
一般口演B

[OB-11] 一般口演 B

欠損値と過分散を伴う経時データに対する状態空間モデルを用いた
解析-生存時間解析への影響

2019年6月8日(土) 13:00 ~ 13:30 第1会場 (熊本市民会館 1F 大ホール)

[OB-11] 欠損値と過分散を伴う経時データに対する状態空間モデルを用いた
解析-生存時間解析への影響

永田 桂太郎 (高知大学)

大会欠損値と過分散を伴う経時データに対する 状態空間モデルを用いた解析-生存時間解析への影響

永田 桂太郎^{*1}, 兵頭勇己^{*1}, 畠山豊^{*1}, 渡部輝明^{*1}, 奥原義保^{*1}

堀野太郎^{*2}, 松本竜季^{*2}, 寺田典生^{*2}

^{*1} 高知大学医学部附属医学情報センター, ^{*2} 高知大学医学部内分泌代謝・腎臓内科

State-space model for missing data and overdispersion in time-course data - effect on survival analysis

K. Nagata^{*1}, Y. Hyodo^{*1}, Y. Hatakeyama^{*1}, T. Watabe^{*1}, Y. Okuhara^{*1},
T. Horino^{*2}, T. Matsumoto^{*2}, Y. Terada^{*2}

^{*1} Center of Medical Information Science, Kochi Medical School, Kochi University,

^{*2} Department of Endocrinology, Metabolism and Nephrology, Kochi Medical School,
Kochi University

抄録：検査の経時データを研究利用する際には欠損値や過分散の課題が伴う。状態空間モデルによる欠損値推定および過分散の除去を行い、生存時間解析への影響を調べる。例として、糸球体濾過量(GFR)の経時データを考察する。状態モデルとしてGFRの時間に関する一階微分方程式を仮定し、また、観測値はモデル推定値を平均とする正規分布に従うと仮定する。測定間隔のばらつきを補正するため、360日ごとのGFRをモデルを用いて推定する。GFRの低下をエンドポイントとして、標準の方法、検査値の線形補間、および状態空間モデルの3つの方法で生存時間を求め、Kaplan-Meier曲線を比較する。線形補間は左打ち切りの補正効果のみを持つものに対して、状態空間モデルでは、これに加え、過分散の除去の効果が得られることがわかった。

キーワード 欠損値、過分散、経時データ、状態空間モデル、生存時間解析

1. はじめに

検査値の経時変化は患者状態の推移を知るための有用な情報である。医学研究においては、生存時間解析などに应用され、また、検査値の時間変化のモデル化は発症予測に向けた重要なテーマとして研究されている。

診療データの研究利用には測定間隔のばらつきが起す欠損値の問題や、観測されない要因がもたらす過分散の問題がつきまとう[1]。これらは解析にも影響を与えることがあり、例えば、欠損値は生存時間解析に左打ち切りをもたらし、過分散はデータに適合するモデル構築を困難にする。

本研究では、状態空間モデルを用いて欠損値推定および過分散除去を行い、その生存時間解析への影響を調べる。

2. 方法

データ

応用例として[2]で用いたデータ、高知大学医学部附属病院において急性腎障害(AKI)を発症し、回復した患者のeGFRのデータを考察する。AKIはKDIGOガイドラインに従って定義した。回復はeGFRがAKI前の値まで上昇する事と定義した。

モデル

AKIからの回復日をゼロ日、t日後のeGFRの検査値を y_t とする。 y_t は真の腎機能 α_t に誤差 s_y を加えたものと仮定する、すなわち

$$y_t \sim N(\alpha_t, s_y).$$

α_t は1日あたり ϵ 変化すると仮定する。恒常的な変化以外の、生活や疾病等の影響は正規分布に従うと仮定する、すなわち

$$\alpha_t \sim N(\alpha_t + \epsilon \Delta t, s_a)$$

ここで、 $\Delta t = t' - t$ は測定間の経過日数を表す。

患者ごとの測定間隔の違いを補正するため、360日ごとに α_t を推定する。ただし、測定期間が360日未満の患者に対しては、測定期間を4等分した各時点における腎機能を推定する。測定が2度の場合は線形補完を行なった。

