早稲田大学理工学術院

2022 年 1 月 21 日

概 要

コホート別の死亡率予測に対する全く新しいアプローチ "SEM" が Shimizu *et al.* [17] によって導入された．これは，人間の死が「生命エネルギー」の消滅によって起こると仮定し，生命エネルギー過程を確率過程でモデル化することで死亡時刻を予測するというものである．一見オカルト的とも思えるこの仮説は一種の方便であって，ある種の確率過程のクラスにおいて "ゼロ" への初期到達時刻の分布族を考えると，それが驚くほど良い予測分布を与えるという筋書きになっている．本稿では SEM 予測法の概略を紹介し，その利便性と今後の展望について議論する．最後に，この予測法に興味があり実務的な応用上の共同研究に参画していただける方は是非著者までご一報ください．

1　生命エネルギー仮説と構造的アプローチ

この数十年の間，人類の寿命は驚くほど伸びている．例えば，1950 年における日本女性の平均寿命は 60 歳程度であったが 2020 年には 87 歳に達しているし，先進国中最も伸び率の低い米国でも女性の平均寿命は 1950 年の 71 歳から 2020 年には 80 歳にまで伸びている．このような長寿傾向は「長寿革命（longevity revolution）」と言われるほど近年では顕著になっているようで（例えば [22]），このことは一方では良いことではあるが，他方では社会福祉や年金・経済問題，高齢者医療などさまざまな面で問題を抱えることにもなるため，長寿予測は世界的にも重大な課題になっている．

死亡率予測の問題は古くからアクチュアリーの興味の対象であり，Gompertz の法則，Makeham の法則，Heligman-Pollard の法則といった決定論的モデルが古典として残っている（例えば Olivieri [14] を見よ）．これらの考え方は「死」というものがある種のポアソン過程の最初のイベントであるとみなすことが基本になっている．すなわち，x 歳の人の余命を T_x（ポアソン過程の初期イベント時刻）とするとき，ある関数 $\mu(x,t)$ を用いて

$$\mathbb{P}(T_x > t+1 \,|\, T_x > t) = \exp\left(-\int_t^{t+1} \mu(x,s)\,\mathrm{d}s\right), \quad t > 0.$$

という関係を仮定するのである．特に，$\mu(t,x) \equiv \lambda > 0$(定数) であれば $\mathbb{P}(T_x > t+1 \,|\, T_x > t) = e^{-\lambda}$ であり，これは強度 λ のポアソン過程（T_x が平均 $1/\lambda$ の指数分布に従う）を考えていることになる．この関数 μ は保険数理の文脈で「死力 (force of mortality)」と言われるが，確率過程（特に点過程）の分野では「強度関数」，生存時間解析では「ハザード関数」などと言われるものであり，統計学でも馴染みの深い関数である．古典的には μ として非確率的なモデルが考えられてきたが，近年では $\{\mu(x,t)\}_{t\geq 0}$ を確率過程とみなすようなモデルが研究されている（例えば，Biffis [1], Cairns *et al.*

*E-mail: `shimizu@waseda.jp`

[5], Hainaut and Devolder [10], Biffis *et al.* [2], Blackburn and Sherris [3] およびそれらの参考文献を参照されたい）．ここで少しモデルを単純にして，$\mu(x,\cdot)$ が区間 $(t, t+1]$ の間は一定値 $m(x,t)$ をとるものとし，これに確率モデル（回帰モデル）を仮定するのが実務的な主流であり，これらがお馴染みの Lee-Carter モデル [8]，RH (Renshaw-Haberman) モデル [16], CBD (Cairns-Blake-Dowd) モデル [6, 7] などである．このようなモデルは，特に「死亡」という事象がなぜ起こるかには頓着せず，ざっくりとその確率だけに注目する方法論であり，この種のモデリングは信用リスク解析における「デフォルト」のモデリングにも見られる[1]．そこから用語を拝借して，これを「縮約的アプローチ (reduced-form approach)」と呼んでいる．

これに対して，Shimizu *et al.* [17] は「構造的アプローチ (structural approach)」を新たに提案した[2]．これは，人間が「生命エネルギー (survival energy)」を持って生まれてくると仮定し（**生命エネルギー仮説**），そのエネルギーの消滅によって「死亡」が起こるという考え方である．つまり，生命エネルギーの変動過程をモデリングできれば，その初期到達時刻によって死亡時刻を定義できる．このように，構造的アプローチでは「死」が起こる構造的原因に着目する．構造的モデルを考えるとき，世代によって初期エネルギーやエネルギー変動は異なると考えるのが自然であるから，モデルはコホート毎に与えるのがよい．例えば，コホート c に対する生命エネルギー過程を $X^c = (X_t^c)_{t\geq 0}$ と書けば，死亡時刻 τ^c は X^c の 0 への初期到達時刻 $\tau^c = \inf\{t > 0 \,|\, X_t^c < 0\}$ で定義され，その分布関数

$$q_c(t) := \mathbb{P}(\tau^c \leq t) \tag{1.1}$$

を **(コホート c に対する) 死亡率関数 (mortality function)** と呼ぶ．そこで，X^c に確率モデルを与えることで，実際に $q_c(t)$ の計算が可能になる．X^c に対する確率モデルを，以下では **SEM（Survival Energy Model，生命エネルギーモデル）** と呼ぶことにする．

