

2019年10月26日 日本アクチュアリー会
2019年度 第1回研究集会
Vine Copulaを用いた多国間の死亡率の従属関係のモデル化

プルデンシャルジブラルタファイナンシャル生命保険株式会社
数理チーム 鈴木 理史

本発表の構成

- ▶ 研究の背景・モチベーション
- ▶ Vine copulaとは
- ▶ 死亡率の予測
- ▶ Vine copulaを用いた同時分布の推定
- ▶ 計算結果と考察
- ▶ モンテカルロシミュレーション
- ▶ まとめと今後の課題

研究の背景・モチベーション

- ▶ 複数の国で生命保険ビジネスや年金ビジネスを行う保険会社は、必要資本の計算やストレステストの実施の際に、多国間の死亡率の従属関係に関心を持つ。
- ▶ 死亡率の従属関係を表現する方法として、コンピュータを用いた手法が数多く研究されている。
- ▶ しかしながら、代表的な楕円コンピュータやアルキメデスコンピュータでは、分布のテイル部分の従属関係を柔軟に表現できず、ストレス時のような複雑な依存関係をとらえることが難しいことが知られている。
- ▶ そこで本稿では、vine copulaを用いた多国間の死亡率の従属関係のモデル化を提案する。

代表的なコピュラの欠点

- ▶ アルキメデスコピュラでは、変数間で従属関係の違いを変えられず、多国間の死亡率の従属性のモデル化には不向き。
- ▶ 正規コピュラは裾従属性がないためテイルリスクを過小推定してしまう。
- ▶ tコピュラは自由度によって裾従属性を柔軟に変更できるも、リスクの従属関係が左右対称であること、裾での依存性を表すパラメータが1つ(自由度)しかないことから従属性の柔軟なモデル化が難しい。

コピュラ	パラメータ数	下裾従属係数	上裾従属係数
ガンベルコピュラ	1	0	$2 - 2^{1/\alpha}$
クレイトンコピュラ	1	$2^{-1/\alpha}$	0
正規コピュラ	$n(n-1)/2$	0	0
tコピュラ	$n(n-1)/2+1$	$2t_{v+1} \left(-\sqrt{\frac{(v+1)(1-\rho)}{1+\rho}} \right)$	$2t_{v+1} \left(-\sqrt{\frac{(v+1)(1-\rho)}{1+\rho}} \right)$

n : 変数数、v : t分布の自由度、ρ : ピアソンの相関係数

Vine copulaとは

3次元確率ベクトル (X_1, X_2, X_3) の同時分布関数は条件付確率を用いて分解すると

$$f(x_1, x_2, x_3) = f_3(x_3) f_{2|3}(x_2|x_3) f_{1|2,3}(x_1|x_2, x_3) \quad \dots \textcircled{1}$$

Sklarの定理より

$$f(x_2, x_3) = c_{2,3}(F_2(x_2), F_3(x_3)) f_2(x_2) f_3(x_3) \quad \dots \textcircled{2}$$

同時密度関数と条件付密度関数の関係より

$$f(x_2, x_3) = f_{2|3}(x_2|x_3) f_3(x_3) \quad \dots \textcircled{3}$$

②式と③式を組み合わせると条件付密度関数 $f_{2|3}(x_2|x_3)$ は

$$f_{2|3}(x_2|x_3) = c_{2,3}(F_2(x_2), F_3(x_3)) f_2(x_2) \quad \dots \textcircled{4}$$

と表せる。同様に

$$f_{1|2,3}(x_1|x_2, x_3) = c_{1,3|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) f_1(x_1|x_2) \quad \dots \textcircled{5}$$

$$f_1(x_1|x_2) = c_{1,2}(F_1(x_1), F_2(x_2)) f_1(x_1) \quad \dots \textcircled{6}$$

