

一般化 fused lasso を用いたクラス料率の自動セグメンテーション

統計数理研究所 野村俊一

保険業界では古くから、保険契約ごとの保険料を決めるのにタリフ (tariff) と呼ばれる保険料率表が利用されてきた。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は、アクチュアリーと呼ばれる保険数理の専門家が、過去のクレームデータに基づいて算定しており、近年ではその方法論として一般化線形モデルが主流となっている。

一方で、タリフ分析において障害になる要素として、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の扱いが挙げられる。料率区分数が多くなる分だけ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という問題が生ずる。このような問題に対しては、リスク水準の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対処法として行われてきた。しかし、ファクターの区分数が増えるほど、そのような組合せの数は指数関数的に増大するため、現実には全ての組合せを検討することは計算量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営業施策上の都合やアクチュアリーの経験などに基づいて決められてきた。

本発表では、スパース回帰手法である一般化 fused lasso を利用して、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。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。ここでは特に、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グループ正則化項を導入する。L1 正則化項付き対数尤度からのモデルパラメータ推定には交互方向乗数法を利用し、また、正則化パラメータは交差確認法により選択する。

解析例として、提案手法を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、都道府県を同水準のクレーム頻度または規模をもつクラスターへとグループ化を行う。

参考文献

野村俊一 (2017). Group fused lasso による料率区分の自動セグメンテーション, 「ジャーナル・ジャーナル」, 第 9 巻 第 1 号, pp.1-20.