

中央大学博士論文

オペレーショナルリスク損失金額データの
分析手法の提案

鈴木晋史

Shinji Suzuki

博士（工学）

中央大学大学院

理工学研究科

経営システム工学専攻

令和4年度

2023年3月

目次

1. 序章	4
1.1. 背景	4
1.2. 本論文の目的	7
1.3. 本論文の構成	8
1章参考文献	8
2. 損失金額分布, 依存構造のモデル化とパラメータ推定手法の提案	10
2.1. はじめに	10
2.2. モデル	17
2.3. ベイジアン MCMC 法アプローチ	19
2.4. 数値実験	25
2.4.1. 数値実験におけるデータ作成	25
2.4.2. MCMC 法によるパラメータ推定結果	28
2.4.3. モデルの選択とリスク量の比較	34
2.5. 2章のまとめと今後の課題	36
Appendix 2. 用語の説明 (信用区間, NIF, CD)	37
2章参考文献	38
3. 離散型データにおける D-vine t コピュラで構成されたモデルのモデル選択	42
3.1. はじめに	42
3.2. モデル	44
3.3. ベイジアン MCMC 法, RJMCMC 法アプローチ	45
3.3.1. 潜在変数 U の事後分布, 提案分布, 採択確率の導出	45
3.3.2. RJMCMC 法による独立コピュラの検出とパラメータ推定	45
3.4. 数値実験	49
3.4.1. 5次元 D-vine t コピュラとモデルベクトル	49
3.4.2. 数値実験事例とその結果	50
3.5. 3章のまとめと今後の課題	57
Appendix 3. RJMCMC 法による遷移アルゴリズム	58
3章参考文献	59
4. 結論	62
4.1. 本研究のまとめ	62
4.2. 今後の研究課題	63
謝辞	63

目次

☒ 2-1. Parameters sampling flowchart by MH within Gibbs.	20
☒ 2-2. An example is given for the case where the number of bins is two. The suffix j is omitted for the measurement unit, bin number h for π and the data number i for x, u are written. And the X -coordinate is set to 0,1,2 for simplicity.	21
☒ 2-3. Scatter plot of $u_{i,j}$	26
☒ 2-4. Cumulative probability of bin of observation data $x_{i,j}$	27
☒ 2-5. Scatter plot of observed data $x_{i,j}$	28
☒ 2-6. Sample path, posterior probability density, and sample autocorrelation for bin parameters.	32
☒ 2-7. Sample path, posterior probability density, and sample autocorrelation for D-vine t copula parameters.	34
☒ 3-1. Scatter plot of $u_{i,j}$	51
☒ 3-2. Sample path, posterior probability density, and sample autocorrelation for D-vine t copula parameters.	53
☒ 3-3. Model switching path.	55
☒ 3-4. No. 960 Posterior probability density.	56
☒ 3-5. No. 842 Posterior probability density.	56
☒ 3-6. No. 63 Posterior probability density.	57

表目次

表 2-1. Results of a questionnaire on how to create scenarios for major domestic and foreign financial institutions.....	11
表 2-2. Basic statistics of observation data $x_{i,j}$	27
表 2-3. Top 10 bins with the largest divergence ratio.....	29
表 2-4. Estimation results of bin parameters	30
表 2-5. Acceptance rate of latent variable $U_{i,j}$	32
表 2-6. Hyperparameter.	32
表 2-7. Estimation results of D-vine t copula parameters.	33
表 2-8. Comparison of the percentile between different discrete distribution. ...	34
表 2-9. Parameter estimation results.	35
表 2-10. Amount of risk for probability distribution.	36
表 3-1. Acceptance rate of latent variable $U_{i,j}$	52
表 3-2. Hyperparameter.	52
表 3-3. Estimation results of D-vine t copula parameters.	52
表 3-4. Acceptance rate of latent variable $U_{i,j}$	54
表 3-5. Posterior model probability.....	54
表 3-6. Acceptance rate of latent variable $U_{i,j}$	54
表 3-7. Posterior model probability.....	54
表 3-8. Acceptance rate of latent variable $U_{i,j}$	54
表 3-9. Posterior model probability.....	55
表 3-10. No. 960 Estimation results of D-vine t copula parameters.	55
表 3-11. No. 842 Estimation results of D-vine t copula parameters.	56
表 3-12. No. 63 Estimation results of D-vine t copula parameters.	57

1. 序章

1.1. 背景

金融のグローバル化や国際競争が加速していく中で、金融システムの安定、国際競争での公平性の維持、開示情報の透明性の確保を目的として、バーゼル銀行監督委員会（BCBS）¹が国際的に活動する金融機関に対して各種規制を課している。その代表例として、自己資本比率規制が挙げられ、リスク対象は信用リスク、市場リスク、オペレーショナルリスク（オペリスク）である。自己資本比率規制は、銀行の自己資本を分子、リスクの大きさを分母とする比率（自己資本比率）が国際的に活動する国際統一基準金融機関には8%以上²であることを求めており（海外拠点を持たない国内基準金融機関は4%）、日本では1993年3月末から適用された。

本研究では、市場リスク、信用リスク、オペリスクのなかで、オペリスクに焦点を当てる。オペリスクを選択した理由は、1980年代半ば以降、オペリスクにかかわる巨額損失事件が再三発生し³銀行経営者のオペリスク管理の重要性が認識されたことによる。オペリスクはBCBS[2]において、『内部プロセス・人・システムが不適切であること若しくは機能しないこと、または外生的事象が生起することから生じる損失に係るリスク』と定義され、金融機関では狭義には「事務リスク」「システムリスク」「法的リスク」、広義には「市場リスク、信用リスク、流動性リスク以外のリスク」と多様に解釈されている。日本銀行[9]では、オペリスクを顕在化させる事例として、現金事故、横領、不正融資、マネーロンダリング、インサイダー取引、市場運用の損失隠し、コンピュータ・システム障害、自然災害、テロ行為、ハラスメントによる被害等を挙げている。