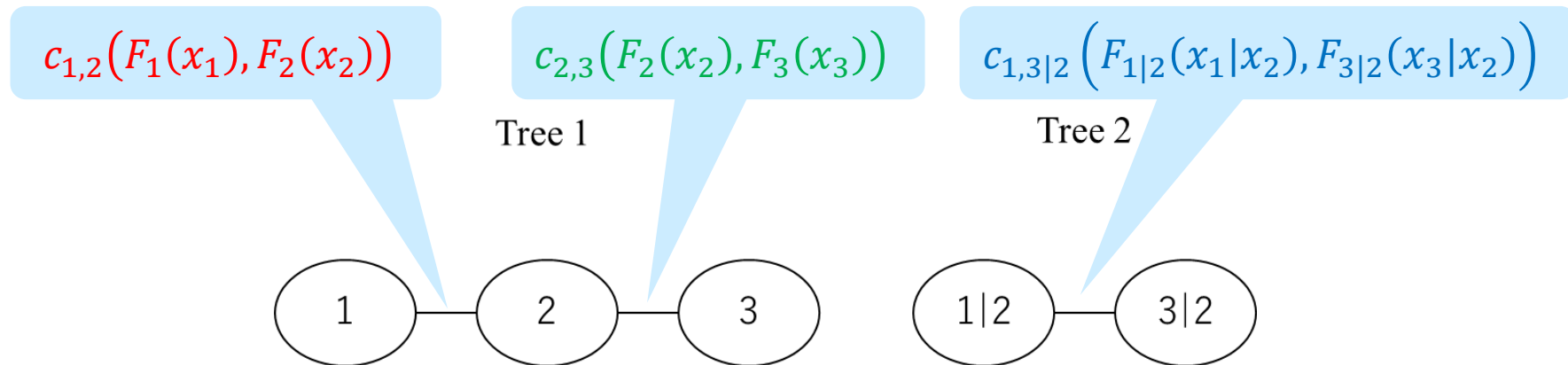
④、⑤、⑥式を①式に代入すると、3次元確率ベクトル (X_1, X_2, X_3) の同時分布関数は

$$f(x_1, x_2, x_3) = c_{1,3|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) c_{1,2}(F_1(x_1), F_2(x_2)) \\ \times c_{2,3}(F_2(x_2), F_3(x_3)) f_1(x_1) f_2(x_2) f_3(x_3)$$

同様の方式で4次元以上の同時分布に拡張することができ、一般に4次元以上の同時密度関数についても、2変量コピュラ密度関数と周辺密度関数の積で表現できる。

Vine copulaとは

- ▶ このような従属構造はグラフ理論の概念であるvineを用いて表現できる。
- ▶ 頂点が確率変数を表し、辺が条件付2変量コピュラを表している。
- ▶ 変数の間の従属関係を階層（tree）構造で表現し、それぞれの辺に条件付2変量コピュラを適用することで、同時分布を求める手法である。
- ▶ 複雑な従属関係をフレキシブルに表現できる手法として近年注目が集まっている。



$$f(x_1, x_2, x_3) = c_{1,3|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) c_{1,2}(F_1(x_1), F_2(x_2)) \\ \times c_{2,3}(F_2(x_2), F_3(x_3)) f_1(x_1) f_2(x_2) f_3(x_3)$$

- ▶ 6 vine copulaの性質に関する詳細はGruber(n.d.)を参照。

死亡率の予測

使用する死亡率データとモデル

使用するデータ	:	Australia(AUS) + 欧州の12カ国の国民死亡率
使用する年齢	:	60歳～85歳（男性のみ）
性別	:	男性のみ
使用する年	:	1921年～2009年
データ元	:	Human Mortality Database

