2020年10月27日　統計数理研究所　オープンハウス

テンソル分解を利用した都道府県別生命表解析

野村　俊一　　モデリング研究系　助教

概要

本研究では，機械学習におけるテンソル分解の考え方を用いて，Lee-Carterモデルに小地域の次元を加えて拡張したモデルを提案する．国立社会保障・人口問題研究所が公表している日本版死亡データベースの都道府県別生命表の死亡率データに提案モデルを適用し，都道府県による死亡率の特徴の違いを議論する．

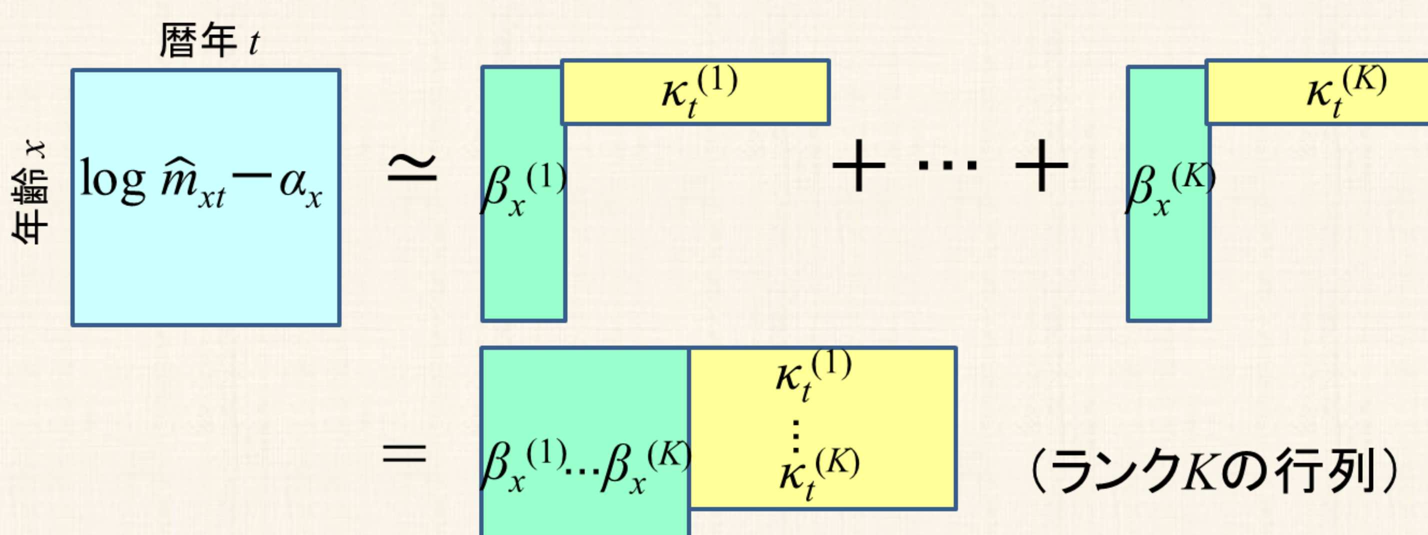
Lee-Carterモデル

死亡率推計のための生命表解析は，将来人口推計をはじめ，保険料算定，年金計算などを目的として発展してきた．特に，Lee and Carter (1992) を皮切りに，双線形型の回帰モデル（以下，Lee-Carter モデル）が広く用いてきた．Renshaw & Haberman (2003)による拡張Lee-Carter モデルは

$$\begin{cases} D_{xt} \sim \mathrm{Poisson}(E_{xt}\, m_{xt}) \\ \log m_{xt} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)} + \cdots + \beta_x^{(K)} \kappa_t^{(K)} \end{cases}$$

と表される．ここで，母集団の暦年t における満年齢 x 歳の生存延べ年数をE_{xt}，死亡数をD_{xt}，死亡率をm_{xt}としており，死亡率m_{xt} を切片α_x とK個の年齢効果×暦年効果のファクター$\beta_x^{(1)}\kappa_t^{(1)},\ldots,\beta_x^{(K)}\kappa_t^{(K)}$ の和によりモデル化している．