Shimizu *et al.* [17] は，SEM としてある拡散過程 (diffusion process) のパラメータ族を仮定することにより，将来コホートに対する死亡率関数を精度よく予測できることを示している．このことは「生命エネルギー」なるものの存在を暗示している... と言うとオカルト的で面白いかもしれないが，これは一種の方便であり，主張したいことは単に死亡事象分布の良い候補モデル（母数分布族）を与えることが可能であるということであって，オカルト研究ではないのでご注意を．

本稿では，いくつかの SEM とその推定法，将来予測の手法について概説し，従来モデルと比べたときの SEM の実務的利便性と将来展望を述べる．細かい数値例などは本稿では触れないので Shimizu *et al.* [17] などを参照されたい．加えて，拙清水研究室のメンバーで遂行している死亡率関数の計算プロジェクト (**SEM Project**) についても紹介する[3]．

2 SEM の推定と予測に関する一般的手順

SEM アプローチでは $\vartheta \in \Theta\,(\subset \mathbb{R}^p)$ をパラメータとして持つ X^c の確率モデル $X^{c,\vartheta} = (X_t^{c,\vartheta})_{t\geq 0}$ を与え（具体的なモデルは次節），その死亡率関数のパラメータ族

$$q_c(t,\vartheta) = \mathbb{P}(\tau_\vartheta^c \leq t), \quad \tau_\vartheta^c := \inf\{t > 0 \,|\, X_t^{c,\vartheta} \leq 0\},$$

を考えていく．以下では，真の死亡率関数 q_c に対して，パラメータ ϑ の真値 (ϑ_c とする) が存在して

$$q_c(\cdot, \vartheta_c) \equiv q_c(\cdot), \quad \vartheta_c \in \Theta$$

を満たすとする．ただし，Θ は $\mathbb{R}^p$ の有界開集合とする．

[1]例えば，Jarrow and Turnbull [11] など．
[2]信用リスク解析における「構造的アプローチ」の代表例は Merton [12, 13] など．
[3]https://www.shimizu.sci.waseda.ac.jp/smzlab/semproject/

2.1 パラメータ推定

パラメータの真値 ϑ_c の推定には，以下の“条件付き”死亡率関数の理論形と経験形をフィットさせる最小二乗法（いわゆるカリブレーション）によって行う．

ある年齢 $S > 0$ に対して，

$$q_c(t|S) := \mathbb{P}(\tau^c \le t \,|\, \tau^c > S) = \frac{q_c(t)}{1 - q_c(S)}$$

とおく．この条件付確率 $q_c(\cdot|S)$ は，もし個別の死亡データがあれば以下のように経験推定できる：コホート c の母集団において，i 番目の人の死亡時刻 τ_i^c が得られれば，

$$\widehat{q}_c(t|S) := \frac{\sum_{i=1}^{n_c} \mathbf{1}_{\{S<\tau_i^c \le t\}}}{\sum_{i=1}^{n_c} \mathbf{1}_{\{\tau_i^c > S\}}}. \tag{2.1}$$

ただし，n_c はコホート c の母集団人数である．このような個別データは一般には入手不可能であるが，各国の死亡率データが掲載されている **Human Mortality Database (HMD)** [20] のデータセットを用いることで，以下のようにして (2.1)（の近似値）を得ることができる．

HMD によるデータ加工の手順：(`R` による自動変換コード [21] が利用可能)
(HMD で得られる) 西暦 c 年の生命表において x 歳の 1 年以内死亡率を $q_x^{(c)}$ とし，生まれ年 (コホート) が c 年である x 歳の人の 1 年以内死亡率を q_x^c とする．このとき，西暦 c 年の生命表において，以下が成り立つ：

$$q_0^{(c)} = q_0^c, \quad q_1^{(c)} = q_1^{c-1}, \quad \ldots, \quad q_\omega^{(c)} = q_\omega^{c-\omega},$$

ただし，ω は生命表における最終年齢である (HMD では $\omega = 110$).