- ▶ それぞれの国の死亡率をLee-Carterモデルでモデル化。
- ▶ 計算は、統計ソフトRのパッケージ"StMoMo"を使用。

Lee-Carter モデルの概要

年齢毎の κ_t による死亡率改善の度合い

$$\log(q_{x,t}) = \alpha_x + \beta_x \cdot \kappa_t + e_{x,t}$$

基準死亡率

暦年 t における死亡率改善

$q_{x,t}$: x 歳の暦年 t における死亡率

α_x : 各年齢の観測期間の平均的な死亡率の対数值

β_x : κ_t が変化するときの各年齢の死亡率の影響度合い

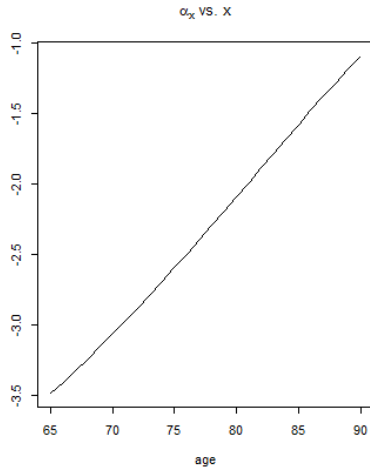
κ_t : 各暦年における死亡率への影響の程度

$e_{x,t}$: 誤差項

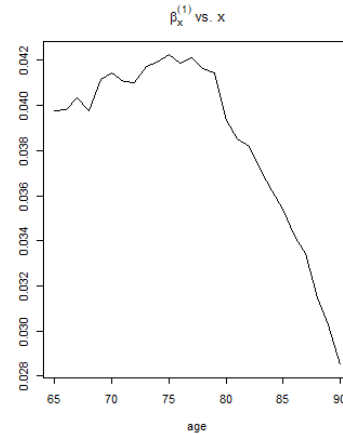
過年度の実績データをもとに α_x , β_x , 過年度の κ_t をそれぞれ推定する。

将来の κ_t の推定では時系列モデル (ARIMAモデル等) が用いられるのが一般的である。

パラメータの推定結果 (Franceの結果を抜粋)

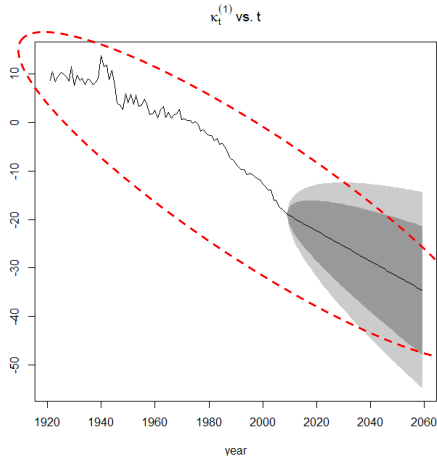


α_x : 平均的な死亡率の対数値



β_x : κ_t が変化するときの死亡率の影響度合い

$$\log(q_{x,t}) = \alpha_x + \beta_x \cdot \kappa_t + e_{x,t}$$



κ_t はおおむね直線的に低下している
 \Rightarrow ドリフト項付ランダムウォークでモデル化

$$\kappa_t = \kappa_{t-1} + \mu + \varepsilon_t$$

κ_t : 各暦年における死亡率への影響の程度

コンピュータを用いた 同時分布の推定

データを経験コピュラに変換

対象とする国($j=1,2,\dots,13$)の死亡率を以下のLee-Carterモデルで推定。

$$\log(q_{x,t}^j) = \alpha_x^j + \beta_x^j \kappa_t^j + e_{x,t}^j$$
$$\kappa_t^j = \kappa_{t-1}^j + \mu^j + \varepsilon_t^j$$

κ_t^j の残差項に正規分布を仮定しているため、サンプルデータを標準偏差で正規化。

$$\mathbf{z}_j = \frac{1}{\sigma_j} \left(\varepsilon_{1921}^j, \varepsilon_{1922}^j, \dots, \varepsilon_{2009}^j \right)'$$

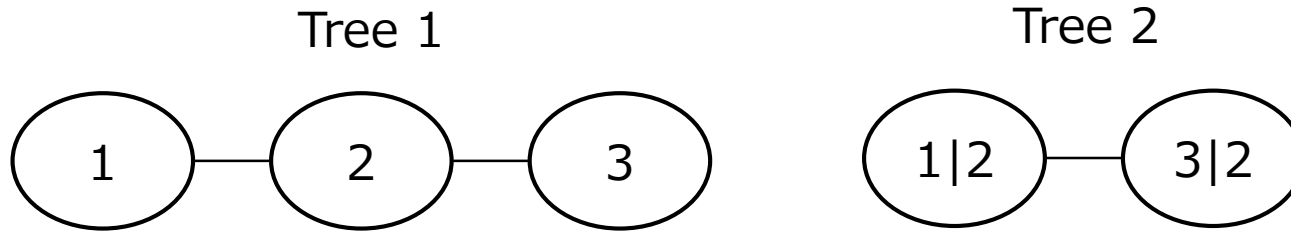
正規分布の分布関数を用いて、経験コピュラに変換。

$$\mathbf{U} = \left(\Phi(\mathbf{z}_1), \Phi(\mathbf{z}_2), \dots, \Phi(\mathbf{z}_{13}) \right)'$$