金融庁[6]によれば、自己資本比率算出式における「オペリスクによって発生しうる損失推定額（オペリスク量⁴）」の推定手法一つとして「先進的計測手法」が

¹金融機関を対象とした国際的なルールを協議・決定し、継続的な協力を行うためにG10（主要10カ国）諸国の中央銀行総裁会議での合意によって創設された機関のこと。日本を含む28の国と地域の中央銀行、銀行監督当局で構成し、スイスのバーゼルに本部を置く国際決済銀行（BIS）のなかに事務局を設けている。

²鎌倉[5]によれば、自己資本比率＝

$$\frac{\text{自己資本}}{\text{信用リスクに応じて重み付けした資産額} + \text{市場リスクによって発生しうる損失推定額} \times 12.5 + \text{オペリスクによって発生しうる損失推定額} \times 12.5} \geq 8\%$$

³西山[10,11,12]によれば、例えば、巨額不正事件としては、1985年に英国の証券会社では損失金額が8億ポンド以上で清算に追い込まれ、1995年邦銀のNY支店では11億ドル以上、同年日本の商社では2850億円、1997年に英国の銀行での9000万ポンドのなどの事例が知られている。

⁴「オペリスク相当額」と呼ばれる場合もある。

知られている。日本銀行[8]によれば、「先進的計測手法」の中で「損失分布手法」が最も一般的であり、オペリスク量にはオペリスクの損失金額分布の Value at Risk(VaR)が採用されている⁵。また、オペリスク量の VaR は自己資本比率規制におけるオペリスク量として使用するだけでなく、内部的なリスク管理ではオペリスク VaR と位置づけ、リスク削減策の策定等リスク管理実務で標準的に使用されるリスク指標である。

金融庁[6]によれば、オペリスク量の推定において、おもな収集すべき損失金額データは、金融機関内部で顕在化した「内部損失金額データ」、専門的な知識や経験にもとづいて損失金額とその頻度が推定される潜在的な「シナリオ損失金額データ」である。⁶「シナリオ損失金額データ」は低頻度高額損失の内部損失金額データを補完し、まだ損失として顕現化していないが、現場のリスク認識に即したリスクの状況に関する情報を計量化に取り込むことを目的としている。オペリスク量の推定に際し、シナリオ損失金額データで内部損失金額データを補いモデルに投入する場合、低頻度高額損失の多くをシナリオ損失金額データに依存すれば、シナリオ損失データがリスク量の太宗を占めることとなり、シナリオ損失金額データの設定次第でリスク量が大きく変化することから、計量化への取り込み方法は非常に重要となる。

内部損失金額データとシナリオ損失金額データで構成されるオペリスクの損失分布は、連続型データと離散型データが混在した状況で推定することが求められる。内部損失金額データは、良く知られた分布形でモデル化され連続型データとして取り扱われる事例が多く、一方で、BCBS[3]ではシナリオ損失金額データが損失金額と頻度を見積もる際に、1点で推定（あるシナリオに対し、1組の損失金額とそれに対応する頻度を想定）され離散型データとして取り扱われる事例が多いことを紹介している。例えば、100年に一度10億円の損失が発生するシナリオ損失では、 X を損失金額の確率変数とすると、 $\Pr(X = 10 \text{ 億円}) = 1/100$ と1組の損失金額とそれに対応する頻度を設定される場合が多い。

日本銀行[8]によれば、オペリスク損失金額データの特徴として、その頻度と金額から、2種類の類型に分けられる。一つは事務ミスなど「比較的高い頻度で発生する小規模なトラブル（高頻度低額損失）」ともう一つはシナリオによる天災、テロ、おおがかりな不正行為など、「滅多に発生しないが発生すると影響の大きいトラブル（低頻度高額損失）」である。この特徴から、オペリスク1件あたりの損失金額の分布をグラフ化すると「裾の厚い」形状になることが知られている。

オペレーショナルリスク管理高度化勉強会（オペリスク勉強会）[13]によれ

⁵オペリスクのリスク量の計測手法の詳細については、2.1節脚注11を参照のこと。

⁶金融庁[6]では内部損失金額データ、シナリオ損失金額データの他、外部損失金額データを収集しなければならない。詳細は2.1節を参照のこと。

ば、オペリスクは、金融機関の各種の業務や損失事由など多種多様なリスク要素により構成されている。リスク計量化に際しては、例えば、損失金額の分布形、損失事象間の独立性など、前提を置くことが必要となるが、多種多様なリスク要素を含むオペリスクを共通の前提の下で計量化することは容易ではない。そこで、オペリスク損失事象を、例えば、各種業務（ビジネスライン）と各種損失事由（イベントタイプ）の組合せからグループ（計量単位⁷）に分け、各グループに共通する前提の下で、それぞれのグループを計量化することを考える。従って、社内全体のオペリスク量を推定する場合には、計量単位を周辺分布（推定されるべき損失金額分布）とする多次元の損失分布から推定する必要がある。

オペリスク勉強会[13]によれば、オペリスク管理のリスク計量化手法について業界標準的な考え方に収斂しておらず、オペリスク管理の高度化に向け、下記の2点について取組みや議論が望まれている。

第1に、オペリスク損失金額データには高頻度低額損失と低頻度高額損失とあり、この両者を的確に捉える分布形を選定することが難しく、損失分布形の選択、およびパラメータの推定手法次第で、リスク量が大きく異なる可能性がある。このような特徴あるデータにおいても、安定したリスク量が推定できる分布形、推定手法が望まれる。

第2に、計量単位の設定⁸、その計量単位間の依存関係をどう定式化するかが、リスク計量結果に大きな影響を与える。計量単位の依存関係の定式化には、依存関係の程度が、例えば損失金額によって大きく異なる可能性（具体的には、分布のボディ部分での依存関係は小さい一方、テール部分の依存関係が非常に高いケース）を考慮する必要がある。リスク量には99%、99.9%といった高い信頼水準（損失方向に大きな水準）のVaRを推定することから、分布全体をならした依存関係を表す計数（例えば、一つの数値（線形相関）であらわした相関係数）を用いることが適切でないケースも考えられる。また、Artzer et al.[1]によれば、「損益額分布の形状によっては、VaRは劣加法性の公理を満たさない」など金融リスクを捉える際の理論的な問題も知られている。一方で、オペリスク勉強会[13]によれば、「多くの金融機関が社内全体のリスク量を計量単位毎のリスク量を単純合計して推定している事例が多い」ことが知られており、これは、社内全体の全計量単位をまとめたリスク量が計量単位毎の単純合計のリスク量を上回るケース