(1) 上記から，$q_k^c = q_k^{(c+k)}, \quad k = 2, 3, \ldots, \omega.$

(2) 生存確率 $p_k^c := 1 - q_k^c$ for $k = 0, 1, \ldots, \omega$ を用いれば，

$$\mathbb{P}(\tau^c > t|\tau^c > S) = p_S^c \cdot p_{S+1}^c \cdots p_{t-1}^c, \quad t = S+1,\ S+2,\ \ldots,\ \omega.$$

(3) (2) によって，$\widehat{q}_c(t|S) = 1 - \prod_{k=S}^{t-1} p_k^c$（$t = S+1,\ S+2,\ \ldots,\ \omega$）

データ加工は面倒な作業ではあるが，HMD から csv 形式でダウンロードしたデータを，`R` によって一発変換で SEM の形式に変換し，`dataframe` 化できるコードを [21] に用意しているので，適宜利用されたい．

定義 2.1 (最小二乗推定量, LSE)**.** 年齢 $S > 0$ に対して $q_c(S) > 0$ とし，$t_d > \cdots > t_2 > t_1 > S$ なる d 個の年齢を任意にとるとき，ϑ_c に対する (条件付き) 最小二乗推定量 (LSE) を以下のように定める：

$$\widehat{\vartheta}_c := \arg\min_{\vartheta \in \overline{\Theta}} \sum_{i=1}^{d} |q_c(t_i, \vartheta|S) - \widehat{q}_c(t_i|S)|^2.$$

この推定量の一致性と漸近正規性については Shimizu *et al.* [17] に示されている．

2.2 将来コホートに対する死亡率関数予測

今，S 歳から ω 歳（HMD の最終年齢）までデータがしっかり揃っている m 個のコホート $c_1 < c_2 < \cdots < c_m$ に対して，各パラメータが $\widehat{\vartheta}_{c_1}, \widehat{\vartheta}_{c_2}, \ldots, \widehat{\vartheta}_{c_m}$ と推定されたとする．次は c_m 以降の将来コホートに対する $q_c(t, \vartheta_c|S)$ $(c > c_m)$ を予測したい．そこで，将来コホートのパラメータが次のような回帰式で決まると仮定してみる：

$$\vartheta_c = h(c) + \epsilon_c, \quad \epsilon_c \sim N_p(0, \sigma_\epsilon^2), \tag{2.2}$$

ただし，h はある (未知) 関数である．パラメータの推定値 $\widehat{\vartheta}_{c_1}, \widehat{\vartheta}_{c_2}, \ldots, \widehat{\vartheta}_{c_m}$ らを，確率変数 ϑ_{c_i} $(i = 1, \ldots, m)$ の実現値であるとみなし，h を適当な方法で推定することにする[4]．

関数 h が推定されれば,

$$\widehat{\vartheta}_{c'} = \widehat{h}(c'), \quad c' > c_m, \tag{2.3}$$

によって将来コホート c' に対するパラメータを予測し，$q_{c'}(\cdot, \widehat{\vartheta}_{c'})$ を予測死亡率関数とすればよい．このとき，$\vartheta_{c'}$ に対する予測区間ができて,

$$\widehat{I}_\alpha^{c',m} := \left[\widehat{\vartheta}_{c'} - z_{\alpha/2}\widehat{\sigma}_\epsilon,\ \widehat{\vartheta}_{c'} + z_{\alpha/2}\widehat{\sigma}_\epsilon\right], \tag{2.4}$$

となる．ただし，$\widehat{\sigma}_\epsilon$ は (2.2) の σ_ϵ の推定量であり，z_α は $N(0,1)$ の $(1-\alpha)$-パーセンタイルである．すなわち，以下が成り立つ：

$$\lim_{m\to\infty} \mathbb{P}\left(\vartheta_{c'} \in \widehat{I}_\alpha^{c',m}\right) = \alpha.$$

注意 2.2 (修正死亡率関数)．上記で予測死亡率関数 $q_c(t, \widehat{\vartheta}_{c'})$ ができたが，$\vartheta_{c'}$ に対する予測区間に基づいてこれを修正することができる．すなわち，$q_c(t, \vartheta|S)$ の ϑ を (2.4) の予測区間 $\widehat{I}_\alpha^{c',m}$ の中で動かして，コホート c' に対する既存のデータ $\widehat{q}_{c'}(t|S)$ $(t = t_1, \ldots, t_{d'})$ にカリブレートさせるように改めて最小二乗法を行うのである：

$$\widetilde{\vartheta}_{c'} = \arg \min_{\vartheta \in \widehat{I}_\alpha^{c',m}} \sum_{i=1}^{d'} |q_{c'}(t_i, \vartheta|S) - \widehat{q}_{c'}(t_i|S)^2, \tag{2.5}$$

