

JALS 脂質データ解析に向けた具体的方策

1. はじめに

本抄読会では使用データに関する説明を行ったのち、スプライン関数を用いたスムージングと、プログラムによるその実行方法について発表する。

2. 使用データに関して

本研究では、TC (total cholesterol) 値、HDL-C (high-density lipoprotein cholesterol) 値、non-HDL-C (non-high-density lipoprotein cholesterol) 値、LDL-C (low-density lipoprotein cholesterol) 値の 4 つを用いて AMI (acute myocardial infarction) のリスク推定を行うことを目指す。解析方法としては、ノンパラメトリックなモデルを仮定し 2 つの説明変数をスムージングすることでリスク曲面を推定する方法を用いる。

今回使用するデータのうち、LDL-C 値は Friedewald の近似式 ($LDL-C = TC - HDL-C - TG/5$) によって算出される。この式は本来、空腹時に測定されたデータにのみ適用可能であるが、イベント数の関係からこの制約を緩めることを検討している。

使用するデータについてまとめた資料を発表当日に配布する。

3. スムージングについて

スムージング法の一つであるスプラインの基本的な考え方とは、 (x, y) からなるデータに関して、 y の期待値を基底関数と呼ばれる x の関数の和で表そうとすることである。この基底の数や種類、基底同士の接続点 (節点と呼ばれる) の選び方などの違いによりスプラインは回帰スプライン、平滑化スプライン、罰則付きスプラインといった様々な種類に分類される。

また基底の積や回転を考えることで、平面で考えたスプラインを 3 次元空間に拡張することも可能である。この回転の考え方に基づいて編み出されたスプラインの一つである薄板平滑化スプライン (thin plate regression splines) は、他のスプライン関数と比べ計算量も少ないことから、JALS データのような n の大きいデータにも適用しやすいと考えられる。

スプラインを扱うプログラムは SAS の GAM プロシージャや TPSPLINE プロシージャ、R の assist, gss, SemiPar, mgcv パッケージなど数多く存在している。今回扱うデータは n が大きく、またコントロールの人数に対してケースの人数が少ないため、プログラムによっては計算量が非常に多くなってしまいモデル当てはめがうまく実行出来ない。こういった例として SAS の GAM プロシージャがあり、コントロールをサンプリングすれば推定が可能であった。一方 mgcv パッケージではサンプリングすることなく全データを用いて推定

をすることが出来たが、今回はデフォルトの設定でのみ解析を行ったので、今後細かい指定の方法等を勉強し、再度推定を行いたい。

4. 参考文献

- [1] Tanabe N, Iso H, Okada K, Nakamura Y, Harada A, Ohashi Y, et al. Serum total and non-high-density lipoprotein cholesterol and the risk prediction of cardiovascular events - the JALS-ECC -. *Circ J* 2010; 74: 1346-56.
- [2] NIPPON DATA80 Research Group. Risk Assesment Chart for Death from Cardiovascular Disease Based on a 19-Year Follow-up Study of a Japanese Representative Population - NIPPON DATA80 -. *Circ J* 2006; 70:1249-55.
- [3] Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem* 1972; 18: 499-502.
- [4] Nordestgaard BG, Benn M. Fasting and Nonfasting LDL Cholesterol: To Measure or Calculate? *Clin Chem* 2009; 55:845-7.
- [5] Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge University Press, New York, 2003.
- [6] Eilers PHC, Marx BD. Flexible Smoothing with *B*-Splines and Penalties. *Stat Sci* 1996; 11: 89-121.
- [7] Hastie TJ, Tibshirani RJ. *Generalized Additive Models*. Chapman and Hall/CRC, Boca Raton, 1990.
- [8] Wood SN. Thin-plate regression splines. *J. R. Statist. Soc. B* 2003; 65: 95-114.
- [9] Wand MP, Coull BA, French JL, Ganguli B, Kammann EE, Staudenmayer J, Zanobetti A. SemiPar 1.0. R package. <http://cran.r-project.org> 2005.