⁷「計量単位」は多変量分布における「周辺分布」に相当する。本研究では「計量単位」と「周辺分布」の用語を併用している。やや厳密さに欠けるが、オペリスクを意識している場合の損失金額の周辺分布を「計量単位」、一般的な議論の場合「周辺分布」を使用している。

⁸計量単位の設定方法については、BCBS[2]では「イベントタイプ×ビジネスラインの組合せ」の分類が当局報告様式として規定されており、本研究では、計量単位の依存関係のモデル化、推定手法について取り上げる。

も想定され、過小評価していることを示唆している。オペリスク量を過小評価することを回避するためには計量単位の依存構造を反映したうえで、各計量単位のリスク量を合算することが望まれる。

多変量金融資産のポートフォリオの評価、リスク管理など多くの研究において周辺分布間の依存構造のモデル化にコピュラが用いられている。コピュラには従来の多変量コピュラと Vine コピュラが知られている。Vine コピュラは従来の多変量コピュラではモデル化できない異なるペアの複雑な依存構造を表現できるが、一方でペアコピュラ項数が 2 次的に増加すること⁹が知られている。Vine コピュラは、Bedford and Cooke[4]によって提案された多変量コピュラで、依存構造を 2 変量コピュラ (ペアコピュラ) に分解したものである。前田[7]によれば、「2 変量の場合は変量間の依存構造に合わせて 2 変量コピュラをあてはめ、分布の裾のリスク量の合算が可能となるが、3 変量以上になると従来の一つの変量コピュラ (多変量 t コピュラなど) では分布の裾のリスク量の複雑な依存関係を捉えることが難しい」ことから、本研究ではペアコピュラである Vine コピュラを選択する。

周辺分布と Vine コピュラのパラメータの推定について、Smith and Khaled[14]では、離散型データの場合、同時分布の尤度関数項数は指数的に増加すること¹⁰、ならびに最適化の計算が複雑なることから、周辺分布とコピュラパラメータの最尤法による同時推定は現実的でないと指摘している。

以上から、Vine コピュラでモデル化する場合の計算負荷軽減手法、パラメータの同時推定手法等の課題についても検討が必要である。

1.2. 本論文の目的

本研究では、これらの課題に対して、内部損失金額データおよびシナリオ損失金額データを離散型データとして取り扱い、周辺分布を離散分布で近似すること、依存構造を Vine コピュラでモデル化することを提案している。

第 1 に、周辺分布および依存構造をモデル化して、両者のパラメータの同時推定手法を提案する。具体的には、下記の通りである。

- (1) オペリスク損失金額データの特徴を考慮して、内部損失金額データ、シナリオ損失金額データを離散型データとして取り扱い周辺分布を離散分布で近似する。
- (2) リスク量が VaR であることを考慮して、分布の裾のリスク量の複雑な依存関係を捉えるペアコピュラである D-vine t コピュラでモデル化する。

⁹ 「ペアコピュラ項数の 2 次的な増加」の説明については 3.1 節を参照のこと。

¹⁰ 「尤度関数項数の指数的な増加」の説明について 2.2 節(1)を参照のこと。

(3) 周辺分布とコピュラパラメータの同時推定手法について, Markov Chain Monte Carlo(MCMC)法を採用する.

(4) 提案したモデル, パラメータ推定手法の精度を模擬データから数値実験することで, その有用性を検証する.

第2に, 離散型データの場合には, 周辺分布を離散分布でモデル化, 依存構造をペアコピュラである Vine コピュラでモデル化し, パラメータの同時推定手法に MCMC 法を選択すると, ペアコピュラの項数が2次的に増加することから, その計算負荷が非常に高い. そこで, 離散分布と Vine コピュラのパラメータの同時推定の前段階として, Reversible Jump MCMC (RJMCMC) 法によりペアコピュラの組合せの中で, 独立コピュラ¹¹を探索する手法を提案する. これにより, 独立コピュラと判定されたペアコピュラは, パラメータ推定の対象外となり計算負荷軽減が期待できる. また, 提案した RJMCMC 法による独立コピュラの探索精度を模擬データから数値実験することにより, その有用性を検証する.

1.3. 本論文の構成

2章では, 鈴木[15]に則って, まず, オペリスクの損失金額データの特徴, 損失事象の計量単位の分類手法について詳細を述べる. 次に, 周辺分布のオペリスクの損失金額分布を離散分布で近似, 依存構造を D-vine t コピュラでモデル化し, 最後に, 模擬データを用いた数値実験で, MCMC 法により離散分布, D-vine t コピュラパラメータを同時推定し, 提案した手法の精度を検証し, その有用性を確認する.

3章では, 鈴木[16]に則って, まず, 離散型データの場合に, 依存構造を D-vine t コピュラでモデル化した事例で, パラメータ推定時の計算負荷軽減のための RJMCMC 法による独立コピュラの判定方法を提案し, 次に, 模擬データを用いた数値実験で, 提案した手法の精度を検証し, その有用性を確認する.

4章では, 本研究の総括を行い, 今後の展望を述べる.

1章参考文献

- [1]Artzner, P., F. Delbaen, J. M. Eber, and D. Heath, Coherent Measures of Risk, *Mathematical Finance*, **9**(3), (1999), 203-228.
- [2]バーゼル銀行監督委員会 (Basel Committee on Banking Supervision), *Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework*, 2004.
<https://www.bis.org/publ/bcbs107.htm>
- [3]バーゼル銀行監督委員会 (Basel Committee on Banking Supervision). *Results*

¹¹ 独立コピュラの説明については 3.3.2 項(1)を参照のこと.

from the 2008 Loss Data Collection Exercise for Operational Risk, 2009.