こうしてできた $q_{c'}(\cdot, \widetilde{\vartheta}_{c'})$ は，最初の $q_{c'}(\cdot, \widehat{\vartheta}_{c'})$ よりも予測を改善することが多い．以下ではこれを**修正予測死亡率関数 (modefied predicted mortality function)** と呼ぶことにする．

3 SEM の具体例

3.1 非斉時拡散過程モデル：ID-SEM

Shimizu *et al.* [17] は以下のような非斉時拡散過程 (inhomogeneous-diffusion process) を用いて生命エネルギー過程をモデル化した．これを **ID-SEM** と呼ぶ：

$$X_t = x_c + \int_0^t U_c(s, \vartheta_c)\,\mathrm{d}s + \int_0^t V_c(t, \vartheta_c)\,\mathrm{d}W_s,$$

[4]実際には $\widehat{\vartheta}_{c_i}$ らに推定誤差が入るので，これはかなり雑な議論ではあるが，Lee-Carter モデルによる予測などでも標準的に行われているいわゆる二段階予測である．ただし，このような予測法は理論的には問題もある：Leng and Peng [9].

ただし，W はウィナー過程（標準ブラウン運動）で，x_c は初期エネルギーを表す正の定数である．ここで，U_c, V_c を

$$\frac{U_c(t)}{V_c^2(t)} = \frac{\kappa_c}{2} \in \mathbb{R}, \quad \inf_{t>0} V_c^2(t) > 0, \tag{3.1}$$

を満たすように与えると，死亡率関数が具体的に表現可能となるのだが，Shimizu *et al.* [17] らは以下のようなモデルを提案している:

$$U_c(t,\vartheta_c) = \alpha_c + \beta_c \exp\left(\gamma_c(t - T_c)\right) \mathbf{1}_{\{t \geq T_c\}}.$$

ただし，T_c は生命エネルギーのドリフト（トレンド）が急激に変化する変化点である．x_c や T_c などは，本来推定されるべきパラメータであるが，技術的な理由から [17] では適当な定数として与えており，$\vartheta_c = (\alpha_c, \beta_c, \gamma_c, \kappa_c)$ をパラメータとして扱っている．このとき，母数空間は

$$\Theta \subset \{(\alpha,\beta,\gamma,\kappa) \in \mathbb{R}^4 \mid \alpha < 0,\, \beta < 0,\, \gamma > 0,\, \kappa < 0\}$$

であり，死亡率関数は以下のように陽な表現を持つ．

定理 3.1. ID-SEM の死亡率関数は以下で与えられる：

$$q_c^{ID}(t,\vartheta_c) = 1 - \int_0^\infty f(z,t|\vartheta_c)\,\mathrm{d}z, \quad t \geq 0,$$

where $f(z,t|\vartheta_c) = G_{\vartheta_c}(z - x_c, t) - e^{-\kappa_c x_c} G_{\vartheta_c}(z + x_c, t)$;

$$G_{\vartheta_c}(y,t) := \frac{1}{2\sqrt{\pi S(t,\vartheta_c)}} \exp\left(-\frac{(y - M(t,\vartheta_c))^2}{4S(t,\vartheta_c)}\right);$$

$$M(t,\vartheta_c) = \int_0^t U(s,\vartheta_c)\,\mathrm{d}s; \quad S(t,\vartheta_c) = \frac{1}{2}\int_0^t V^2(s,\vartheta_c)\,\mathrm{d}s.$$

3.2 新しい SEM の提案: IG-SEM

本稿では，ID-SEM 以外の新たな SEM 候補も紹介する（これはまだ未発表）．その導入として次のような確率変数を考える：

定義 3.2. 確率変数 Y が，平均 a，分散 a^2/b の**逆ガウス分布** (inverse Gaussian distribution) に従うとは，その確率密度関数が

$$f_Y(y;a,b) = \sqrt{\frac{b}{2\pi y^3}} \exp\left(-\frac{b(y-a)^2}{2a^2 y}\right), \quad y > 0.$$

で与えられることであり，このとき $Y \sim IG(a,b)$ のように書く．

定義 3.3 (IG-SEM; Inverse Gaussian)**.** 以下の X^c で表される確率過程による生命エネルギーモデルを *IG-SEM* という．

$$X_t^c = x_c - Y_t^c, \quad t \geq 0, \tag{3.2}$$

ここに，$x_c > 0$ は初期エネルギーで，$Y^c \sim IG(\Lambda_c, \sigma_c)$ は平均関数 Λ_c とパラメータ $\sigma_c > 0$ を持つ**逆ガウス過程**，すなわち，$Y_0^c = 0\ a.s.$ で Y^c は独立増分過程，かつ，各 $t > s > 0$ に対して，Λ_c は単調増加で，

$$Y_t^c - Y_s^c \sim IG\left(\Lambda_c(t) - \Lambda_c(s), \sigma_c(\Lambda_c(t) - \Lambda_c(s))^2\right), \quad \Lambda_c(0) = 0.$$

