

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*

2018年11月10日

概要

本稿では、機械学習におけるテンソル分解の考え方をを用いて、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。

キーワード： 都道府県別生命表, Lee-Carter モデル, テンソル分解, 日本版死亡データベース

1 はじめに

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。Lee-Carter モデルは、母集団の暦年 t における満年齢 x 歳の死亡率 m_{xt} に対して

$$\log m_{xt} = \alpha_x + \beta_x \kappa_t \quad (1)$$

を当てはめるものである。 α_x は時点によらない平均的な各年齢の対数死亡率、 κ_t は各時点における対数死亡率の全体的なトレンド、 β_x はトレンドに対する年齢ごとの対数死亡率の応答を表しており、これら全てが未知パラメータとして推定されることとなる。ここで、誤差が正規分布に従うものとしてモデル (1) を最小二乗法で推定する場合、 α_x は観察期間にわたる各年齢の対数死亡率の標本平均、 κ_t, β_x は $\log m_{xt} - \alpha_x$ を成分とするデータ行列の特異値分解における左第一特異ベクトルと右第一特異ベクトルがそれぞれ最小二乗推定量となる。一方で、実際の年齢別死亡数の観測モデルとしてポアソン分布などを仮定する場合には、最尤法に基づき、上記の最小二乗推定値を初期値にして対数尤度を最適化することにより推定値が求まる。

その後、Lee-Carter モデルはさらに拡張され、コホート項の導入 (Renshaw and Haberman, 2006) や双線形項を複数持たせたモデル (Brouhns et al., 2002) などが提案されてきた。特に、後者の双線形項を複数持たせたモデルは

$$\log m_{xt} = \alpha_x + \sum_{k=1}^K \beta_x^{(k)} \kappa_t^{(k)} \quad (2)$$

のように定式化され、 $\kappa_t^{(k)}, \beta_x^{(k)}$ ($k = 1, \dots, K$) に対しては第一特異ベクトルから第 K 特異ベクトルまでが最小二乗推定量となる。観測モデルにポアソン分布などを仮定した場合は前述のように最尤法で推定されるこ

* 統計数理研究所