<http://www.bis.org/publ/bcbs160.htm>

- [4] Bedford, T. and Cooke, R. M., Vines — A new graphical model for dependent random variables, *The Annals of Statistics*, **30** (2002), 1031–1068.
- [5] 鎌倉治子, 新 BIS 規制とオペレーショナル・リスク, *レファレンス*, **632**, (2003), 55-67.
https://dl.ndl.go.jp/view/download/digidepo_999975_po_063203.pdf?contentNo=1
- [6] 金融庁, 銀行法第十四条の二の規定に基づき, 銀行がその保有する資産等に照らし自己資本の充実の状況が適当であるかどうかを判断するための基準, 告示第十九号, 2006.
http://www.fsa.go.jp/policy/basel_ii/01.pdf
- [7] 前田寿満, ヴァイン・コンピュータを用いた統合リスク管理, *金融研究*, **35(2)**, (2016), 37-66.
<https://www.imes.boj.or.jp/research/papers/japanese/kk35-2-2.pdf>
- [8] 日本銀行, オペレーショナル・リスク管理の高度化, 2005.
https://www.boj.or.jp/research/brp/ron_2005/fsk0507a2.pdf
- [9] 日本銀行, オペレーショナルリスク管理の現状と高度化への課題, 金融高度化セミナー「金融機関の経営管理の高度化—理論と実践」, 2014.
https://www.boj.or.jp/announcements/release_2014/data/rel140828a9.pdf
- [10] 西山都, 連載オペレーショナル・リスク入門の入門(1) なぜいま注目されているのか, *週刊金融財政事情*, 4月23日号(2001), 52.
- [11] 西山都, 連載オペレーショナル・リスク入門の入門(2) オペレーショナル・リスクの定義, *週刊金融財政事情*, 5月21日号(2001), 52.
- [12] 西山都, 連載オペレーショナル・リスク入門の入門(3) なぜ管理が必要なのか, *週刊金融財政事情*, 6月4日号(2001), 52.
- [13] オペレーショナルリスク管理高度化勉強会, オペレーショナルリスク管理の更なる高度化に関する議論(1), (2), オペレーショナルリスク管理高度化勉強会, 2006.
http://www.boj.or.jp/research/other_release/fsc0606a.pdf
http://www.boj.or.jp/research/other_release/fsc0608e.pdf
- [14] Smith, M. S. and Khaled M.A. , Estimation of Copula Models with Discrete Margins via Bayesian Data Augmentation, *Journal of the American Statistical Association*, **107** (2012), 290-303.
- [15] 鈴木晋史, オペレーショナルリスク損失金額データ分析手法の提案, *日本応用数学会論文誌*, **31**(2021), 244-277.
- [16] 鈴木晋史, 離散型データにおける D-vine t コピュラで構成されたモデルのモデル選択とパラメータ推定手法の提案, *日本応用数学会論文誌*, **32**(2022), 20-45.

2. 損失金額分布， 依存構造のモデル化とパラメータ推定手法の提案

2.1. はじめに

2004年6月に公表されたBCBSが定める「自己資本の測定と基準に関する国際的統一化：改訂された枠組」BCBS[4]基準では、オペリスクが従来の市場リスクと信用リスクに加えて、銀行のリスク量の計算対象となった。

金融庁[22]によれば、オペリスク相当額(リスク量)の算出*¹²において、3種類の損失金額データを収集しなければならない。第1に銀行の内部で生じたオペリスク損失に関する情報である内部損失金額データ、第2に銀行の外部から収集したオペリスク損失に関する情報であって、銀行におけるオペリスクの管理に資する外部損失金額データ、第3に重大なオペリスク損失の額および発生頻度について、専門的な知識および経験ならびにオペリスクに関する情報に基づいて推計するシナリオ分析によるシナリオ損失金額データである。また、同告示では、内部損失金額データの要件は、銀行の全ての業務における一定の閾値以上のオペリスク損失のデータが全て含まれていること、その閾値は、100万円以下で銀行が設定した金額とすることと定められている。*¹³

内部、外部損失*¹⁴については、顕在化した損失金額データを収集する。一方、シナリオ損失は潜在的な損失事象であり、「顕現化の頻度」と「影響度の大きさ」を反映したデータを作成しなければならない。その方法については、内外金融業界標準的な考え方への収斂が見られないのが現状であるが、BCBS[5]においては、内外主要金融機関からのアンケート結果が取り纏められており、表2-1のように3種類のシナリオ作成方法の中で、内外主要金融機関で半数以上が、個別シナリオアプローチ、すなわち、あるシナリオに対し、1組の損失金額とそれに対応する頻度を想定している。また、連続分布を仮定する場合においては、インターバルアプローチなどで、パラメトリックモデルの最尤法によるパラメータ推定に

*¹² オペリスク相当額の計測手法は3種類に分類され、うち2種類は財務諸表の損益計算書から算出され、もう1種類は各種損失金額データを使用し、その分布形を推定、オペリスクの場合は、上側99.9%分位点(バリュー・アット・リスク(VaR))がリスク量と定義される事例が多い。なお、本章では各種損失金額データの周辺分布と依存構造の定式化と推定手法を議論するもので、損失金額データの要件、収集方法などについては、これに従うが、オペリスク相当額の算出方法についての説明は省略する。

*¹³ 閾値を調整するモデリングの煩雑さなど考慮し、多くの金融機関では、1円以上の全ての損失金額データを対象としている。

*¹⁴ 他の金融機関の損失事象から損失金額データを作成する場合、その所在国、規模、業態などによって、金額等を調整する場合があるが、その手法についての説明は省略する。

において最適化が上手くいかない場合には、シナリオに離散分布を仮定することが言及されている。

表 2-1.BCBS [5] Results of a questionnaire on how to create scenarios for major domestic and foreign financial institutions.