注意 3.4. 上記，逆ガウス過程において，$\Lambda(t)=t$ とすると，Y はジャンプをもつレヴィ過程の一種であり，単調増加でジャンプが正となるようないわゆる従属過程 (subordinator) である．つまり，ID-SEM は連続なパスをもつモデルであるのに対し，IG-SEM は不連続なパスを許容するモデルとなっている．

このようなモデルは，工学系システムの故障時刻をモデリングするのに用いられている（Ye and Chen [19]）すなわち，ある種のダメージ過程 Y_t^c がシステムに蓄積し，それがある閾値 x_c を超えると故障するというモデルであり，その故障時刻は $\tau^c=\inf\{t>0\,|\,Y_t^c>x_c\}$ で与えられる．IG-SEM はこれを模倣したモデルであり，このとき，死亡率関数は以下で与えられる．

定理 3.5. IG-SEM による死亡率関数は以下で与えられる：

$$q_c^{IG}(t,\vartheta_c)=\Phi\left(\sqrt{\frac{\sigma_c}{x_c}}(\Lambda_{\vartheta_c}(t)-x_c)\right)-e^{2\sigma_c\Lambda_{\vartheta_c}(t)}\Phi\left(-\sqrt{\frac{\sigma_c}{x_c}}(\Lambda_{\vartheta_c}(t)+x_c)\right),$$

where $\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,\mathrm{d}z$.

我々が提案する平均関数 Λ_{ϑ_c} のモデルは,

$$\Lambda_{\vartheta_c}(t)=e^{a_ct}+b_ct-1,\quad \vartheta_c=(a_c,b_c,\sigma_c)\in\Theta,$$

であり，母数空間は以下である：

$$\Theta\subset\{(a,b,\sigma)\in\mathbb{R}^3\,|\,a>0,\ b>0,\ \sigma>0\}.$$

3.3 データ解析と SEM Project

本稿は SEM の取り扱い説明書なので，ここで詳細なデータ解析について議論することは避けるが，興味のある読者は論文 [17] のデータ解析を御覧いただきたい．スウェーデン，オランダ，日本，イギリス (ウェールズ) のデータを用いて分析を行っており，ID-SEM を用いることで数十年先のコホートに対する死亡率関数を良く予測することがわかるであろう．また，[17] に示されているように，これらの結果は従来の Lee-Carter モデルよりも良い結果を与える．

論文 [17] では ID-SEM のみを議論しているが，現在，筆者の研究室で IG-SEM を用いた分析も行っており，こちらのほうがより良い予測を与えることが多い．変化点 T_c なども考慮する必要がなく，パラメータの少なさの点でも使い勝手の良いモデルになっている．これらについては研究室メンバーで論文を執筆中である．

現在，拙研究室では，各国・各コホートの SEM を用いた予測死亡率関数を計算し，それを WEB 上に掲載するプロジェクト「**SEM Project**」を遂行中で，ここでは，HMD[20] 掲載のデータを用いて前述の手順によって求めた修正予測死亡率関数のグラフとパラメータ値をあわせて掲載している：

`https://www.shimizu.sci.waseda.ac.jp/smzlab/semproject/`

現在のところ，1937 年生～1980 年生の日本の男女別死亡率関数のグラフとパラメータが利用可能であるが，今後徐々に国やコホートを増やしていく予定である．読者個人で推定・予測などの作業を行う必要はないので，実務や研究などに気軽に利用されたい．また，共同で実証的な検証・研究をしていただける方も募集しているので，興味のある方はご一報いただければ幸いである．

4 SEM の利便性について

ここでは SEM を用いるいくつかの利点について，すぐに思いつく事を述べてみたい．

4.1　アクチュアリー記号の簡便化

アクチュアリー数学の初学者にとって，アクチュアリーが用いる独特の記号は一つの参入障壁といって良いかもしれない．例えば，契約年齢 x 歳の年 m 回年金期始払い n 年確定年金の一時払 (純) 保険料が $\ddot{a}^{(m)}_{x:\overline{n}|}$ のような記号でかかれたり，保険金即時払いの n 年定期保険の一時払い保険料は $\bar{A}^{1}_{x:\overline{n}|}$ などと，独特な記号で表現される．鈴木 [18] によると，このような一連のアクチュアリー記号は，1898 年に行われた第 2 回の国際アクチュアリー大会 (the 2nd International Congress of Actuaries, ICA1898) によって業界で合意されたもので，国際的には *Halo notation* と言われているようである．これらの記号群は，アクチュアリー数学に馴染みのない確率論や統計学の研究者にとっても理解が難しい記号であり，これらの研究への参入を少なからず阻んでいるかもしれない．

