線形時間尺度関数の推定のためのスケール変換について

電気通信大　情報理工　山本渉
電気通信大　情報理工　金　路

製品の設計開発段階では、製品を均質な条件下に置いた寿命試験によって、寿命分布が推定される。また複数の条件下で行われた寿命試験のデータには、製品の使用条件と故障との関係を探るために、生存時間に関する回帰モデルがよく用いられる。寿命が、何かに晒されていることによって加速される場合には、加速寿命時間モデルやその拡張である累積曝露モデルが用いられる。

その故障時間が、複数の時間尺度で記録されると、問題が複雑になる。その場合の戦略は 2 つで、多変量寿命分布を推定するか、あるいは何らかの低次元の仮想的な寿命時間に変換して一次元の問題に帰着させるかを考えることになる。適切な多変量寿命分布がない場合には、コピュラが用いられるようになってきた。しかし、複数のリスク源への曝露の蓄積がそれぞれ記録されている場合には、それらの和が寿命の適切な時間尺度となることもある。線形時間尺度関数 [1] はそのような場合に用いられるモデルである。

$X_1, \ldots, X_p$ を暦時間を含む寿命に関係のある確率変数とすると、線形時間尺度関数は

$$T = \sum_{j=1}^{p} \beta_j X_j = \boldsymbol{X}'\boldsymbol{\beta} \tag{1}$$

と定義される。回帰モデルのようだが、右辺は線形結合で得る確率変数であり、観測はされない。変換の母数 $\boldsymbol{\beta}$ には、T の分布に尺度母数がある場合に識別可能とするために、$\sum_j \beta_j = 1$ という条件を課す。これらは各 X_j の T への貢献の度合いを表す母数である。

T が従う確率分布が母数を除いて既知の場合には、最尤法を用いてその分布の母数および $\boldsymbol{\beta}$ を推定する。しかし変換の前から寿命分布の関数形が既知とすることには、データ解析に用いる上では少し違和感が残る。そのために、セミパラメトリックな推定法 [2] が提案されている。この推定法は、T に関する分布を特定せずに $\boldsymbol{\beta}$ の推定を可能とする。

ところで、この推定法だが $X_1, \ldots, X_p$ それぞれの単位が異なる場合に、$\beta_1, \ldots, \beta_p$ が T への貢献度を表すという解釈が困難になる。そこで線形時間尺度関数を少し修正し、各変数をその期待値 $\mu_j = E\left[X_j\right]$ で除し、

$$T* = \sum_{j=1}^{p} {\beta_j}^* \frac{X_j}{\mu_j} = \boldsymbol{X}^{*\prime}\boldsymbol{\beta}^* \tag{2}$$

とし、制約も $\sum_j {\beta_j}^* = 1$ とすることを提案する。モデル 1 とモデル 2 は、等価ではあるが、後者の方が母数の解釈が容易となる。本発表では、打ち切りがあるデータについて、$\mu_1, \ldots, \mu_p$ をカプランマイヤー推定量を用いて推定してから、$\boldsymbol{\beta}$ をセミパラメトリック推定することを提案する。

参考文献

[1] Duchesne, T. and Lawless, J. L. (2000) Alternative time scales and failure time models, *Lifetime Data Analysis*, Vol. 6, pp. 157–179.

[2] Duchesne, T. and Lawless, J. L. (2002) Semiparametric inference methods for general time scale models, *Lifetime Data Analysis*, Vol. 8, pp. 263–276.