Item/Method	Individual Scenario Approach	Interval Approach	Percentile Approach
Summary	Output from the individual scenario approach consists of one severity estimate linked with a probability of occurrence(frequence)for each scenario.	The interval approach output consists of frequency estimates for series of distinct severity ranges.	The percentile approach results in severity estimates for specified percentiles of a loss distribution.
Adoption ratio	52%(=34banks / 65banks)	20%(=13banks / 65banks)	28%(=18banks / 65banks)
Distribution Assumption, Parameter Estimation Method etc	The probability of occurrence and one estimated loss size are linked. $\Pr(X = x) = p$	It is mentioned that the parameters are estimated by the maximum likelihood method assuming a lognormal distribution for the intervals of the various scenarios, and that a discrete distribution is assumed when an optimal solution cannot be obtained. $\Pr(x_1 < X < x_2) = p$	Quantile Matching Method $\Pr(X > x) = p$

表 2-1 は BCBS [5]VI Scenario Analysis Results の概略を記述したものである。表中の数式表示は、筆者の記述による。ここで、確率変数 X は損失金額、 $x, x_1, x_2 (x_1 < x_2)$ はその実現値、 p はオペリスク損失事象の発生確率である。

日本銀行[36]によれば、これらの損失金額データから構成されるオペリスクの事例や特徴は、その頻度と金額から、2種類の類型に分けられ、一つは「比較的高い頻度で発生する小規模なトラブル(高頻度低額損失)」であり、これには事務ミス(少額の送金ミス、窓口の誤払い等)、業務用端末の軽微なトラブル等が含まれる。もう一つは「滅多に発生しないが発生すると影響の大きいトラブル(低頻度高額損失)」であり、例えば、勘定系コンピュータ・システムの大規模障害、天災(大地震等)、テロ、大掛かりな不正行為等がこれに当たる。こうした特徴があるため、オペリスクの1件あたりの損失金額の分布をグラフ化すると、「裾の厚い」形状になることが知られている。なお、以降は、高頻度低額損失の特徴を有する事象が多い内部損失、リスク管理上重要な位置づけとなっている低頻度高額損失の特徴を有する事象が多いシナリオ損失の2種類の損失金額データで議論する。^{*15}

*15 外部損失金額データは大小いずれの損失金額の事象も存在する。損失分布推定手法に関して議論する本研究では2種類のタイプの損失金額データが存在することが重要な点であり、内部損失金額データ、シナリオ損失金額データの2種類の損失金額データで議論してもその本質は変わらない。

オペリスク勉強会[37]によれば、リスク計量化に際しては、例えば、損失金額の分布形、損失事象間の独立性等の仮定を置くことが必要となるが、オペリスクは多種多様なリスク要素により構成されており、共通の前提の下で計量化することは容易でない。そこで、オペリスク事象をグループ分けしたものを計量単位と呼び、各グループに共通する前提の下で、それぞれのリスク、相互の依存性を計測することを考える。この計量単位設定に留意すべきことは、同一計量単位内のデータの属性、原因が共通であり同一の分布に従い、1つのモデルで扱えること、同一計量単位で括られたデータが相互に独立していること、1つの計量単位内に十分な量のデータが確保できること、および、計量単位と実際のリスク管理の単位の整合性など、内部管理上において有用であることが挙げられる。また、その分類の切り口に関する内外主要金融機関の実務をみると、イベントタイプ(損失事象の種類)別、ビジネスライン(業務区分)別、イベントタイプ×ビジネスライン、原因別、リーガルエンティティ別、全データを一括に扱う等、多種多様であり、また、リスク計量結果に大きな影響を与える各計量単位に帰属する損失金額データの依存関係の設定、モデル化についても、実務において業界標準的な考え方へは収斂していない。

そこで、BCBS[4]では「イベントタイプ(7種類)とビジネスライン(8種類)*¹⁶の組合せ」に基づく56種類の分類が当局への報告用として規定されており、金融機関は自社の業態に合致したイベントタイプ、ビジネスタイプを選択する。日本銀行[36]によれば、5~7区分程度の計量単位を設定している事例が多く、本研究ではこれを計量単位と考える。また、計量単位の依存関係については、各計量単位に帰属する個々の損失金額データ間の相互の因果関係の共通性や連鎖性がある事例を想定しモデル化する。

オペリスクの研究分野では、筆者の知る限り、損失金額データ分布の分析においては、よく知られた連続確率分布を仮定している事例が殆どである。実務の観点からは、損失金額分布の選択とそのパラメータの推定手法により、また、計量単位間の依存関係をどう設定するかにより、リスク計量結果に大きな影響を与えることから、これらが研究対象となっているのかどうかに着目し整理する。

まず、依存性が考慮されていないが、複数の分布、推定手法を取り上げている事例として、三菱信託銀行[29]、森他[30]、小林他[21]、Fontnouvelle et al.[16]、Chapelle et al.[12]、Moscadelli[32]、Dutta and Perry[14]がある。

*¹⁶ イベントタイプには、「内部の不正」、「外部からの不正」、「労務慣行および職場の安全」、「顧客や商品および取引慣行」、「有形資産に対する損傷」、「事業活動の中断およびシステム障害」、「注文等の執行、送達およびプロセスの管理」の計7種類、ビジネスタイプには「リテールバンキング」、「コマースバンキング」、「決済業務」、「リテールブローカレッジ」、「トレーディングおよびセールス」、「コーポレートファイナンス」、「代理業務」、「資金運用」、「その他」の計8種類が存在する。

