

# Copulaの多重従属な競合リスクモデルにおける 識別可能性問題への応用

米山 寛二, 葉 光毅<sup>1</sup>, 松崎 友美<sup>2</sup>, 夏 皓清<sup>1</sup>, 石井 博昭<sup>3</sup>

社会環境部門, 中華民国国立成功大学<sup>1</sup>, 東京都立大泉桜高等学校<sup>2</sup>, 大阪大学大学院情報科学研究科<sup>3</sup>

Application of Copulas to the Identifiability Problem of the Multidependent Competing Risks

Kanji YONEYAMA, Kuang-Yih YEH<sup>1</sup>, Yumi MATSUZAKI<sup>2</sup>, Hao-Ching HSIA<sup>1</sup>, Hiroaki ISHII<sup>3</sup>

School of Human Science and Environment,

University of Hyogo,

1-1-12 Shinzaike-honcho, Himeji, 670-0092 Japan

National Cheng Kung University, Taiwan<sup>1</sup>

Ohizumi Sakura Senior High School, Tokyo, Japan<sup>2</sup>

Graduate School of Information Science and Technology, Osaka University, Japan<sup>3</sup>

**Abstract** In this paper we will deal with the case where we assume a known copula for the joint distribution of latent failure times to overcome the identifiability problem of a competing risks model in probabilistic risk assessment. After defining competing risks, basic probability functions, identifiability, copula and dependence we will study the effect of complete elimination of five competing risks on the overall survival function. For illustrative purposes, we have applied our methodology to the female Japan population, using cause specific mortality data set, provided by Journal of Health and Welfare Statistics.

keywords: competing risk, identifiability, copula, dependence, elimination

## 1 はじめに

近年、グローバリゼーションの進行によってリスク管理に関する研究が益々増えてきた。確率的リスク解析 (Probabilistic Risk Analysis : PRA), 定量的リスク解析 (Quantitative Risk Analysis : QRA) あるいは確率的安全解析 (Probabilistic Safety Analysis : PSA)が最近では輸送, 建設, エネルギー, 化学処理, 宇宙産業, 軍事を含む多くの分野に広く適用され, プロジェクト計画やファイナンスにまで応用されるようになってきた。解析的PRA手法は, 安全に対する要求を正当化するためやさらなる改善要求を示すために適用されることが増えてきている。これらすべての分野の傾向はPRAが意思決定管理のためのツールをサポートし, リスク管理の新しいエリアを形作ることにある [1]。

独立な故障時間確率変数を持つ競合リスクモデルが生物統計, 計量経済学, 医学, 人口統計学や保険統計学における多くの研究者によって考えられてきた。何故なら独立でない場合, 識別可能でないからである。しかしながら, たいていの実生活では, 特に人間の寿命に関しては, リスクに関する独立性の仮定は道理にかなっておらず, 疑わしい [2]。さらに死に至る時間や死因が与えられた時, 独立性を検定することすら不可能である。

保険統計学において多重減少モデルと呼ばれる競合リスクモデルが故障時間間の独立性の仮定の下で考えられてきた。Carrierは2変量Gaussian copulaを用いて, 人間の死亡数における2つの競合する死

因の1つの完全原因除去の効果をモデル化した [3]。2変量Frank copulaが解剖癌データに適用された [4]。Kaishev等は人間の死亡数における死因の部分原因除去と完全原因除去の両方についてその効果をモデル化した。さらに心臓疾患, 癌, 呼吸器疾患とその他の死因の4つの競合リスクの同時分布をモデル化するために4次元copulaの適用可能性を研究した。彼らの方法論はアメリカ合衆国・国立保健統計センター (National Center for Health Statistics : NCHS) によって提供された1989年~91年・死因別寿命統計に適用された [5]。

本論文ではKaishev等によって研究された多変量copulaベースの競合リスクモデルを拡張する。我々は糖尿病, 高血圧性疾患, 脳血管疾患, 肝疾患, 心臓疾患の5つの競合リスクの同時分布をモデル化するために5次元Frank copulaの適用可能性を研究する。

次節では潜在故障時間を用いた競合リスクの定型的な定義を与える。第3節では競合リスク問題で重要な幾つかの数学的関数を定義する。第4節では識別可能性を定義したあと識別可能性問題を克服する方法について述べる。第5節ではcopulaの定義とその性質について述べる。copulaは多変量分布関数を周辺分布を持つ部分モデルに分割することと変量間の依存性の意味を与える。もっとも広く使われている従属性尺度としてKedallの $\tau$ を紹介し, 第6節でこれとcopulaとの関係を述べる。Archimedean copulaは目下用いられているすべてのモデルの中で最も有名であるので, これについての基礎を第7節で与える。第8節では提案する方法論をどのように

