

一般化 Fused Lasso によるクラス料率算定

野村 俊一*, 高橋 宏典†

2017 年 12 月 2 日

概要

本稿では、スパース回帰手法である Lasso の拡張版（一般化 Fused Lasso）を用いて、リスクファクター内の料率区分を同等のリスク水準をもつグループへと自動的にクラスタリングする新たな料率算定手法を提案する。タリフ分析では、信頼できる最良推計を得るために、多数のクラスをもつ料率ファクターを少数のカテゴリーへと再グループ化することが多い。しかしながら、料率区分の分割方法の組合せは膨大になることが多く、そのため可能な全ての分割方法を検討することは計算量的に困難である。このような状況下で、fused lasso と呼ばれる L1 正則化法は、推定の過程の中でリスク水準に有意差のない隣接区分を自動的に統合してくれるため非常に適している。特に、料率ファクター間の交互作用がある場合や、クレーム頻度と規模を同時にモデル化する場合の料率算定のための一般化線形モデルに対する幾つかの正則化法を提案し、日本の自動車損害賠償責任保険のクレームデータへと適用して各提案手法の比較を行う。

キーワード：タリフ分析、一般化線形モデル、一般化 fused lasso、交互方向乗数法、自動車損害賠償責任保険

1 はじめに

保険契約の契約条件や属性情報から保険料を決めるのにタリフ (tariff) と呼ばれる保険料率表が古くから利用されている。タリフでは料率ファクターと呼ばれる保険リスクに関わる属性情報について、各ファクター内で属する料率区分から該当契約の保険料が参照できるようになっている。タリフに定める保険料を過去のクレームデータから算定するための方法論は、現代の推測統計学とともに発展を遂げており、近年では Nelder and Wedderburn(1972) による一般化線形モデルが主流となっている。

タリフ形式の料率算定においては、被保険者年齢や居住地域、職業などの区分数の多いファクターが障害となる。料率区分数が多くなると区分ごとの過去のクレーム実績のデータ量は減少するため、信頼のおける料率の最良推計が得られない問題が発生する。また、現在でも冊子となっているタリフが利用されることは多く、区分数が多いとタリフ適用も煩雑となる。そのため、リスク水準に近い料率区分を統合して、より少数からなるカテゴリーへとグループ化することが従来から実務的に行われてきた。最適な料率区分のグルーピングを行うのに、妥当と考えられる全てのグルーピングの組合せを試すことは計算時間的に不可能であることが多く、従来より料率区分の統合のしかたは主に営業施策上の都合やアクチュアリーの実験に基づいて決められて

* 統計数理研究所 E-mail: nomura@ism.ac.jp

† 東京工業大学 情報理工学院 E-mail: takahashi.a.ar@m.titech.ac.jp