これらの研究では、指数分布、対数正規分布、ワイブル分布、ガンマ分布、一般化パレート分布(GPD)等を損失分布に仮定している。パラメータは、最尤法、モーメント法、最小二乗法等によって推定され、分布形の選択や推定手法によってリスク量が大きく異なることに言及している。極値理論に基づく閾値超過モデル(POT) *¹⁷の適用例についても紹介しており、その手法は、閾値を経験的に与え、パラメータを最尤法により推定している。なお、Dutta and Perry[14]ではPOTについて、他の手法に比べて突出して大きなリスク量が推定される場合があること、閾値の設定によりリスク量が大きく変動することを指摘している。POTにおいて、閾値とGPDパラメータの同時推定にMCMC法によるベイズ手法が用いられている例として、Behrens et al. [7], Nakatsuma[34]などが挙げられ、前者は、閾値未満のデータをガンマ分布に従うと仮定し、分析データとして、閾値未満をガンマ分布から、閾値以上をGPDからサンプリングしたデータを、また実証分析として株価収益率を損失金額データに見立て Nasdaq100*¹⁸のデータを使用している。後者は、閾値未満のデータを正規分布、ガンマ分布に従うと仮定し、金融機関から提供された内部損失金額データを使用しモデル選択を行っている。次に、計量単位間の依存性についてコピュラ以外でモデル化された事例として、Badescu et al. [3]では、頻度分布にパラメータがアーラン分布に従う多変量混合ポアソン分布を採用して計量単位の依存構造を反映させ、損失金額分布に混合アーラン分布を採用し、パラメータはEMアルゴリズムで推定している。実証分析については、Operational Riskdata eXchange Association (ORX)のデータを利用し、提案したモデルの適合性が高いことを示している。また、計量単位間の依存性のコピュラによるモデル化の事例として、Valle[42]では、周辺分布に指数分布、ガンマ分布、GPDを仮定し、依存構造を従来の多変量正規および多変量tコピュラでモデル化し、MCMC法によるベイズ手法でパラメータを推定している。Mittnik and Yener[28]は計量単位の損失金額を定数として、閾値モデルを潜在変数モデルとして、頻度を計数過程によってモデル化している。依存構造はコピュラでモデル化し潜在変数に反映させ、頻度のみの相関の強弱とリスク量の増減を分析している。数値実験では、完全相関でない場合が必ずしもリスク量の減少を意味するものでなく、リスク量は増加する場合があるので、相関関係の影響を評価することが重要であると指摘している。Brechmann et al. [11]は計量単位の依存構造のモデル化において、損失金額データが存在しない場合を考慮して、zero-inflated dependence model (ゼロ過剰依存モデル)を提案している。また、イタリアの

*¹⁷ 極値理論、POT手法については森本[31]に詳しい。

*¹⁸ 米国のナスダック市場に上場する金融セクター以外の銘柄のうち、流動性が高く時価総額の大きい上位100銘柄で構成する株価指数であり、時価総額加重方式によって算出される。

金融機関のオペリスクデータベース (DIPO) により実証分析を実施し、ペアにおける不均一な依存性，テール依存性，解釈可能なパラメータ，計算の容易さの観点から，individual Student's *t* コピュラを最適なコピュラとして挙げている．頻度分布については，QQ プロット，適合度検定から負の二項分布を，損失分布には対数正規分布を選択している．Mittnik et al. [27] は計量単位の損失金額の依存構造をパラメトリック，ノンパラメトリックコピュラでモデル化した．また，DIPO データによる実証分析で，まず，計量単位全体のリスク量を推定する場合，VaR は劣加法性の公理を満たさないことから各計量単位の相関を考慮せずリスク量を単純合計すると過小評価する場合があります，次に線形相関や順位相関は安定的でなく，さらに，モンテカルロシミュレーションの試行回数が十分でない場合，VaR はばらついて，安定性に問題があることを指摘し，理論的な VaR の上限を推定している．Brechmann et al. [10] は計量単位の損失発生件数の依存構造を Vine コピュラによってモデル化している．件数は離散型データであり，その対応として Panagiotelis [38] の提案した手法により Vine コピュラを差分形式で展開している．また，DIPO データを使用した実証分析において，QQ プロット，適合度検定から損失金額に対数正規分布を，頻度に負の二項分布を，依存構造については，AIC によって正規，クレイトン，ガンベル，フランクペアコピュラの中からの Vine コピュラの組合せを選択している．

先行研究の問題点として，次の 2 点が挙げられる．

第 1 の点は，分布形の選択，推定手法によってリスク量が大きく異なることである．内部損失とシナリオ損失金額データの両者を考慮した損失分布を推定する場合，1 件あたり 1 円から数千億円におよぶ場合がある損失金額範囲のデータから構成され*¹⁹，よく知られた分布を適用した先行研究では，分布形の選択，推定手法によってリスク量が大きく異なることに言及している．また，POT の事例においても閾値を経験的に与え，パラメータを最尤法により推定する場合，他の手法に比べて突出して大きなリスク量が推定される場合があること，閾値の設定によりリスク量が大きく変動することが指摘されている．オペリスク勉強会 [37] によれば，このように，分布形の選択，およびパラメータの推定手法次第で，リスク量が大きく異なる可能性があり，それらの選択は重要な問題であるにもかかわらず，業界標準的な考え方に収斂していない理由として，「オペリスク損失金額データには高頻度低額損失と低頻度高額損失とあり，この両者を的確に捉える分布形を選定することが難しいこと，仮に標本数が多くても，特定のビジネスラインや，イベントタイプ，エンティティに偏っていたり，リスク量の大宗を占めるシナリオ損失金額データが乏しかったりすること」を挙げている．

第 2 の点は，依存構造が従来の多変量コピュラによりモデル化されていること

*¹⁹ 金融機関の規模，業務によって，損失金額データの範囲は多種多様である．

である。オペリスク研究分野における依存性について、先行研究では従来の正規、 t コピュラでモデル化しているが、全てのペアの順位相関関係を 1 種類のパラメータで調整することになり、複雑な依存構造を表現できない。また、オペリスクを考慮した場合、分布の中心部だけでなく、計量単位において巨額損失の同時発生の可能性も考慮すべきであり、その場合、各周辺分布の複雑な裾依存性もモデル化する必要があるが、正規コピュラの場合、上側、下側のいずれでも漸近独立であり同時的な大規模損失事象を十分に表現できないことを示唆している。