実装するかを述べる。第9節では日本人女性の死因別寿命統計データに本方法論を適用する。幾つかのグラフでは5つの死因の任意の組合せの完全原因除去によって寿命がどれだけ延びるかを示す。最後に、第10節ではまとめを述べる。

## 2 競合リスク

形式的にペア  $(J, T)$  を観測する。ただし、 $J \in \{1, \dots, p\}$  は故障のタイプを表し、 $T > 0$  は故障時間を表す。かくして  $p$  個の異なる故障タイプが存在し、各故障は  $p$  タイプのいずれかに所属することが仮定される。ここで「故障」という言葉は一般的な言葉で実際はアプリケーションに応じて決まる。また「時間」はカレンダー時間である必要はないが、原理的には運転時間やサイクル数やクラック数のようなカレンダー時間を持つ非減少な適当な尺度である [6]。

$p$  個のリスクを持つ競合リスクを記述する直感的方法は故障時間  $T_j, j = 1, \dots, p$  に付随される。これら  $p$  個の時間はもし他のリスクが存在しないならば仮説的な故障時間と考えられ、それらは潜在故障時間と呼ばれる。すべてのリスクが存在する時、システムの故障までの観測された時間は故障の原因となる故障時間のうちの最も小さい値となる。かくして  $T = \min\{T_1, \dots, T_p\}$  で、もし  $T = T_p$  ならば  $J = p$  となり ペア  $(J, T)$  が観測される。ここで最小値  $T_p$  は唯一に決まると仮定する。

例えば、医学では、 $J$  は死因で、 $T$  は死亡時刻である。信頼性理論では、 $J$  はシステムにおける故障部品を同定し、 $T$  は稼働開始から故障までの運転時間を表す。工学分野では、我々の関心は修理なしの直列システムの信頼性のことと同じである。経済学では、 $T$  は失業記録に登録されている期間を表し、 $J$  は失業記録から抹消された理由を表す [2]。

## 3 基本確率関数

確率変数  $T$  の分布関数  $F$  と生存分布関数  $\bar{F}$  はそれぞれ  $F(t) = pr(T \leq t)$ ,  $\bar{F} = pr(T > t)$  によって定義される。したがって、すべての  $t$  に対して  $F(t) + \bar{F}(t) = 1$  である。

密度関数は以下のように定義される。

$$f(t) = dF(t)/dt = -d\bar{F}(t)/dt \quad (1)$$

モデルの識別可能性の見地から  $J$  と  $T$  の同時分布に関心がある。これはいわゆる部分分布関数

$$F(j, t) = pr(J = j, T \leq t) \quad (2)$$

あるいは部分生存分布関数

$$\bar{F}(j, t) = pr(J = j, T > t) \quad (3)$$

によって明らかにされる。これらの関数には

$$F(j, t) + \bar{F}(j, t) = p_j \quad (4)$$

なる関係が成り立つ。ただし、

$$p_j = pr(J = j) = F(j, \infty) = \bar{F}(j, 0) \quad (5)$$

は  $J$  の周辺分布を与える。

かくして  $t = \infty$  で1ではなく値  $p_j$  となることから

$F(j, t)$  は proper な分布関数ではない。陽には

$p_j > 0$  で  $\sum p_j = 1$  とする。一般的には  $\bar{F}(j, t)$  は

故障タイプ  $j$  について  $T > t$  の確率に等しくなく、

その確率は条件付き確率  $pr(T > t | J = j)$  で

$\bar{F}(j, t)/p_j$  と表現される。連続変数  $T$  に対する部

分密度関数  $f(j, t)$  は  $-d\bar{F}(j, t)/dt$  となる。

競合リスクの文脈では  $T_j$  は特別な意味を持つ

が、本節ではそれらを非負の確率変数として扱う。

同時分布関数は

$$H(t_1, \dots, t_p) = pr(T_1 \leq t_1, \dots, T_p \leq t_p) \quad (6)$$

で定義され、同様に同時生存分布関数は

$$\bar{H}(t_1, \dots, t_p) = pr(T_1 > t_1, \dots, T_p > t_p) \quad (7)$$

で定義される。もし  $T_j$  が連続であるならば、

それらの同時密度関数はそれぞれ

$$\begin{aligned} & \partial^p H(t_1, \dots, t_p) / \partial t_1 \dots \partial t_p, \\ & (-1) \partial^p \bar{H}(t_1, \dots, t_p) / \partial t_1 \dots \partial t_p \end{aligned} \quad (8)$$

