

Profile Likelihood for Cure-Death Hazard Ratio on Competing Risks Model and Its Application to Estimation of the Case Fatality Rate of SARS

競合リスクモデルに基づく治癒対死亡ハザード比のプロフィール (profile) 尤度法とその SARS 死亡率の推定の応用

生産科学研究科 陳 征

諸言

SARS (severe acute respiratory syndrome、重症急性呼吸器症候群) は 21 世紀初頭に社会に大きな影響を与えた新興感染症である。SARS には様々な研究すべき問題があるがその一つに死亡率の推定方法がある。死亡率 (Case fatality rate) とは、「疫病に罹った人々の中でその疫病が原因で死亡した人々の割合」と定義される。従って SARS の終息後、精確な死亡率が計算可能となる。疫病発生中における死亡率の推定は困難であるが疫学的に重要な情報なので、様々な推定法が提案されている。2003 年 5 月 7 日に WHO は 3 つの推定法を公表した：①累積死亡数/累積患者数、②累積死亡数/(累積死亡数+累積退院数)と③生存時間解析法に基づく方法である。同じ日に Donnelly et al. (Lancet, 2003) はパラメトリック法とノンパラメトリック法の 2 種類の推定法を提案した。Ghani et al. (2005) と Jewell et al. (2006) は Kaplan-Meier 法に基づく推定値を与えた。最後の 2 つは個人データを用いている。一方、Yip et al. (2005) は毎日、その日までのサマリーデータをもとに、その日の死亡率を求めた。WHO と Lancet によるノンパラメトリックな死亡率推定法はバイアスが大きく、またパラメトリック法は検証困難な仮定を用いている。余計な仮定を必要とせずバイアスの小さい死亡率の推定法は未だ発見されずにいる。本研究は、治癒例を考慮した競合生存時間解析モデルを用いた、プロフィール尤度 (profile likelihood) 法による死亡率の推定法を提案する。これはサマリーデータにも個人データにも適用可能なである。シミュレーションと数学的考察により既に提案された上記推定法と精度を比較した。また北京、香港及びシンガポールのデータに応用し有効性を検証した。

データおよび推定方法

サマリーデータの場合：入院者数(n)、死亡者数(d)と治癒者数(c)、観察対象者数 (a) から 2×2 表を作り、2 つのエンドポイントである死亡と治癒、それぞれの 1 日当りハザードを毎日、 t_i , $i=1,2,\dots$ 、求めた。もし t_i における治癒ハザード(h_{2i})と死亡ハザード(h_{1i})の比が定数 θ ならば最終死亡率は $\pi_1 = 1/(1+\theta)$ と推定される。

死亡ハザード(h_{1i})は日毎に変わりうる定数、死亡対治癒ハザード比 (cure-death hazard ratio) は一定を仮定し、martingale 理論と競合リスク理論に基づくプロフィール対数尤度：

$$pl(\theta, h_{1i}) = \sum_i \{d_i \log h_{1i} + c_i \log(\theta h_{1i}) + (a_i + d_i + c_i) \log(1 - h_{1i} - \theta h_{1i})\}$$

を導出した。

個人データの場合：共変量 \mathbf{Z} を持つ個人についての時間 t_i におけるタイプ別ハザードを $\lambda_j(t_i, \mathbf{Z}) = \lambda_{j0}(t_i) \exp[\boldsymbol{\beta}_j^T \mathbf{Z}]$ ($j=1$: 死亡 ; $j=2$: 治癒) と書くことにする。二つハザードの比が時間に依存しない定数 $\theta(\mathbf{Z}) = \theta_0 \exp[(\boldsymbol{\beta}_2^T - \boldsymbol{\beta}_1^T) \mathbf{Z}]$ ならば、死亡率は $\pi_1(\mathbf{Z}) = 1/(1 + \theta(\mathbf{Z}))$ である。ここに θ_0 はベースラインハザード比と呼ばれる。同日の死亡や治癒 (tie) があるときでも正しい、全尤度に基づく θ_0 、 $\boldsymbol{\beta}_1$ と $\boldsymbol{\beta}_2$ のプロフィール対数尤度

$$pl(\theta_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \propto \sum_i \left(-(\Delta_{1i} + \Delta_{2i}) \log \left\{ \sum_{k \in R_i} (\exp[\boldsymbol{\beta}_1^T \mathbf{Z}_k] + \theta_0 \exp[\boldsymbol{\beta}_2^T \mathbf{Z}_k]) \right\} \right. \\ \left. \sum_{k \in R_{1i}} \boldsymbol{\beta}_1^T \mathbf{Z}_k + \Delta_{2i} \log \theta_0 + \sum_{k \in R_{2i}} \boldsymbol{\beta}_2^T \mathbf{Z}_k \right)$$

を導出した。ここに Δ_{1i} と Δ_{2i} は t_i における死亡者数と治癒者数、 R_1 、 R_{1i} 、 R_{2i} は t_i における観察対象者、死亡者、治癒者の集合である。

$\mathbf{Z}=0$ または $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ のときは、上記プロフィール対数尤度は簡単に次式で書ける。

$$\overline{pl}(\theta_0) \propto \sum_{i=1}^s \left\{ -(\Delta_{1i} + \Delta_{2i}) \log(1 + \theta_0) + \Delta_{2i} \log \theta_0 \right\}$$

シミュレーションによる比較と実際 SARS データへの応用

シミュレーション結果によると、プロフィール尤度法をサマリーデータ並びに共変量を持つ個人データに適用したときの結果は標本サイズが 200 と小さいときでも死亡率の推定値とその分散の推定値は近似的に不偏であることが確認された。共変量を持たない個人データについては、3つの方法 (プロフィール尤度法、Kaplan-Meier 法、パラメトリック治癒モデル法) が適用可能なので、それらをシミュレーションにより比較した結果、Kaplan-Meier 法は分散の推定値が偏っており、パラメトリック治癒モデル法はモデルが正しくかつピーク時を過ぎた時期でのみほぼ正しい推定値を提供する。一方、プロフィール尤度法の推定値は治癒死亡ハザード比が一定のときは常に高い精度の推定値を提供することが確認された。

SARS 終息後の香港、シンガポール、北京の SARS 死亡率 (%) の真値は 17.10, 15.53, 7.62 であった。SARS 発生後の早期まで遡って本論文の方法をそれぞれのデータへ適用したところ、推定値は香港とシンガポールではほぼ真値と一致し、また 95%信頼区間は発病中の全ての時期において真値を含む。これにより本論文の方法の有効性は検証されたと言える。一方、北京の回帰式は直線ではないので、折れ線回帰モデルを応用したフェーズ別死亡率を求めた。

考察

ノンパラメトリック競合リスクモデルの理論に基づくプロフィール尤度法を提案し、既存の方法との優劣を比較した。応用上重要な共変量を持つデータに適用可能なのは、本方法だけである。共変量の無いデータについて既存の方法とシミュレーションにより比較した結果、本方法の適用範囲が広く精度の高いことが確認された。今後新たに発生する新興感染症については、まず治癒対死亡ハザード比の散布図をもとにその比が時間に依らず一定かどうかを調べ、一定ならば本論文の提案方法を適用し、そうでないときは、折れ線回帰法によるフェーズ別死亡率を求めればよい。これから起きる新感染症に、もし治癒対死亡ハザード比が定数ならば、この方法での推定値の精度が最高い。