オペリスク以外の分野ではあるが、まず、離散型データとコピュラの関係について、Smith and Khaled[39]は、アマゾンの閲覧回数とその購入金額の依存構造を D-vine t コピュラによりモデル化、周辺分布にパラメータが 1 個のベルヌイ分布、負の 2 項分布を仮定している。また、Nelsen[35]のコピュラの差分による確率関数の定義を引用して、以下の点を指摘している。第 1 に離散型データをコピュラでモデル化した場合、 n をデータ数、 m をデータの次元とすると、その同時分布の尤度関数項数は $n2^m$ *²⁰ と指数的に増加することから、データ数が多く、高次元の場合には膨大となることである。第 2 に高次元でない場合でも、最尤法において最適化の計算が複雑となることから、周辺分布とコピュラのパラメータの同時推定は現実的でないことである。そこで、潜在変数をデータ拡大法で生成する MCMC 法を提案している。Genest and Nešlehová[17]では、事例を挙げた理論的な解説で、離散分布に従う周辺分布の依存性をコピュラでモデル化する際に、連続型データとして疑似尤度によりパラメータを推定すると、両者の推定値は大きく乖離する場合があることを指摘している。次に、性質の異なる多変量の依存関係をモデル化するために、Vine コピュラ*²¹を用いた研究が行われている。Vine コピュラは、Bedford and Cooke[6]によって提案された多変量コピュラであり、その依存構造を 2 変量コピュラ（ペアコピュラ）に分解したもので、R-vine、その特殊ケースである C-vine、D-vine が知られている。これらの主な違いは、ペアコピュラの結合形態の違いであり、C-vine、D-vine はコピュラの構造が簡単に記述でき、また、その特徴について、岩永[20]では「C-vine は特定の変数が各変数間の相互作用の中心となることが明らかな場合にメリットがあり」、塚原[40]では「D-vine から構成される分布は各変数が同等に扱われている」と紹介されている。最後に、従来の多変量コピュラと D-vine コピュラの比較について、Aas and Berg[1]、Fischer et al.[15]は、2 変量 t コピュラを使った D-vine 型コピュラは、例えば階層構造多変量アルキメディアンや多変量 t コピュラよりも多変量金融データへの適合度*²²が高いことを報告している。

*²⁰ 詳細は 2.2 節(1)を参照のこと。

*²¹ コピュラ全般については、Nelsen[35]、戸坂・吉羽[41]に、Vine 型については、Kurowicka and Joe[24]、Aas et al.[2]が詳しい。

*²² コピュラに関する適合度検定手法 (goodness-of-fit testing) については Berg[8]に詳しい。

研究目的は、まず、オペリスク損失金額データが高頻度低額損失、低頻度高額損失の特徴があり、リスク量の大宗を占めるシナリオ損失が離散型データとして扱われる事例が多いことなどの特徴を反映し、安定的なオペリスク量を推定できる分布形を選択し、次に、複雑な依存構造、裾依存性を反映できるコピュラによってモデル化することを提案する。さらに、周辺分布、コピュラのパラメータの同時推定手法を提案し、最後に、その精度を数値実験で検証することにある。

本研究では、まず、多変量オペレーショナルリスク損失金額データの周辺分布を離散分布で近似する。次に、依存関係のモデル化については、構造が簡単で、各変数が同等に扱われており、適合度が高く、周辺分布の確率変数が「上側・下側で漸近依存」の関係にあり、吉羽[43]によれば「 t コピュラにおいては裾依存性の強さを表す自由度パラメータ ν を持つ」D-vine t コピュラで定式化する。さらに、観測データが離散型データの場合の尤度関数の項数、周辺分布、D-vine t コピュラコピュラパラメータの同時推定の場合の最適化の難しさなどを考慮して、MCMC法を用いたベイズ手法を用い、最後に、その精度を数値実験で検証する。離散型データの場合のデータ拡大法はSmith and Khaled[39]を、D-vine t コピュラパラメータのサンプリング手法はMin and Czado[25]を参照する。なお、前者では周辺分布が1個のパラメータであるモデルを前提としているが、本研究では離散分布を構成するビン数に拡張する。

提案した手法の精度検証のための数値実験では、まず、設定したパラメータでAas et al. [2]のアルゴリズムによりコピュラの引数を、一般化逆関数法に類似した手法で高頻度低額、低頻度高額の特徴を有する模擬損失金額データを生成する。次に、MCMC法によりパラメータを推定し、設定値が再現されることを示す。

2章における貢献として、Smith and Khaled[39]では、周辺分布に1個のパラメータを前提としていたが、離散分布を構成するビン数にパラメータ数を拡張し、離散分布およびD-vine t コピュラのパラメータの同時推定手法を提案したことが挙げられる。

今般、金融庁[23]に記載されているとおり、BCBSは、リスク量の計算において、国際統一基準金融機関は2023年3月（国内基準期間は2024年3月末）に先進的計測手法を廃止し、新しいStandardised Measurement Approach（標準的手法）に一本化することを決定した。これにより、リスク量の計算手法は変更されるが、金融機関のリスク管理実務で標準的なオペリスク VaR は内部的な管理手法として継続されることから、離散型データとして取り扱われる事例が多いシナリオ損失金額データと連続型データとして取り扱われる事例が多い内部損失金額データが混在する損失金額データを対象とする損失分布推定手法の一つとして、本研究は必要不可欠と考える。

本章の構成は以下の通りである。2.2節では、損失金額データの同時分布、周辺分布の離散化とその依存性を表現するD-vine t コピュラを実装したモデル

説明, 2.3 節では, ビンと D-vine t コピュラのパラメータのベイズ推定法を説明, 2.4 節では, 模擬データ生成方法や数値実験の説明, および, 提案した手法が有効であることを示す. 2.5 節では, まとめと今後の課題を述べる.

2.2. モデル

(1) \mathbf{X} の同時分布

m 次元離散型データ \mathbf{X} の同時分布とコピュラ関数の関係を示す.