によって計算される。

$T_j$  の独立性は次式によって定義される。

$$H(t_1, \dots, t_p) = \prod_{j=1}^p G_j(t_j), \quad (9)$$

$$\bar{H}(t_1, \dots, t_p) = \prod_{j=1}^p \bar{G}_j(t_j),$$

ただし、 $G_j(t_j) = pr(T_j \leq t_j)$  は  $T_j$  の周辺分布関数

で、 $\bar{G}_j(t_j) = pr(T_j > t_j)$  は  $T_j$  の周辺生存分布関数

である。

## 4 識別可能性

競合リスク解析の主な関心事は潜在故障時間  $T_1, \dots, T_p$  の同時分布ならびに周辺分布にある。しかし、問題は観測されるペア  $(J, T)$  の分布が一般的に

潜在故障時間の分布を決定しないということである。

標準的な言葉で言えば、 $T_1, \dots, T_p$  の同時および周辺

分布関数は  $(J, T)$  の観測からは識別不可能と言われる。

次のCrowderの定理は  $(J, T)$  の同じ分布を与

える  $T_1, \dots, T_p$  の幾つかの異なる同時分布が存在することを述べている [2] .

### [定理 1 ](Crowder)

$\bar{G}_j(t_j)(j = 1, \dots, p)$  を 1 変量周辺分布関数の集合とし, 部分生存分布関数  $\bar{F}(j, t)$  が分かっているものとする. (i) 部分生存分布関数は観測データにマッチしており, (ii) 確率ゼロの開集合を全く含まないような  $R^p$  で開集合となるある集合  $\Omega$  が存在する, このような同時生存分布関数  $\bar{H}(t_1, \dots, t_j)$  を発見できると仮定すると, (i) と (ii) を満たす可算個の異なる同時生存分布関数が存在する.

識別可能性の数学的定義は次のように与えられる.  $\Phi$  を同時生存分布関数の族,  $\Psi$  を部分生存分布関数の族とする.  $\Gamma$  を  $\Phi$  から  $\Psi$  への写像とし,  $\Gamma(\bar{H}) = \bar{F}$  のように定義する.  $\Gamma$  が単射であるときはいつでも  $\Phi$  は  $\Psi$  によって識別可能であると言う. すなわち, もし  $\Gamma(\bar{H}_A) = \Gamma(\bar{H}_B)$  であるならば,  $\bar{H}_A = \bar{H}_B$  である. しかし, 一般に  $\Gamma$  は単射でないことが知られている. この状況はいわゆるノンパラメトリックの場合に常に発生する.

潜在故障時間表現による競合リスクの下で識別可能性問題を克服する以下の 4 つの方法がある [6] .

- ①独立なリスクを仮定する.
- ②潜在故障時間に対して既知のcopulaを仮定する.
- ③周辺生存分布関数に対する境界を求める.
- ④パラメトリック・モデルを用いる.

我々は②の同時生存分布関数に対して既知のcopulaを仮定する方法を採用する.

## 5 Copula

統計学におけるcopulaの研究とその応用は比較的新しい現象である. ごく最近まで統計学の文献にcopulaという言葉が配置することすら困難であった. copulaとは何なのか? ある観点から言えば, copulaとは多変量分布関数をそれらの 1 次元周辺分布関数に結合またはカップルさせる関数である. 別の表現をとれば, copulaはその 1 次元周辺分布が区間  $(0, 1)$  上で一様な多変量分布関数である. 何故, copulaが

確率論や統計学の研究者たちに関心があるのだろうか? FisherがEncyclopedia of Statistical Science の第 1 回改訂版で述べているように, 「copulaは 2 つの主な理由から統計学者に関心がある. 1 つはスケールフリーな従属性尺度を研究する方法として, そして 2 つ目はしばしばシミュレーションの観点から 2 変量分布関数の族を構成するスタートポイントとして」. copulaという言葉はラテン語の名詞で「リンク, 結び目, 絆」を意味し, 文法や主部と述部を結合する命題の役割を表す論理学で用いられる. copulaという言葉は定理の中でAbe Sklarによって数学的意味あるいは統計学的意味で初めて採用された [7] .