経験コピュラをもとにvine copulaで同時分布を推定。

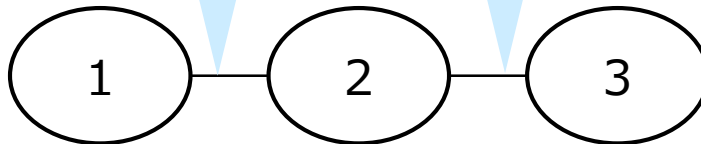
Vine copulaによるモデリングの手順

① tree構造を決定する。



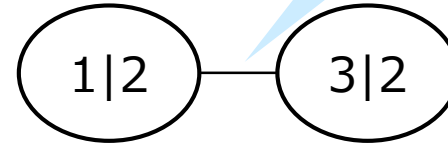
② 各辺に適用するコピュラファミリーを決める。

ガンベル



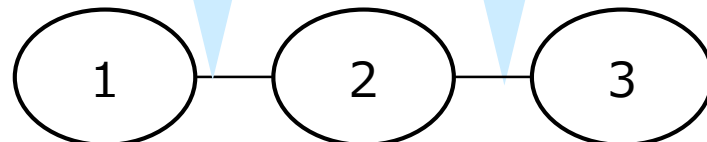
正規

Tree 2 クレイトン



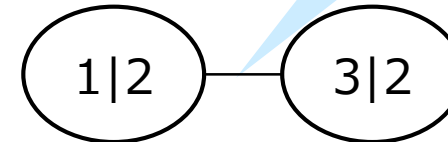
③ 各コピュラのパラメータを決める。

ガンベル($\alpha = 0.2$)



正規($\rho = 0.5$)

Tree 2 クレイトン($\alpha = 0.3$)



Vine copulaによるモデリング

- ▶ vine copulaの強みは、フレキシブルなモデリングだが、モデルの組み合わせが非常に多い。
- ▶ 変数の数 (n)を増加させていくと指数関数的に構造・コピュラの数が増加する。
- ▶ 数多くの構造の中からヴァイン構造を選択する必要があるため、ある程度自動化されたアルゴリズムが必要である。

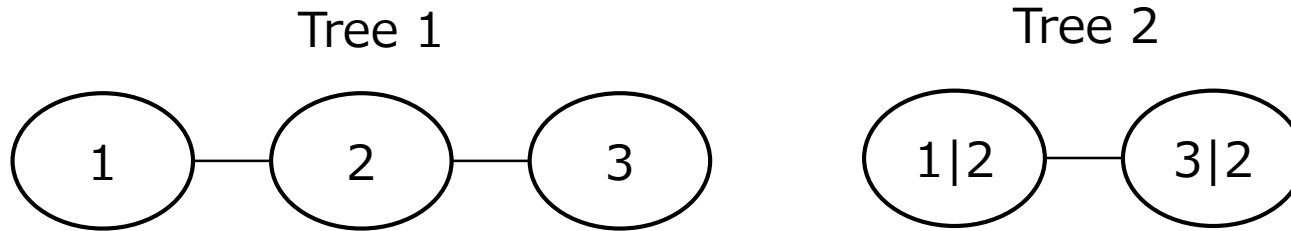
変数の数 (n)	tree 構造数	tree 構造×コピュラ数*
2	1	7
3	3	1,029
4	24	2,823,576
5	480	1.3559 e+11
6	23,040	1.0938 e+17
7	2,580,480	1.4413 e+24
8	660,602,880	3.0387 e+32
9	3.8051e+11	1.0090 e+42
10	4.8705e+14	5.2118 e+52

*copula の候補を 7 種類とした場合

- ▶ 14 Gruber (n.d.)等、既に多くの先行研究でアルゴリズムの研究が行われている。

本稿でのvine copulaのモデリング手順

① **ケンドールのタウの絶対値の和が最大**になる木構造を選択。

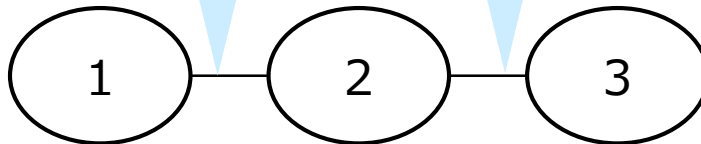


② 各辺のペアコピュラとして**AICが最小になるコピュラファミリー**を選択。

ガンベル

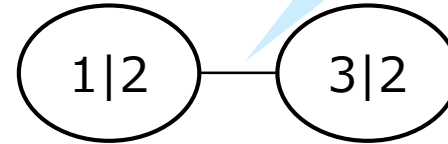
Tree 1

正規



Tree 2

クレイトン

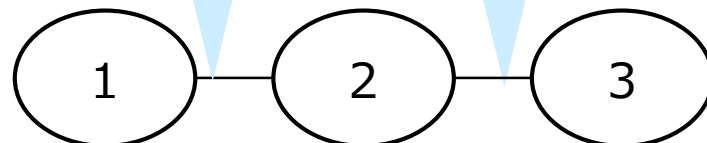


③ 各コピュラのパラメータを**最尤法**で推定する。

ガンベル($\alpha = 0.2$)