一方，SEM notation であれば，このような記号を全てコホート毎の生存時間に対する分布関数として表現することになるので，少なくとも数学者や統計学者にとっては理解が（多少）優しくなるであろう．例えば ${}_tp_x$ のような簡単な記号ですら，保険数理外の研究者には馴染みがないが，

$$
{}_tp_x := \frac{1-q_c(x+t)}{1-q_c(x)} = \mathbb{P}(\tau^c > x+t | \tau^c > x),
$$

のように書けば何を表すかは一目瞭然であり，"x 歳における (条件付き) 生存関数" と言えば済むであろうし，いわゆる「死力」と呼ばれる関数 μ_{x+t} については，

$$
\mu_{x+t} := -\frac{\mathrm{d}}{\mathrm{d}t}\log \mathbb{P}(\tau^c > x+t | \tau^c > x) = \frac{\partial_t q_c(x+t)}{1-q_c(x+t)}
$$

のように書いて，"x 歳における (条件付き) ハザード関数" という方が統計学者には通りがよいであろう．このように，Halo notation を死亡率関数や τ^c を用いて書き直すのは，コホートごとの計算をするのに役立つだけでなく，アクチュアリアルな量に対する直感的理解を確率論・統計学の研究者共通のものにするのに役立つと思われる．例えば，先の例であれば以下のように書ける：

$$
\ddot{a}^{(m)}_{x:\overline{n}|} = \frac{1}{m}\sum_{s=0}^{mn-1} v^{\frac{s}{m}}\frac{1-q_c\left(x+\frac{s}{m}\right)}{1-q_c(x)}; \qquad \bar{A}^{1}_{x:\overline{n}|} = \int_0^n v^t \frac{\partial_t q_c(x+t)}{1-q_c(x)}\,\mathrm{d}t,
$$

ただし，v はディスカウント・ファクター (例えば，金利 $r>0$ に対して $v=(1+r)^{-1}$) であり，$\partial_t = \partial/\partial t$ である．SEM によってコホート毎の将来死亡率が陽に与えられれば，このような保険料や年金原資のコホート毎計算が容易に行える．

4.2　死亡率の推定 (予測) 誤差について

SEM の最大の利点は，アクチュアリアルな推定や予測に対する簡便さではないかと思う．今，簡単な例として，x 歳の終身年金に対する一時払い保険料 A_x を考えよう．これを Hola notation と SEM notation の 2 通りに書いてみると，

$$
\begin{aligned}
A_x &:= \sum_{k=1}^{\infty} v^k\, {}_{k-1|}q_{x+k-1} \quad \text{(Halo notation)} \\
&= \sum_{k=1}^{\infty} v^k \frac{q_c(x+k)-q_c(x+k-1)}{1-q_c(x)} \quad \text{(SEM notation)}
\end{aligned}
$$

となる．これを例えば，従来の Lee-Carter モデルを用いて死亡率を予測するならば，

$$
A_x = \sum_{k=1}^{\infty} v^k\left[1-\exp\left(-m_{x+k-1,t}(\alpha_{x+k},\beta_{x+k})\right)\right],
$$

のように書かれる．ただし，

$$\log m_{x,t}(\alpha_x, \beta_x) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t},$$

であり，α_x, β_x は年齢 x にのみ依存するパラメータ，κ_t は時間依存するパラメータである．κ_t については，通常，離散時系列モデルを用いて予測されることが多いため，そのノイズ過程にも未知パラメータが混入するので，A_x の計算のためには，多くのパラメータ：$\{(\alpha_y, \beta_y)\}_{y=x,x+1,\ldots}$ に加えて κ_t のモデルの中の未知パラメータ，の推定（予測）が必要となり，誤差の増幅は避けられない．

しかしながら，SEM を用いれば，

$$A_x = \sum_{j=1}^{\infty} v^j \frac{q_c(x+j, \vartheta_c) - q_c(x+j-1, \vartheta_c)}{1 - q_c(x, \vartheta_c)}$$

であり，A_x は x 歳その人のコホート毎に計算されるだけでなく，パラメータの推定は（j によらない）ϑ_c を求めれば良いだけである．例えば IG-SEM を用いれば，推定すべきパラメータはたったの３つであり，その誤差評価は随分と容易である（次節を参照）．これだけでも SEM を使う利点は大きいであろう．

4.3 感度分析

パラメータの推定量や予測子はかならず誤差を含むものであり，それらを代入して計算される量もまた誤差を持つ．したがって，推定・予測のズレに対して，対象としている量がどれくらいズレるか，その程度を把握することは重要で，この文脈はいわゆる「感度分析」に相当する．前節で見たように，多くのアクチュアリアルな量は SEM の表現によって少数のパラメータによって予測可能であるが，このことはパラメータに対する感度分析の上でも好都合である．