### [定理 2 ](Sklar)

$H$  が周辺分布  $G_1, \dots, G_n$  を持つ  $n$  次元分布関数であるならば,  $\bar{R}_n$  上のすべての  $(x_1, \dots, x_n)$  に対して, 次式を満たす  $n$ -copula  $C$  が存在する.

$$H(x_1, \dots, x_n) = C(G_1(x_1), \dots, G_n(x_n)) \quad (10)$$

もし  $G_1, \dots, G_n$  がすべて連続であるならば,  $C$  は唯一である. そうでない場合,  $\text{Ran}G_1 \times \dots \times \text{Ran}G_n$  上で唯一決定される. ただし,  $\text{Ran}G_j$  は  $G_j, j = 1, \dots, n$  のレンジである. 逆に,  $C$  が  $n$ -copula で  $G_1, \dots, G_n$  が分布関数であるならば, (10) で定義される関数  $H$  は周辺分布  $G_1, \dots, G_n$  を持つ  $n$  次元分布関数となる.

$X$  と  $Y$  のcopulaを  $C$  とすると, 次式を得る.

$$\begin{aligned} \bar{H}(x, y) &= 1 - G_1(x) - G_2(y) + H(x, y) \\ &= \bar{G}_1(x) + \bar{G}_2(y) - 1 + C(G_1(x), G_2(y)) \quad (11) \\ &= \bar{G}_1(x) + \bar{G}_2(y) - 1 + C(1 - \bar{G}_1(x), 1 - \bar{G}_2(y)) \end{aligned}$$

したがって,  $I^2$  から  $I$  への関数  $\hat{C}$  を

$$\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v) \quad (12)$$

で定義するならば,

$$\bar{H}(x, y) = \hat{C}(\bar{G}_1(x), \bar{G}_2(y)) \quad (13)$$

を得る. まず, 第1に  $\hat{C}$  はcopulaで,  $X$  と  $Y$  の survival copula と言う. 第2に  $\hat{C}$  はcopulaが同時分布関数とその周辺分布関数を結合した方法と完全に同じ方法で, 同時生存分布関数とその1次元周辺分布関数をカップルする.

## 6 従属性

本節では確率変数間の従属性あるいは提携にどのようにcopulaが用いられるかを示す. 従属性を議論し, また測る方法が幾つかある. これらの性質と尺度の多くはHoeffdingの言葉を借りれば, 確率変数の狭義増加変換の下で不変であるという意味の scale-invariant (スケール不変性) という言葉で示される [7]. Correlation (相関) という言葉は2変量間のあらゆる種類の従属性を記述する横断的な用語としてしばしば用いられ, 誤用されている. 我々はここでは線形相関の意味に限ってこの用語を用い,  $\rho$  を次式で定義する.

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}X} \sqrt{\text{Var}Y}} \quad (14)$$

ただし  $X$  と  $Y$  はゼロでない有限分散を持つ2つの確率変数で,  $\text{Cov}$  と  $\text{Var}$  はそれぞれ共分散と分散を表す.  $\rho$  は線形従属の尺度で,  $Y$  に線形的に従属 (すなわち  $Y = aX + b$ , ただし  $a, b \in \mathbb{R}$ ,  $a \neq 0$ ) するならば,  $\rho(X, Y)$  の絶対値は1に等しい. また, 線形相関は狭義増加線形変換の下でのみ不変である. さらに, 線形相関は多変量正規分布に対して用いられる時, 正確な従属性尺度となるが, 他の同時分布関数に対しては非常にミスリードな印象を与える. 線形相関係数は2変量間の関連を測る尺度としては貧弱な尺度である. 特に, 狭義非線形変換の下では不変ではなく, このことが従属性のグローバルな尺度として使用する際の主な障害となっている. 従って, 代わりに狭義非線形変換の下でも不変である順位相関係数Kendallの  $\tau$  が用いられる. 例えば, 身長の高い者から順番に並んでいる学生の体重を測定した場合, 身長と体重の相関係数を(14)式では計算できないが, 順位相関係数Kendallの  $\tau$  は計算することができる. 一方, copulaは狭義増加変換の下で不変であるので, Kendallの  $\tau$  はcopula で表現される [8].

$X$  と  $Y$  はそのcopulaが  $C$  である連続確率変数とする.  $X$  と  $Y$  の Kendallの  $\tau$  は次式で与えられる.

$$\tau = 4 \iint_{I^2} C(u, v) dC(u, v) - 1 \quad (15)$$

## 7 多変量copulaの構成

Copulaの文献では多くの2変量分布関数の族が紹介されているが, 多変量の分布関数はほとんどない. しかも, これらを扱うのはやさしくない. 2変量の分布族の幾つかはより高い次元に拡張されるが, 一般的に2変量以上の分布関数の構成は難しい問題である. この問題を克服する方法として以下の4つの方法がある [9].