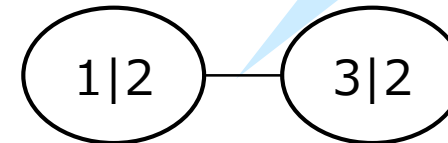
Tree 1

正規($\rho = 0.5$)



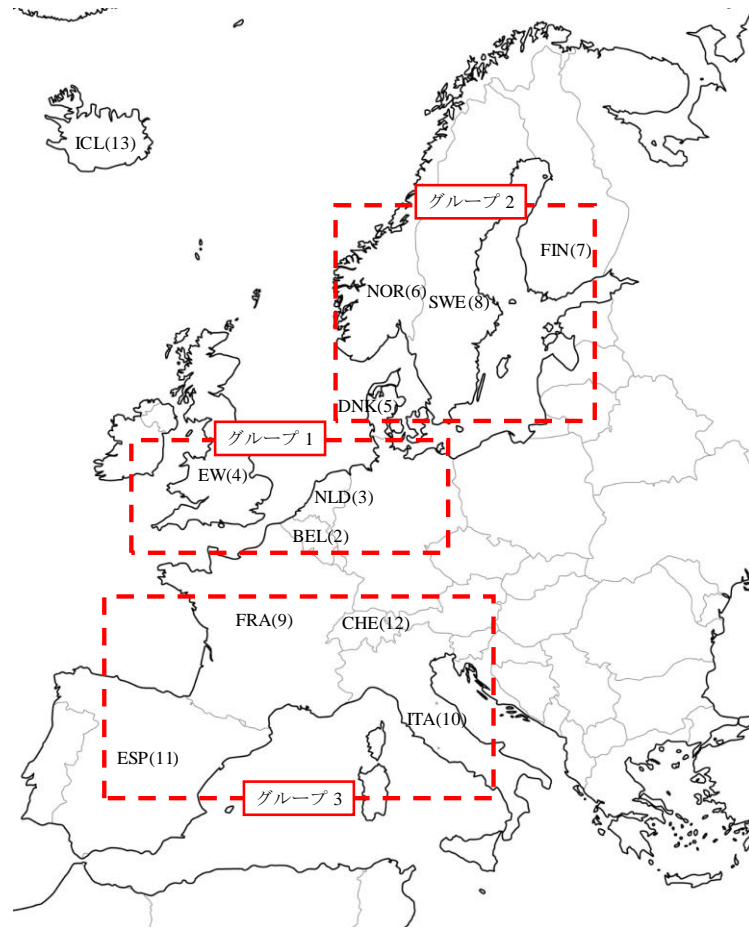
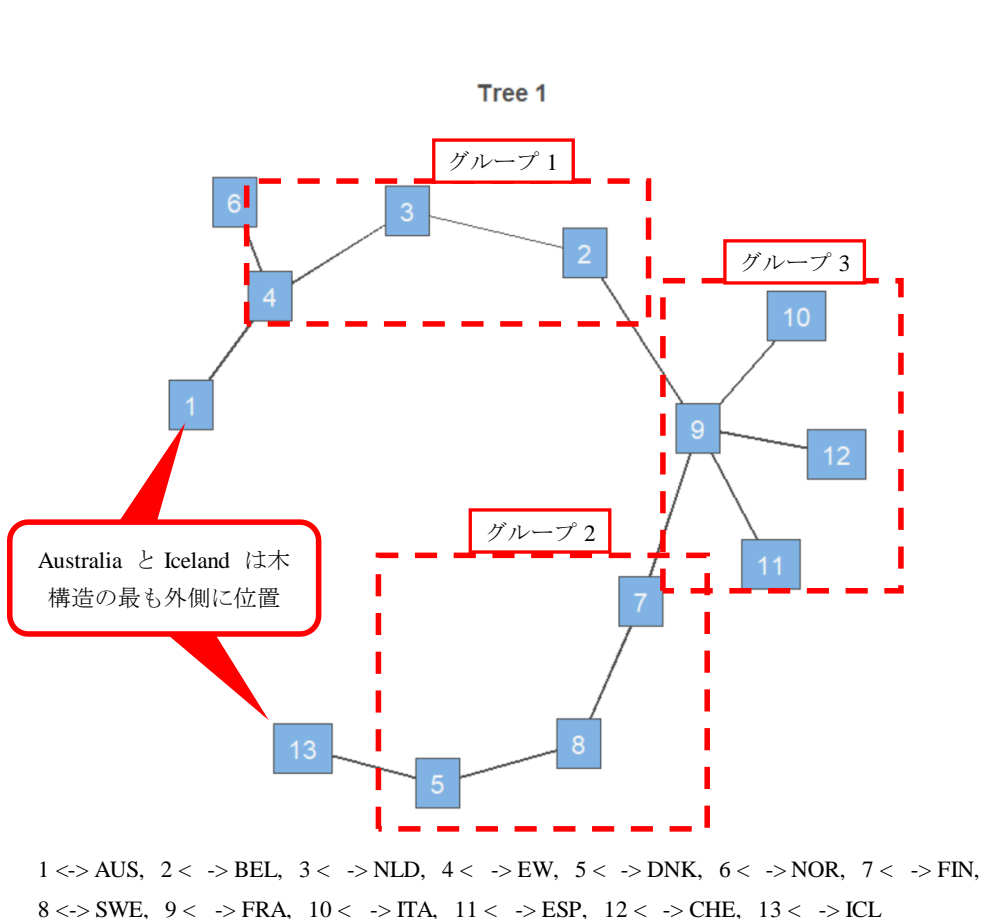
Tree 2

