

有限混合モデルを用いた競合リスクの解析について

東京理科大学 下川 朝有

東京理科大学 宮岡 悦良

はじめに

本研究では、イベント発生時間と打ち切り時間の依存、及び競合するリスクを仮定した下での解析に着目する。観測時間及び打ち切り指標という一般的な生存データが与えられた下、イベント発生時間と打ち切り時間の依存は、追加情報無しに同定不可能であることは広く知られている。故に、その依存の違いによる推定量の感度分析が重要なテーマの一つとなる。本研究では、生存関数を各リスクの有限混合モデルを用いてモデル化し、また、あるリスク原因が与えられた下でのイベント発生時間及び打ち切り時間の結合分布を、既知のコピュラを用いてモデル化する。これらの仮定の下、EM アルゴリズムに基づく反復法を利用したパラメータ推定を行うことを提案し、シミュレーションを用いてその性能を比較する。

記述・モデル

T は死亡時間、 C は打ち切り時間、そして $X = \min(T, C)$ は観測時間を表す確率変数とする。また、 $Y \in \{1, 2, \dots, J\}$ は死亡原因（リスク）を表すとする。このとき観測データは $\{(x_i, y_i, \delta_i) : i = 1, 2, \dots, n\}$ で与えられ、ただし $\delta_i = I(X_i = T_i)$ は i 番目の標本に対するイベント指標を表すものとする。加えて、 y_i は $\delta_i = 1$ の標本に対しては観測されるが、 $\delta_i = 0$ の標本に対しては観測されず、 $y_i = 0$ とするものとする。

本研究では死亡原因 $Y = j$ が与えられた下、 T 及び C の周辺分布はそれぞれ定数ハザード ξ_j 及び θ によって与えられると仮定する。このとき T 及び C の条件付き分布関数はそれぞれ以下で与えられる：

$$\begin{aligned} F_j(t) &= P(T \leq t | Y = j) = 1 - \exp(-\xi_j t), \\ G_j(t) &= P(C \leq t | Y = j) = 1 - \exp(-\theta t). \end{aligned}$$

$Y = j$ が与えられた下、 T 及び C の結合関数は以下で与えられると仮定する：

$$P(T \leq t, C \leq t | Y = j) = H\{F_j(t), G_j(t); \alpha\},$$

ただし、 $H\{\cdot\}$ は T 及び C の依存度を表す既知のパラメータ α を持つコピュラとする。

推定

$\pi_j = P(Y = j)$ とすると、完全データが与えられていると仮定した場合における完全尤度は以下で与えられる：

$$L^* = \prod_j \prod_{i \in D_j} \pi_j [1 - H_u\{F_j(x_i), G_j(x_i)\}] \times \prod_j \prod_{i \in E_j} \pi_j [1 - H_v\{F_j(x_i), G_j(x_i)\}]$$

ここで $D_j = \{i; \delta_i = 1, y_i = j\}$ 及び $E_j = \{i; \delta_i = 0, y_i = j\}$ は、それぞれ原因 j によるイベント、打ち切りが観測された標本の集合を表す。また $H_u(a, b)$ 及び $H_v(a, b)$ は、それぞれ

$$H_u(a, b) = \frac{\partial H(u, v)}{\partial u} \bigg|_{(u, v) = (a, b)}, \quad H_v(a, b) = \frac{\partial H(u, v)}{\partial v} \bigg|_{(u, v) = (a, b)},$$

で与えられるとする。

実際には集合 E_j は未知であるため、本研究では EM アルゴリズムに基づく手法を用いて各パラメータを推定することを試みる。またコピュラ $H\{\cdot\}$ は事前に指定しておく必要があるが、本研究におけるシミュレーションでは Clayton copula を用いた場合の結果を示すとする。