

単変量 Cox 回帰にもとづく遺伝子選択と複合共変量による生存期間の予測

国立中央大学（台湾）、統計研究所 江村剛志

名古屋大学大学院医学系研究科 松井茂之

Institute of Statistical Science, Academia Sinica Hsuan-Yu Chen

遺伝子発現量が、がん患者の生存期間の予測に有用であることは広く報告されている。遺伝子発現量と生存期間の関連を調べる研究において、単変量 Cox 回帰を使った遺伝子選択 (Feature selection) を用いる予後予測法が生物医学研究者の間で広く利用されてきた。興味深いことに、単変量 Cox 回帰の結果を線形結合した複合共変量 (Compound covariate) 法は、リッジ回帰法や Lasso 法などの罰則付き多変量 Cox 回帰法に匹敵する予測精度を持つことが報告されている[文献 1]。また複合共変量は、遺伝子選択を含む臨床試験のデザインに組み込むことや[文献 2,3]、患者レベルの個別化予測を実現すること[文献 4]が報告されており、個別医療の発展に貢献することが期待される。しかしながら、単変量 Cox 回帰分析による最適な遺伝子選択法、False Discovery Rate (FDR) の計算法、選択された遺伝子の予測能力の評価を行う方法などを実行する包括的なソフトウェアは開発されていない。本研究発表では、単変量 Cox 回帰分析にもとづく遺伝子選択法のいくつかの方法論を概説し、それら統計手法を実行する R パッケージ「*compound.Cox*」[文献 5]を紹介する。肺癌データ[文献 6]を用いて遺伝子選択法を説明し、FDR の計算やクロスバリデーションによる予測能力の評価を行う統計手法[文献 7]を紹介する。また選択された遺伝子を使用した予後予測の手法、とりわけコピュラ法[文献 8]や縮小推定法[文献 1]などの高度な手法も紹介する。本発表の詳細な解説はウェブページに掲載される「詳細論文」または原論文[文献 5]を参照されたい。

参考文献

- [1] Emura T, Chen YH, Chen HY. Survival prediction based on compound covariate under Cox proportional hazard models. *PLoS One* 2012; **7**(10): e47627. DOI:10.1371/journal.pone.0047627.
- [2] Matsui S, Simon RM, et al. Developing and validating continuous genomic signatures in randomized clinical trials for predictive medicine. *Clin Cancer Res* 2012; **18**(21): 6065-6073.
- [3] Matsui S. Statistical issues in clinical development and validation of genomic signatures. In Matsui S et al. (ed) *Design and Analysis of Clinical Trials for Predictive Medicine* 2015, CRC, pp.207-26.
- [4] Emura T, et al. Personalized dynamic prediction of death according to tumour progression and high-dimensional genetic factors: meta-analysis with a joint model, *Stat Method Med Res* 2017.
- [5] Emura T, Matsui S, Chen HY, compound.Cox: univariate feature selection and compound covariate for predicting survival, in revision *Computer Methods and Programs in Biomedicine*.
- [6] Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 2007; **356**: 11-20.
- [7] Matsui S. Predicting survival outcomes using subsets of significant genes in prognostic marker studies with microarrays. *BMC Bioinformatics* 2006; **7**:156
- [8] Emura T, Chen YH, Gene selection for survival data under dependent censoring, a copula-based approach, *Statist Method Med Res* 2016; **25**(6): 2840-2857.