- (a)統計的あるいは数学的性質
- (b)pair-copula
- (c)Archimedean copulaの生成子
- (d)階層型 Archimedean copula

我々はArchimedean copulaの生成子を用いて多変量化を行う.

$\varphi$  を連続で,  $\varphi(1) = 0$  となるような  $I$  から  $[0, \infty]$  への狭義減少関数とする.  $\varphi$  のpseudo-inverse とは

$$\begin{aligned} \varphi^{[-1]}(t) &= \varphi^{-1}(t) \quad (0 \leq t \leq \varphi(0)) \\ \varphi^{[-1]}(t) &= 0 \quad (\varphi(0) \leq t \leq \infty) \end{aligned} \quad (16)$$

によって与えられる  $\text{Dom}\varphi^{[-1]} = [0, \infty]$  と  $\text{Ran}\varphi^{[-1]} = I$  を持つ関数  $\varphi^{[-1]}$  である. ただし,  $\text{Dom}\varphi^{[-1]}(t)$  は  $\varphi^{[-1]}(t)$  のドメインである.

もし関数  $g(t)$  が連続で符号を変化させるすべての次数の導関数を持つ, すなわち区間  $E$  の内包におけるすべての  $t$  と  $k = 0, 1, 2, \dots$  に対して

$$(-1)^k \frac{d^k}{dt^k} g(t) \geq 0 \quad (17)$$

であるならば, 関数  $g(t)$  は区間  $E$  上でcompletely monotonicである.

次の定理は完全生成子  $\varphi$  が  $n \geq 2$  の Archimedean  $n$ -copulasを生成するための必要十分条件を与える. ただし,  $\varphi(0) = \infty$  ならば  $\varphi$  を完全生成子と呼ぶ.

### [定理 3]

$\varphi$  は  $\varphi(0) = \infty$ ,  $\varphi(1) = 0$  となるような  $I$  から  $[0, \infty]$  への連続な狭義減少関数とし,  $\varphi$  の逆関数を

$\varphi^{-1}$  で表す. もし  $C^n$  (ただし  $C$  の上付き添え字は次元を表す) が次式で表される  $I^n$  から  $I$  への関数ならば,  $C^n$  が  $n \geq 2$  に対して  $n$ -copula となる必要十分条件は,  $\varphi^{-1}$  が  $[0, \infty)$  上で *completely monotonic* であることである.

$$C^n(u_1, \dots, u_n) = \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_n)) \quad (18)$$

## 8 copulaを用いた多重従属性のモデル化

本節では部分生存分布関数の推定値と適当に選択されたcopulaが与えられた時, どのようにして非線形常微分方程式を解くことによって対応する競合リスクに対する周辺生存分布関数を評価するのかを述べる.

### 8. 1 部分生存分布関数・部分生存密度関数

適当な  $p$  個の死因別寿命統計データに基づき, 部分生存分布関数と部分生存密度関数を推定することが可能である. もし各死因に対するデータが既に平滑化されているならば, スプライン補間や他の別の近似法を使って直接内挿される.

### 8. 2 周辺生存分布関数

survival copula  $\hat{C}(u_1, \dots, u_p)$  を固定し, 周辺生存分布関数  $\bar{G}_j(t_j), j = 1, \dots, p$  が分かると, (13)式から同時生存分布関数を評価することができる.

$$\bar{H}(t_1, \dots, t_p) = \hat{C}(\bar{G}_1(t_1), \dots, \bar{G}_p(t_p)) \quad (19)$$

周辺生存分布関数を見つけるために,  $\bar{G}_j(t_j)$  と部分生存分布関数  $\bar{F}(j, t), j = 1, \dots, p$  の関係式を用いることができる [3] [10]. かくして  $t > 0$  に対して,  $u_j \in (0, 1)$  に関する  $\hat{C}(u_1, \dots, u_p)$  の微分可能性と  $t_j > 0$  に関する  $\bar{G}_j(t_j)$  の微分可能性を仮定すると, 次の連立微分方程式が成り立つ.

$$\frac{d}{dt} \bar{F}(1, t) = \hat{c}_1(\bar{G}_1(t), \dots, \bar{G}_p(t)) \times \frac{d}{dt} \bar{G}_1(t)$$

$$\frac{d}{dt} \bar{F}(2, t) = \hat{c}_2(\bar{G}_1(t), \dots, \bar{G}_p(t)) \times \frac{d}{dt} \bar{G}_2(t)$$

$$\frac{d}{dt} \bar{F}(p, t) = \hat{c}_p(\bar{G}_1(t), \dots, \bar{G}_p(t)) \times \frac{d}{dt} \bar{G}_p(t), \quad (20)$$