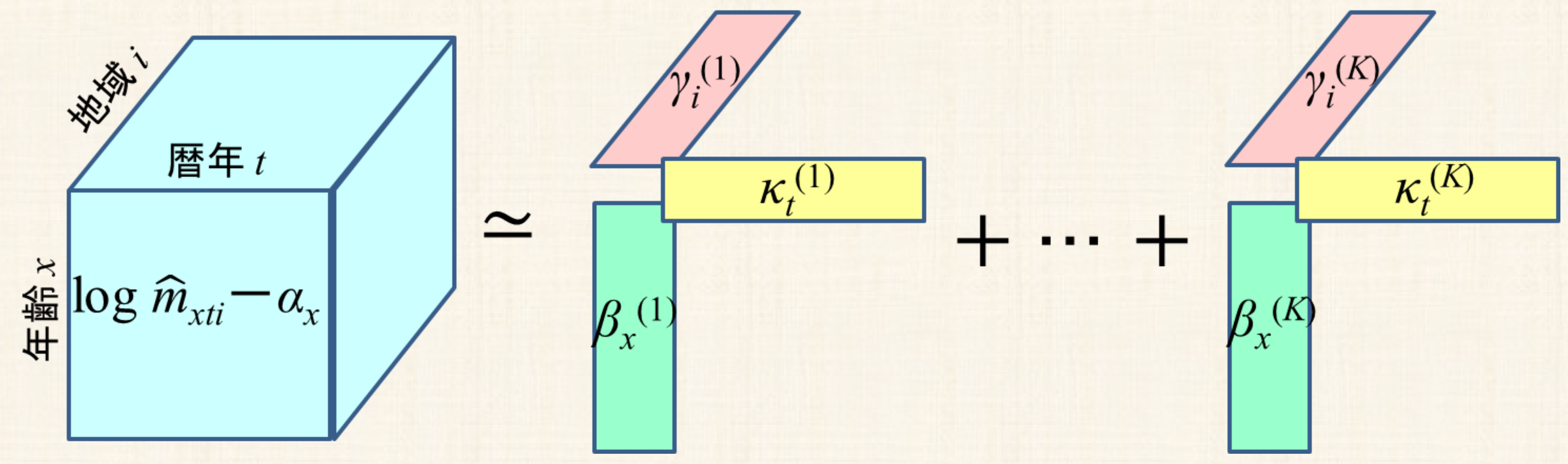
上のモデルは，対数死亡率から切片を引いた $\log m_{xt} - \alpha_x$ を，下図のように低ランク行列で近似したモデルと解釈することができる．実際，上のようにポアソン分布を仮定せず，最小二乗法により切片とファクターを推定する場合，切片は経験対数死亡率$\log \hat{m}_{xt} = \log D_{xt} / E_{xt}$ の年齢別平均で推定され，各ファクターは $\log \hat{m}_{xt} - \alpha_x$ の特異値分解における第1～K 特異ベクトルとして得られる．



提案手法

本研究では，地域別に細分化された生命表データに対するLee-Carter モデルの拡張を提案する．地域 i ごとの暦年t における満年齢 x 歳の生存延べ年数をE_{xti}，死亡数をD_{xti}，死亡率をm_{xti}として，地域効果のファクター$\gamma_i^{(1)},\ldots,\gamma_i^{(K)}$ を加えた次式によってモデル化する．

$$\begin{cases} D_{xti} \sim \mathrm{Poisson}(E_{xti}\, m_{xti}) \\ \log m_{xti} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)} \gamma_i^{(1)} + \cdots + \beta_x^{(K)} \kappa_t^{(K)} \gamma_i^{(K)} \end{cases}$$



上式は，$\log \hat{m}_{xti} - \alpha_x$ をテンソル分解（CP分解）したものと解釈され，切片および各ファクターは次式による更新を対数尤度が収束するまで繰り返すことで推定した．

$$\alpha_x = \alpha_x + \frac{\sum_{t,i}(D_{xti} - \widetilde{D_{xti}})}{\sum_{t,i} \widetilde{D_{xti}}}, \quad x = 0,\ldots,97$$

$$\beta_x^{(k)} = \beta_x^{(k)} + \frac{\sum_{t,i}(D_{xti} - \widetilde{D_{xti}})\kappa_t^{(k)}\gamma_i^{(k)}}{\sum_{t,i} \widetilde{D_{xti}}(\kappa_t^{(k)}\gamma_i^{(k)})^2}, \quad x = 0,\ldots,97,\ k = 1,\ldots,K$$

$$\kappa_t^{(k)} = \kappa_t^{(k)} + \frac{\sum_{x,i}(D_{xti} - \widetilde{D_{xti}})\beta_x^{(k)}\gamma_i^{(k)}}{\sum_{x,i} \widetilde{D_{xti}}(\beta_x^{(k)}\gamma_i^{(k)})^2}, \quad t = 1975,\ldots,2016,\ k = 1,\ldots,K$$

$$\gamma_i^{(k)} = \gamma_i^{(k)} + \frac{\sum_{x,t}(D_{xti} - \widetilde{D_{xti}})\beta_x^{(k)}\kappa_t^{(k)}}{\sum_{x,t} \widetilde{D_{xti}}(\beta_x^{(k)}\kappa_t^{(k)})^2}, \quad i = 1,\ldots,47,\ k = 1,\ldots,K$$