クレイトン($\alpha = 0.3$)



計算結果と考察

Vine copulaによる従属構造（第1階層）



得られた木構造を、地理的な位置関係に基づきグループ1～3に分けた。
実際の地理的な位置関係に極めて近い構造となっており、直感に近いモデルとなっている。

コピュラファミリーとそのパラメータ（第一階層）

変数間	コピュラ	パラメータ 1	パラメータ 2*1	Kendall's τ	λ_U *2	λ_L *2
4-1	逆グンベル	1.10	-	0.09	-	0.12
4-6	フランク	4.05	-	0.39	-	-
3-4	t	0.70	6.96	0.49	0.27	0.27
2-3	フランク	9.75	-	0.66	-	-
9-11	フランク	4.35	-	0.41	-	-
9-2	フランク	8.65	-	0.63	-	-
5-13	フランク	1.62	-	0.18	-	-
8-5	正規	0.65	-	0.45	-	-
7-8	フランク	4.94	-	0.45	-	-
9-7	t	0.64	30.00	0.44	0.01	0.01
9-10	逆グンベル	2.29	-	0.56	-	0.65
12-9	フランク	8.13	-	0.61	-	-

*1 パラメータ 2 は t コピュラの自由度

*2 λ_U は上裾従属係数、 λ_L は下裾従属係数

vine copulaを用いることでそれぞれの変数間で異なるコピュラを適用することができ、変数間ごとに異なる従属関係、特に裾の従属関係を個別に表現することができる。

代表的なコピュラとのパフォーマンス比較

- ▶ すべての指標において、vine copulaが他のコピュラを上回っている。
- ▶ フランクコピュラとクレイトンコピュラはパラメータを1つしか持たず、変数のペア毎に従属の度合いを個別に設定することができないため当てはまりが悪いと考えられる。
- ▶ 正規コピュラとtコピュラは、変数のペア毎に異なる相関係数を設定できるものの、線形かつ対称の従属関係しか表現することができず、vine copulaよりもパフォーマンスが劣っていると考えられる。

	AIC ^{*1}	BIC ^{*1}	対数尤度 ^{*2}
vine copula	-682.1	-548.3	395.1
正規コピュラ	-480.9	-287.6	318.4
t コピュラ	-562.7	-367.0	360.3
フランクコピュラ	-205.5	-203.0	103.7
クレイトンコピュラ	-215.6	-213.1	108.8