ただし,

$$\hat{c}_j(u_1, \dots, u_p) = \frac{\partial}{\partial u_j} \hat{C}_j(u_1, \dots, u_p), \quad j = 1, \dots, p \quad (21)$$

上の方程式は非線形であるが, いったん, copulaのタイプを明示し, 部分生存密度関数が推定されて適当な関数形を取った時, (20)式の左辺に代入されると, 非線形連立微分方程式は周辺生存分布関数  $\bar{G}_j(t_j)$  に関して解くことができる. 実際に周辺生存分布関数  $\bar{G}_j(t_j)$  を得るためには, 効果的な数値解析手法を見つける必要がある.

### 8. 3 全生存分布関数

いったん周辺生存分布関数が得られると, それらを(19)式の右辺のcopulaに代入し, 同時生存分布関数を, 実際は我々の研究において主に関心がある全生存分布関数  $\bar{G}(t, \dots, t)$  を評価することができる.

## 9 数値結果

先に述べたごとく本方法論を厚生労働省「人口動態統計」第54巻第16号「厚生指標」(2007年・特別編集号)の表2にある「生活習慣病に関する死亡数」の女性に適用する [11]. 我々は Kashev 等の方法論を糖尿病 [1], 高血圧性疾患 [2], 脳血管疾患 [3], 肝疾患 [4], 心臓疾患 [5] の5つの競合リスクの場合に適用する.

### 9. 1 5次元Frank copula

Frank copulaは関数方程式  $C(u, v) = \hat{C}(u, v)$  を満たす唯一の Archimedean copula である.  $\varphi_\theta(t) = -\ln[(e^{-\theta t} - 1)/(e^{-\theta} - 1)]$  とおくと, これは2次元Frank copulaとなる. Frank copulaに対して, 生成子generator  $\varphi_\theta^{-1}(t)$  は次式で与えられる.

$$\varphi_\theta^{-1}(t) = -\frac{1}{\theta} \ln[1 - (1 - e^{-\theta})e^{-t}] \quad (22)$$

ところで,  $\theta > 0$  に対して, 関数

$f(x) = -\ln(1-x)/\theta$  は  $(0,1)$  内の  $x$  に対して absolutely monotonic で,  $g(t) = (1-e^{-\theta})e^{-t}$  は  $[0,\infty)$  内の  $t$  に対して completely monotonic であるから,  $\varphi_\theta^{-1}$  は  $[0,\infty)$  で completely monotonic となる.

かくして,  $\theta > 0$  に対して, 2変量Franc copulaを  $n$ -copula ( $n \geq 2$ ) に拡張することができる.  $n = 5$  について次式を得る.

$$\hat{C}(u_1, \dots, u_5) = -\frac{1}{\theta} \ln \left[ 1 + \frac{\prod_{j=1}^5 (e^{-\theta u_j} - 1)}{(e^{-\theta} - 1)^4} \right] \quad (23)$$

上式を  $u_k$  ( $k = 1, \dots, 5$ ) に関して偏微分すると次式を得る.

$$\hat{c}_k(u_1, \dots, u_5) = \frac{\partial \hat{C}(u_1, \dots, u_5)}{\partial u_k} = \frac{e^{-\theta u_k} \prod_{j \neq k} (e^{-\theta u_j} - 1)}{(e^{-\theta} - 1)^4 + \prod_{j=1}^5 (e^{-\theta u_j} - 1)} \quad (24)$$

連立微分方程式(20)を解くためにこれらの式を用いる. 多変量Frank copulaを使用する際, Kaishev等の先行研究で正の相関がある程, 全生存確率は改善されることが分かっているが, 最善のパラメータ推定は今後の課題として, パラメータ  $\theta = 4.3$  とし正の従属性をモデル化した.

### 9. 2 完全原因除去

第8節で述べた手順の第1ステップとして, 我々は通常のスプライン補間法を使って0歳から100歳までの間に対して, 部分生存分布関数  $\bar{F}(j,t)$  ( $j = 1, \dots, 5$ ) の観測値に対する内挿を行う [12].

糖尿病, 高血圧性疾患, 脳血管疾患, 肝疾患, 心臓疾患の5つの競合リスクそれぞれの部分生存分布関数は図1のようになる.