保険料などの多くは q_c の汎関数として，例えば以下のようなスティルチェス型の積分形で書ける：

$$H(\vartheta) := \int_0^\infty Q_{c,x}(t|\vartheta)\,\mathrm{d}h(t), \quad \vartheta \in \Theta$$

ただし，h は $[0,\infty)$ 上の関数で

$$Q_{c,x}(t|\vartheta) := \frac{q_c(x+t, \vartheta)}{1 - q_c(x, \vartheta)},$$

であり，ここでの 積分記号の意味は，$\int_0^\infty := \int_{[0,\infty)}$ とする．以下，$\int_0^\infty$ とパラメータ微分 ∂_ϑ は必要なだけ交換可能であるとし，

$$\partial_\vartheta H(\vartheta) = \int_0^\infty \partial_\vartheta Q_{c,x}(t|\vartheta)\,\mathrm{d}h(t) < \infty, \quad \vartheta \in \Theta,$$

とする．例えば，x 歳加入の終身保険の一時払い保険料 A_x であれば，

$$h(t) = \sum_{k=1}^{\infty} v^k \left(\mathbf{1}_{\{t \geq k\}} - \mathbf{1}_{\{t \geq k-1\}}\right), \quad t \geq 0,$$

に取ればよいし，即時払いバージョンなら

$$\bar{A}_x = \int_0^\infty v^t \frac{\partial_t q_c(x+t)}{1 - q_c(x)}\,\mathrm{d}t$$

なので，

$$h(t) = -v^t - \mathbf{1}_{\{t \geq 0\}}.$$

と取ればよいことになる．実際，部分積分によって

$$
\begin{aligned}
H(\vartheta) &= \int_0^\infty \frac{q_c(x+t,\vartheta)}{1-q_c(x,\vartheta)}\,(-v^t)'\mathrm{d}t - \frac{q_c(x,\vartheta)}{1-q_c(x,\vartheta)} \\
&= \left[-v^t\frac{q_c(x+t,\vartheta)}{1-q_c(x,\vartheta)}\right]_{t=0}^\infty + \int_0^\infty v^t\frac{\partial_t q_c(x+t,\vartheta)}{1-q_c(x,\vartheta)}\,\mathrm{d}t - \frac{q_c(x,\vartheta)}{1-q_c(x,\vartheta)} \\
&= \int_0^\infty v^t\frac{\partial_t q_c(x+t,\vartheta)}{1-q_c(x,\vartheta)}\,\mathrm{d}t = \bar{A}_x.
\end{aligned}
$$

となるからである．

さて，感度分析ではパラメータの真値 ϑ_c に対して，差分 $H(\vartheta)-H(\vartheta_c)$ に興味がある．このとき，（いくつかの正則条件を気にしなければ）Taylor の公式によって

$$
H(\vartheta)-H(\vartheta_c) = \int_0^\infty \partial_\vartheta Q_{c,x}(t|\vartheta_c)\,\mathrm{d}h(t)\cdot(\vartheta-\vartheta_c) + o(\vartheta-\vartheta_c).
$$

と書けるが，積分 $\int_0^\infty \partial_\vartheta Q_{c,x}(t|\vartheta_c)\,\mathrm{d}h(t)$ については，SEM の下で具体的計算が可能である．特に，先述した ϑ_c の LSE $\widehat{\vartheta}_c$（定義 2.1，漸近正規推定量）を用いて評価すれば，

$$
\begin{aligned}
\sqrt{n_c}\left(H(\widehat{\vartheta}_c)-H(\vartheta_c)\right) &= \int_0^\infty \partial_\vartheta Q_{c,x}(t|\vartheta_c)\,\mathrm{d}h(t)\cdot\sqrt{n_c}(\widehat{\vartheta}_c-\vartheta_c) + o_p(1) \\
&\to^d N_p(0,\Sigma_{c,x}), \quad n_c\to\infty,
\end{aligned}
$$

となる．ただし，n_c は (2.1) で用いたコホート c の母集団人数であり，$\Sigma_{c,x}$ は $\widehat{\vartheta}_c$ の漸近分散と $\int_0^\infty \partial_\vartheta Q_{c,x}(t|\vartheta_c)\,\mathrm{d}h(t)$ によって計算可能な量である．これによって，推定量 $H(\widehat{\vartheta}_c)$ の漸近分布が特定され，誤差評価が可能となって以下の信頼区間が得られる：

$$
\mathbb{P}\left(H(\vartheta_c)\in\left[H(\widehat{\vartheta}_c)-z_{\alpha/2}\frac{\widehat{\Sigma}_{x,c}}{\sqrt{n_c}},\ H(\widehat{\vartheta}_c)+z_{\alpha/2}\frac{\widehat{\Sigma}_{x,c}}{\sqrt{n_c}}\right]\right)\approx 1-\alpha,
$$