※ 各更新の前に期待死亡率$\widetilde{D_{xti}}$を次式で再計算しておく

$$\widetilde{D_{xti}} = E_{xti}\, m_{xti} = E_{xti} \exp\{\alpha_x + \beta_x^{(1)}\kappa_t^{(1)}\gamma_i^{(1)} + \cdots + \beta_x^{(K)}\kappa_t^{(K)}\gamma_i^{(K)}\}$$

解析結果

国立社会保障・人口問題研究所の公表する日本版死亡データベースの都道府県別生命表データから3つのファクターを推定した結果を右図に示した．

青線で示す第1ファクターは，全年齢にわたって死亡率が改善させてきたことを表すものであり，都道府県間の較差は小さい．

一方，第2・第3ファクターは，若年層と高齢層の死亡率較差が変化していることを表すものであり，都道府県間の較差が大きいため，都道府県間の死亡率の違いを説明するのに有効と考えられる．

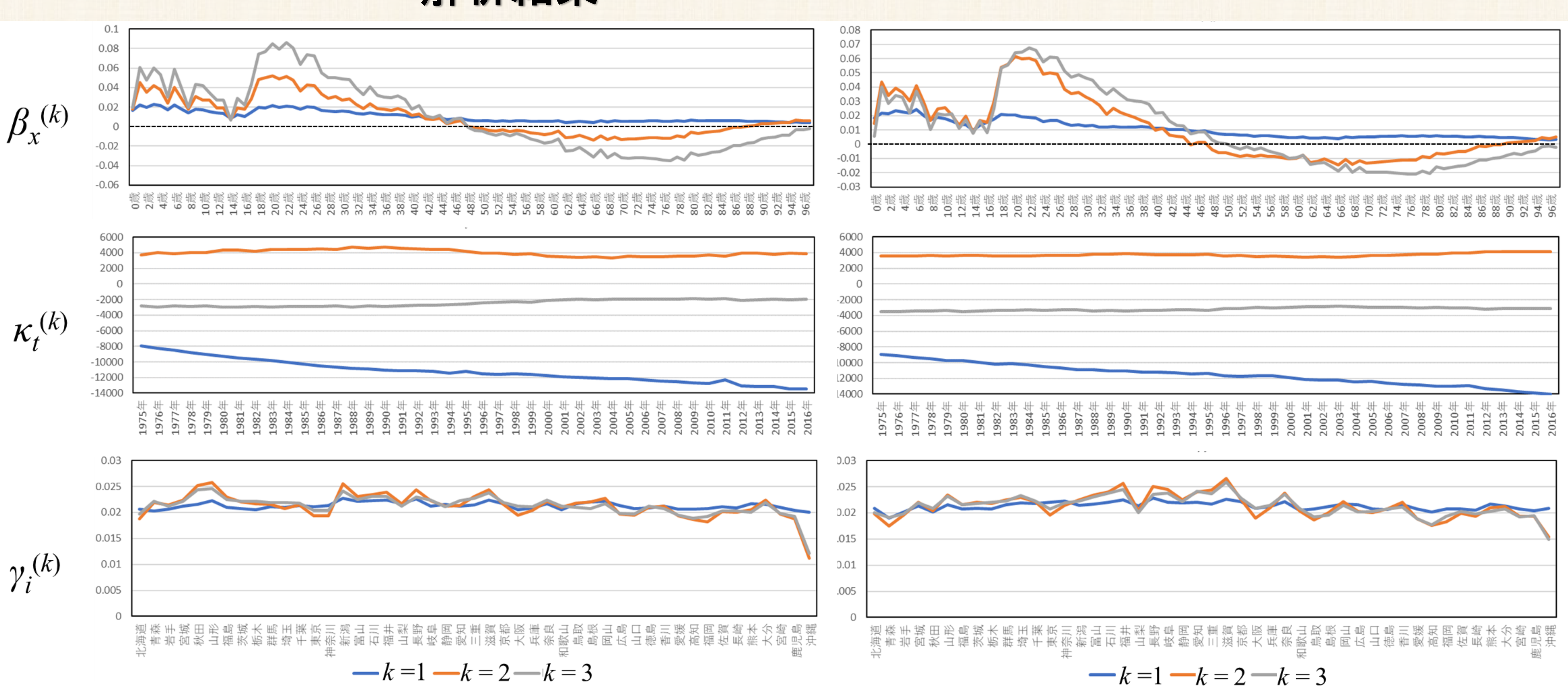


図. 女性（左）と男性（右）の都道府県別生命表から推定された3つのファクター

参考文献

野村俊一,テンソル分解を利用したLee-Carterモデルの拡張と都道府県別生命表解析, JARIP会報―大会プロシーディング特集号―, 2019年, pp.11-16.

大学共同利用機関法人 情報・システム研究機構
統計数理研究所
The Institute of Statistical Mathematics