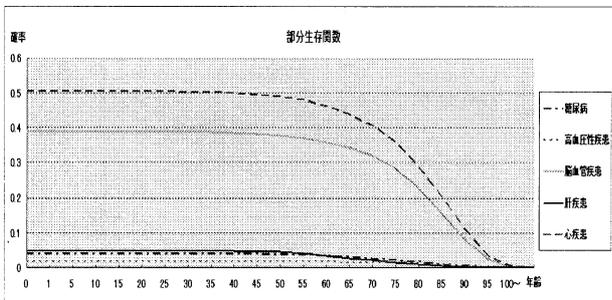


図1 部分生存分布関数

第2ステップとして, Runge-Kutta-Gill法を用いて連立微分方程式(20)を解くことによって周辺生存

分布関数  $\bar{G}_j(t)$  ( $j = 1, \dots, 5$ ) を求める.

第3ステップとして,  $\bar{G}_j(t)$  ( $j = 1, \dots, 5$ ) を(19)に代入して(23)を使って, 同時生存分布関数  $\bar{H}(t_1, \dots, t_5)$  と全生存分布関数  $\bar{G}(t, \dots, t)$  を得る.

明らかに  $r$ -番目の病気を完全除去した全生存分布関数は

$$\begin{aligned} \bar{G}^{(-r)}(t) &\equiv pr(T_1 > t, \dots, T_{r-1} > t, T_{r+1} > t, \dots, T_5 > t) \\ &= \hat{C}(\bar{G}_1(t), \dots, \bar{G}_{r-1}(t), 1, \bar{G}_{r+1}(t), \dots, \bar{G}_5(t)) \end{aligned}$$

であるから

$\bar{G}_r(t) = pr(T_r > t) = 1$  は  $r$ -番目の病気の完全原因除去に対応する.

図2では糖尿病[1], 高血圧性疾患[2], 脳血管疾患[3], 肝疾患[4], 心臓疾患[5]のうちの任意の1つの病気が完全原因除去された時の全生存分布関数を与える. 明らかに心臓疾患[5]が完全原因除去されると生存に関して最も改善される. この結果はKashev等がアメリカ合衆国・国立保健統計センター (National Center for Health Statistics : NCHS) によって提供された1989年~91年・死因別寿命統計の85歳以上の女性に関する結果と同じである [5].

図3では同時に2つの病気が完全原因除去された時の例を示す. 脳血管疾患[3]と心臓疾患[5]が同時に完全除去されると他のどのような2つの病気の組合せよりも全年齢にわたって最も改善される.

図4では同時に3つの病気が完全原因除去された時の例を示す. 年齢が  $0 \leq t < 90$  では脳血管疾患[3], 肝疾患[4], 心臓疾患[5]が, 年齢が  $t \geq 90$  では高血圧性疾患[2], 脳血管疾患[3], 心臓疾患[5]が同時に完全原因除去されると他のどのような3つの病気の組合せよりも最も改善される.

図5では同時に4つの病気が完全原因除去された時の例を示す. 年齢が  $0 \leq t < 90$  では糖尿病[1], 脳血管疾患[3], 肝疾患[4], 心臓疾患[5]が, 年齢が  $t \geq 90$  では糖尿病[1], 高血圧性疾患[2], 脳血管疾患[3], 心臓疾患[5]が同時に完全原因除去されると他のどのような4つの病気の組合せよりも最も改善される.

### 10 まとめ

本論文では競合リスクの識別可能性問題を克服するために潜在故障時間の同時分布として既知のcopulaを仮定する場合を扱い, 全生存分布関数について5つの競合リスクの完全原因除去の効果について研究した. 応用事例として本方法論を財団法人厚生統計協会出版の「厚生指針」(2007年・特別編集号)における「生活習慣病に関する死亡数」の女性に適用した. 本論文はKaishev等の方法論を整理し, 実際に日本人の実データでその方法論を実装したところに意義がある. 今後の課題として, copula

による従属性のモデル化が果たして適しているかのシミュレーションによる検証や適切なcopulaの選択、さらに従属性を的確に表現できるcopulaにおけるパラメータの推定などがある。さらに、完全原因除去による全生存確率の改善に関しては、医学的な知見による検証が必要である。

