

Small data における予測モデルの推定

予測モデルとは、既知のデータを用いて新しいデータや別のデータの結果の予測を行うための統計的モデルである。医療において患者個人や集団の疾病の発生、治療の予後などのイベントを予測することは、意思決定や患者のインフォームド・コンセントの取得などにおいて有用であり、予測モデルが用いられている。一般に予測モデルでは患者の臨床的特徴や人口統計的特徴などを用いて予測を行う。しかし、データが予測因子の数に対して小さすぎる場合、精確で信頼できる予測モデルを作成することは難しい。これはモデルの複雑性がデータの誤差的バラつきまでモデルで説明してしまうために起きる、モデルのデータに対するオーバーフィッティングとしてよく知られている。このオーバーフィッティングは一般にデータを大きくすることで回避できるが、医療における制約のために十分なデータを用意することが困難な場合も存在する。たとえば、興味のある疾患が稀であるときは患者を集めるのが難しく、また興味のあるイベントが稀であるときは患者を集めてもイベントが十分発生しない。

本抄読会では、予測因子に対してイベント数が小さいデータにおいても精度のよい予測をおこなうための方法として、縮小推定法と呼ばれる方法を Cox 回帰をもとに説明する。縮小推定法とは、予測因子の回帰係数を小さく推定することでモデルのデータに対するあてはまりを小さくし、オーバーフィッティングの影響を小さくする手法である。

文献

Ambler G, Seaman S, and Omar R.Z. An evaluation of penalised survival methods for developing prognostic models with rare events. *Statistics in Medicine*. 2012; 31: 1150-61.