アウトカム

イベント発生までの時間(生存時間)と ε を求める。イベントはeGFRが初回値の70%を下回ることと定義した。次の3つの方法を用いて生存時間を求め、Kaplan-Meier(KM)曲線を比較した: (1) 標準の方法(Naive) (検査値が条件を満たした日), (2) 線形補間(Linear Interpolation) (隣接する検査値の線形補間直線がイベント条件式と交差する日), および (3) 状態空間モデルで推定されるイベント日(State Space model)。解析にはRstanを用いた。

3. 結果

条件を満たすn数は2378名となった。図1に、ある患者のeGFRの測定値と推定値を示す。欠損値の推定日を"+"(赤)で表した。図2はKM曲線を表す。線形補間を用いた場合のKM曲線(1)は標準の結果(2)に比べて常に小さくなった。一方、状態空間モデルの結果(3)は通常の結果(1)と交差しており、ほぼ違いがなかった。

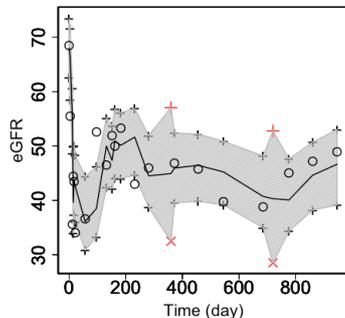


Fig.1 ある患者の eGFR(丸)、状態空間モデルの推定値(平均値: 実線、95%信頼区間: 灰色。ただし、ここでの信頼区間はベイズ信頼区間。)

4. 考察

図1に示したように、状態空間モデルでの推定結果は検査値を信頼区間内に捉えつつデータを平滑化したものとなる。図2で線形補間が標準の方法をどの時刻でも下回る結果は、左打ち切りが補正されたことによる一般的な結果である。この差は系統的なものであり、標本数が増加すれば2つの

KM 曲線に有意差を生じうる。

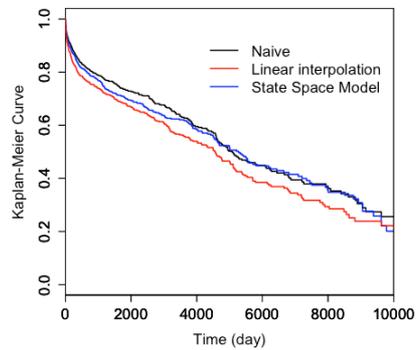


Fig.2 KM 曲線. Naive が方法(1), Linear interpolation が(2), State space model が(3)。

一方、状態空間モデルの結果はかならずしも標準の方法を下回らない。これは、状態空間モデルが左打ち切りの補正だけでなく、過分散を除去し、検査値を平滑化する効果を持つためである。

注意点として、検査値のゆらぎは誤差や生活の影響など非疾患的要因と疾患要因によるものがある。本研究は腎機能低下速度が一定と仮定したため、後者の可能性を考慮できていない。これが本研究の限界である。後者の影響を考えるためには、電子カルテの記載や病歴情報など検査値以外の情報を考慮したモデル改良が必要となる。

5. 結語

経時データの欠損値推定や過分散の除去に対して状態空間モデルが有用である可能性が示唆された。本研究に関して著者らの開示すべき利益相反はない。本研究は高知大学医学部の倫理審査で承認されている(IRB申請番号23-57)。

参考文献

- [1] N. Poh, S. Tirunagari, N. Cole, S. de Lusignan, J. Bio. Info. 76(2017) 69-77.
- [2] 永田桂太郎、畠山豊、堀野太郎、他: 腎性AKIと腎前性AKIの長期腎予後に対する生存時間解析, 第37回医療情報学連合大会論文集, 745-748, 2017.
- [3] 永田桂太郎、畠山豊、堀野太郎、他: 第22回日本医療情報学会春季学術大会抄録集.