参考文献

- [1] Bedford, T. and Cooke, R., *Probabilistic Risk Analysis, Foundation and Methods*, Cambridge University Press, 2001.
- [2] Crowder, M.J., *Classical Competing Risks*, Chapman & Hall, 2001.
- [3] Carriere, J., *Dependent decrement theory*, Transactions, Society of Actuaries, V. XLVI, 45-65, 1994.
- [4] Escarela, G. and Carriere, J., *Fitting competing risks with assumed copula*, Statistical Methods in Medical Research, 12(4), 333-349, 2003.
- [5] Kaishev, V.K., Dimitrova, D.S. and Haberman, S., *Modelling the joint distribution of competing risks survival times using copula functions*, Insurance: Mathematics and Economics, 41(3), 339-361, 2007.
- [6] Lindqvist, B.H., *Competing risks*, in *Encyclopedia of Statistics in Quality and Reliability*, Ruggeri, F., Kenett, R. and Faltin, F.W. eds, Wiley, Chichester, 335-341, 2007.
- [7] Nelsen, R.B., *An Introduction to Copulas*, Second Edition, Springer, 2006.
- [8] Schirmacher, D. and Schirmacher, E., *Multivariate dependence modeling using pair-copulas*, Working Paper n. 183/2008, Department of Applied Mathematics, University of Venice, 2008.
- [9] Lindskog, A.J., *Modelling Dependence with Copulas and Applications to Risk Management*, Master Thesis, 2000, [http://www.math.ethz.ch/research/groups/ifor/publications/2000\\_masterthesis\\_lindskog.pdf](http://www.math.ethz.ch/research/groups/ifor/publications/2000_masterthesis_lindskog.pdf)
- [10] Heckman, J.J. and Honore, B.E., *The identifiability of the competing risks model*, Biometrika, 76(2), 325-330, 1989.
- [11] 財団法人厚生統計協会, 厚生指標, 臨時増刊 生活習慣病と関連統計の年次推移, 第54巻第16号, 104-105, 2007.
- [12] De Boor, C., *A Practical Guide to Splines*, Revised Edition, Springer, 2001

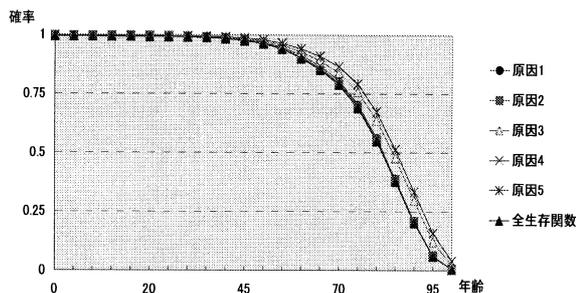


図2 完全原因除去の効果 (1つの場合)

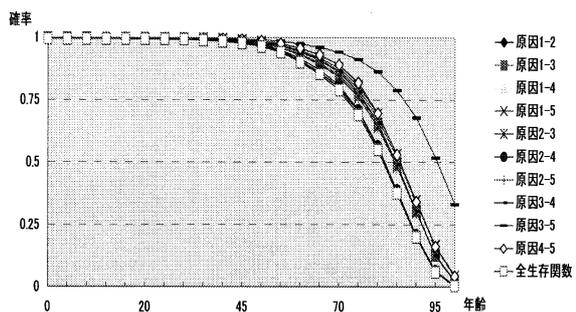


図3 完全原因除去の効果 (2つの場合)

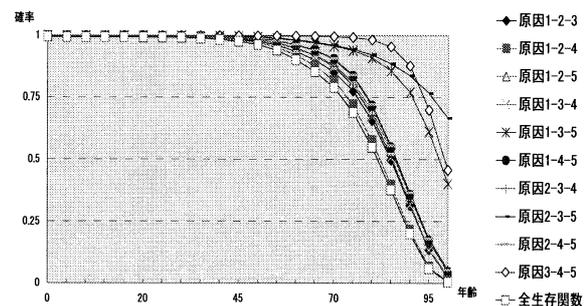


図4 完全原因除去の効果 (3つの場合)

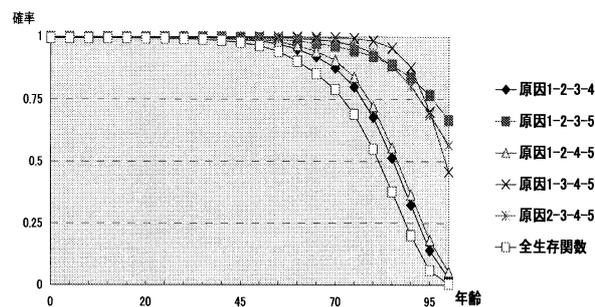


図5 完全原因除去の効果 (4つの場合)

(平成21年 9月25日受付)