ただし，z_α は標準正規分布の上側 100α-パーセンタイルであり，$\widehat{\Sigma}_{c,x}$ は漸近分散 $\Sigma_{c,x}$ の推定量である．

5 結びにかえて

以上見てきたように，SEM によってコホート毎に計算される死亡率関数は，さまざまな計算上・統計推測上の利点を有する.「生命エネルギー」というといかにもふざけているかと思われるかもしれないし，実は筆者も最初は遊びのつもりで趣味の一環であったのだが，実際にデータに当てはめてみると，思いの外良い予測精度をもち，Shimizu *et al.* [17] の解析にもあるように，Lee-Carter モデルよりも良い予測精度を達成することもわかった．今後は，他のコホート効果を持つモデルとの比較や，モデル選択などを通して，より良い SEM を探求する必要があると考えている．SEM Project では，日本以外の死亡率関数の計算も進めながら，より実務的な検証を進めたい.

今後，SEM を用いた実証研究等に関して共同研究をしてくださる方は，

清水泰隆（早稲田大学理工学術院 応用数理学科）`shimizu@waseada.jp` 宛

是非ともご一報をお願いいたします.

参考文献

[1] Biffis, E. (2005). Affine processes for dynamic mortality and actuarial valuation, *Insurance Math. Econom.*, **37**, 443–468.

[2] Biffis, E.; Denuit, M. and Devolder, P. (2010), Stochastic mortality under measure changes, *Scand. Actuar. J.*, (4), 284–311.

[3] Blackburn, C. and Sherris, M. (2013). Consistent dynamic affine mortality models for longevity risk applications, *Insurance Math. Econom.*, **53**, 64–73.

[4] Cairns, A. J. G.; Blake, D. and Dowd, K. (2006a). A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, **73**, 687–718.

[5] Cairns, A. J. G.; Blake, D. and Dowd, K. (2006b). Pricing death: frameworks for the valuation and securitization of mortality risk, *ASTIN Bulletin*, **36**, (1), 79–120.

[6] Cairns, A. J. G.; Blake, D. and Dowd, K. (2008). Modelling and management of mortality risk: a review. *Scandinavian Actuarial Journal*, (2-3), 79–113.

[7] Cairns, A. J. G.; Blake, D.; Dowd, K.; Coughlan, G. D.; Epstein, D.; Ong, A. and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England & Wales and the United States, *North American Actuarial Journal*, **13**, (1), 1–35.

[8] Lee, R.D. and Carter, L. (1992). Modelling and forecasting the time series of US mortality, *Journal of the American Statistical Association*, **87**, 659–671.

[9] Leng, X. and Peng, L. (2016). Inference pitfalls in Lee-Carter model for forecasting mortality, *Insurance: Mathematics and Economics*, **70**, 58–65.

[10] Hainaut, D. and Devolder, P. (2008). Mortality modeling with Lévy processes, *Insurance: Mathematics and Economics*, **42**, 409–418.

[11] Jarrow, R. A. and Turnbull, S. (1995). "pricing derivatives on financial securities subject to credit risk, *Journal of Finance*, **50**.

[12] Merton, R. C. (1974). On the pricing of corporate debt: the risk structure of interest rates, *Journal of Finance*. **29**, (2), 449 –470.

[13] Merton, R. (1976). Option pricing when underlying stock returns are discontinuous, *Journal of Financial Economics*, **3**, 125–144.

[14] Olovieri, A. (2001). Uncertainty in mortality projections: an actuarial perspective. *Insurance: Mathematics and Economics*, **29**, 231–245.

[15] Renshaw, A.E. and Haberman, S. (2003). Lee-Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, **33**, 255–272.

[16] Renshaw, A. and Haberman, S. (2006) A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, **38**, 556–570.

[17] Shimizu, Y.; Minami, Y. and Ito, R. (2020). Why does a human die? A structural approach to cohort-wise mortality prediction under Survival Energy Hypothesis, *ASTIN bulletin*, **51**, (1) 191–219.

[18] Suzuki, S. (2016). Origin of Halo Notation (in Japanese). The 27th Symposium of Mathematics History: `https://www2.tsuda.ac.jp/suukeiken/math/suugakushi/sympo27/27_suzukis.pdf`

[19] Ye, Z. and Chen, N. (2014). The inverse Gaussian process as a degradation model, *Technometrics*, **56**, (3), 302–311.

[20] Human Mortality Database: `https://www.mortality.org/`.

[21] `https://www.shimizu.sci.waseda.ac.jp/smzlab/files/To_semData.r`

[22] `https://www.swisslife.com/en/home/hub/longevity-revolution.html`