*1 小さい方が良い

*2 大きい方が良い

モンテカルロシミュレーション

モンテカルロシミュレーションの前提

- ▶ 65歳から90歳の生存率の加重平均値を表す指標として以下を用いる。

$${}_{90}S_{65} = \sum_{j=1}^n w_j \prod_{x=65}^{90} (1 - q_x^j)$$

- ▶ vine copula、正規コピュラ、tコピュラに対して、2010年（データの最終年度である2009年の1年後）の ${}_{90}S_{65}$ の値を2,000通りシミュレーション。
- ▶ これらのコピュラモデルがストレス時の挙動の予測にどのように異なるかを考察する。
- ▶ シミュレーションは地域的な分散がある場合と、分散が十分になされていない場合の影響を比較するために以下の2つのパターンを比較した。
 - ①すべての国に均等にビジネスが分散している場合
 - ②イタリアとフランスのみにビジネスがある場合

①すべての国に均等にビジネスが分散している場合

- ▶ 通常のリスク状況下である5%tileおよび95%tileはvine copula、正規コピュラ、tコピュラで大きな差の無い結果になっている。(赤枠部分)
- ▶ 一方、長寿リスク方向のテイルである99.5%tileでは、正規コピュラがもっともリスクが小さく、続いてvine copula、tコピュラの順でリスクが大きくなっている。(青枠部分)
- ▶ 正規コピュラは裾従属性が無いため、テイルにおけるリスクを小さく評価。
- ▶ vine copulaは裾従属性のある変数のペアと無い変数のペアで構成。
- ▶ tコピュラは分布の裾において全ての変数の従属性が高まる性質。

tコピュラ > vine copula > 正規コピュラ

大きな差のない結果

	99.5%tile	95%tile	50%tile	5%tile	0.5%tile
vine copula	1.169	1.087	1.000	0.917	0.853
正規コピュラ	1.161	1.086	1.000	0.921	0.848
tコピュラ	1.186	1.084	1.000	0.923	0.847

②イタリアとフランスのみにビジネスがある場合

- それぞれのコピュラはイギリスとフランスの間の従属関係を以下のように表現。

tコピュラ	vine copula	正規コピュラ
全変数を一律の裾従属性で評価	下裾従属性のある逆グンベルコピュラで評価	裾従属性なし

- シミュレーション結果は、①すべての国に均等にビジネスが分散している場合と比較して、99.5%tileのvine copulaとtコピュラの値が逆転している。
- 全体としてはtコピュラの方がテイルのリスクが大きい場合においても、特定の変数間のみに着目した場合にはテイルのリスクがやや低めに出てしまうことがある。
- 実際のビジネスでは、国家間でビジネスが均等に分布していることはまれであり、一部の地域に集中している可能性があり、リスクを過小評価してしまう可能性。

	99.5%tile	95%tile	50%tile	5%tile	0.5%tile
vine copula	1.293	1.138	1.000	0.865	0.756
正規コピュラ	1.282	1.138	1.000	0.868	0.747
tコピュラ	1.291	1.138	1.000	0.865	0.740

まとめと今後の課題

まとめ

<行ったこと>

- ▶ 欧州の12カ国にオーストラリアを加えた計13カ国の死亡率を、それぞれLee-Carterモデルで表現。
- ▶ 各国の死亡率の改善パラメータの同時分布をvine copulaでモデル化。
- ▶ 得られたモデルと代表的なコピュラのモデルとの比較。

<結果>

- ▶ 実際の地理的な位置関係にかなり近い従属構造を持つモデルが得られた。
- ▶ 得られたモデルは、代表的なコピュラを用いたモデルと比較し、AIC、BIC、対数尤度について優れたパフォーマンスを示した。
- ▶ モンテカルロシミュレーションの結果、正規コピュラやtコピュラを用いたモデルでは、特定の国家間のみテイルでの依存関係が高まるような従属関係を表現できないことを示した。

今後の課題

＜使用するデータ期間＞

- ▶ 1921年から2009年までの超長期のデータを用いている。
- ▶ しかし、戦前と戦後では人口動態が大きく異なるといわれているため、戦後のデータを使用した場合も同様の結果が得られるか考察が必要

＜ヨーロッパ以外のデータへの応用＞

- ▶ 例えばアジアのデータに本手法を適用し、同じように地理的な従属関係を再現するようなモデルが示せるかどうかを検証する必要がある。

＜第2階層以下のコピュラに関する考察＞

- ▶ 第2階層以下のコピュラ的作用がvine copulaの大きな利点。
- ▶ 第1階層の構造のみに着目した考察を行ったが、第2階層以下が従属関係に与える影響への考察が必要。