定義 2.1(Nelsen[35] p 43 Definition 2.10.1)

m 次元計量単位における観測データ \mathbf{X} が離散分布に従う場合の確率関数は多変量コピュラ分布関数を C とすれば, (2.1)式で定義される.

$$(2.1) \quad f(\mathbf{x}) = \Pr(X_1 = x_1, \dots, X_m = x_m) = \Delta_{a_1}^{b_1} \Delta_{a_2}^{b_2} \dots \Delta_{a_m}^{b_m} C(\mathbf{v})$$

ただし, $\mathbf{X} = (X_1, \dots, X_m)$ は確率変数, その実現値を $\mathbf{x} = (x_1, \dots, x_m)$ とする. $b_j = F_j(x_j), a_j = F_j(x_j^-)$ を F_j の x_j における左極限值であり, $\mathbf{v} = (v_1, \dots, v_m)$ とすると, $\Delta_{a_k}^{b_k}$ の表記は(2.2)式で定義される.

$$(2.2) \quad \Delta_{a_k}^{b_k} C(u_1, \dots, u_{k-1}, v_k, u_{k+1}, \dots, u_m) = C(u_1, \dots, u_{k-1}, b_k, u_{k+1}, \dots, u_m) - C(u_1, \dots, u_{k-1}, a_k, u_{k+1}, \dots, u_m)$$

例えば, 3次元の場合は(2.3)式のように展開される.

$$(2.3) \quad f(x_1, x_2, x_3) = \Delta_{a_1}^{b_1} \Delta_{a_2}^{b_2} \Delta_{a_3}^{b_3} C(v_1, v_2, v_3) = C(b_1, b_2, b_3) - C(b_1, b_2, a_3) - C(b_1, a_2, b_3) - C(a_1, b_2, b_3) + C(b_1, a_2, a_3) + C(a_1, b_2, a_3) + C(a_1, a_2, b_3) - C(a_1, a_2, a_3)$$

従って, 損失金額データに離散分布を仮定した場合, データ数 n , 計量単位を m とすると, コピュラの尤度関数は $n2^m$ 項数で記述される.

(2) \mathbf{X} の周辺分布

離散型確率変数 \mathbf{X} , その実現値を $\mathbf{x} = \{x_1, \dots, x_n\}$, 第 h 番目の累積確率のビンを π_h とすると, $\pi_h - \pi_{h-1} = \Pr(X = h), h = 1, \dots, l$, その尤度関数は(2.4)式である.

$$(2.4) \quad p(\mathbf{x} | \pi) = \prod_{h=1}^l (\pi_h - \pi_{h-1})^{n_x(h)}$$

ただし, $\pi_0 = 0, \pi_l = 1, n_x(h) = \sum_{i=1}^n \delta_{x_i}(h), \delta_{x_i}(h) = \begin{cases} 1, (x = h) \\ 0, (x \neq h) \end{cases}, 0 < \pi_h \leq 1$ である.

(3) (\mathbf{X}, \mathbf{U}) の同時分布

データ拡大法を導入するために潜在変数 \mathbf{U} と離散型データ \mathbf{X} の (\mathbf{X}, \mathbf{U}) の同時分布を考える。

定義 2.2(Smith and Khaled[39] p 292 (2.3)式)

周辺分布を m 次元, 離散分布と一様分布に従う確率変数を $\mathbf{X} = (X_1, \dots, X_m)$

$\mathbf{U} = (U_1, \dots, U_m)$, $X_j | U_j$ の条件付密度関数を (2.5) 式と定義する。

$$(2.5) \quad f(x_j | u_j) = I(F_j(x_j^-) \leq u_j < F_j(x_j))$$

ここで, $I(A)$ は定義関数であり, A が真の場合 $I(A) = 1$, 偽の場合は $I(A) = 0$ である。

これより, (\mathbf{X}, \mathbf{U}) の同時密度関数は (2.6) 式となる。

$$(2.6) \quad f(\mathbf{x}, \mathbf{u}) = f(\mathbf{x} | \mathbf{u})c(\mathbf{u}) = \prod_{j=1}^m I(F_j(x_j^-) \leq u_j < F_j(x_j)) c(\mathbf{u})$$

ここで, $f(\mathbf{x} | \mathbf{u}) = \prod_{j=1}^m f(x_j | u_j)$ である。

データ番号を $i = 1, \dots, n$, $b_{i,j} = F_j(x_{i,j}; \theta_j)$, $a_{i,j} = F_j(x_{i,j}^-; \theta_j)$, $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$,

$\mathbf{x}_i = (x_{i,1}, \dots, x_{i,m})$, $\mathbf{u} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$, $\mathbf{u}_i = (u_{i,1}, \dots, u_{i,m})$ とすると尤度関数は (2.7) 式となる。

$$(2.7) \quad f(\mathbf{x}, \mathbf{u}) = \prod_{i=1}^n f(\mathbf{x}_i, \mathbf{u}_i) = \prod_{i=1}^n c(\mathbf{u}_i) \prod_{j=1}^m I(a_{i,j} \leq u_{i,j} < b_{i,j})$$

ここで, $a_{i,j} = F_j(x_{i,j}^-; \theta_j)$ は $x_{i,j}$ における F_j の左極限である。

(4) D-vine t コピュラ

多変量コピュラから D-vine コピュラの導出は (2.8) 式のとおりである。

$\mathbf{u} = (u_1, \dots, u_m)$, 周辺分布 j, k の 2 変量のコピュラ密度関数を $c_{j,k}$, そのパラメータを $\phi_{j,k}$, $j = 2, \dots, m$, $k < j$ また $\phi = \{\phi_{j,k}; j = 2, \dots, m, k < j\}$ とすると,

$$(2.8) \quad \begin{aligned} c(\mathbf{u}) &= \prod_{j=2}^m \prod_{k=1}^{j-1} c_{j,k}(u_{j|k+1}, u_{k|j-1}; \phi_{j,k}) \\ u_{j|k} &= C_{j|k, \dots, j-1}(u_j | u_k, \dots, u_{j-1}; \phi) = h_{j,k} \circ h_{j,k+1} \circ \dots \circ h_{j,j-1}(u_j) \\ u_{k|j} &= C_{k|k+1, \dots, j}(u_k | u_{k+1}, \dots, u_j; \phi) = h_{k,k+1} \circ h_{k,k+2} \circ \dots \circ h_{k,j}(u_k) \\ u_j &= C_{j|k, \dots, j-1}^{-1}(u_j | u_k, \dots, u_{j-1}; \phi) = h_{j,j-1}^{-1} \circ h_{j,j-2}^{-1} \circ \dots \circ h_{j,k}^{-1}(u_j) \end{aligned}$$

ここで,

$$\omega_j = C_{j|k, \dots, j-1}(u_j | u_k, \dots, u_{j-1}; \phi), h_{j,k}(v_j | v_k; \phi_{j,k}) = \frac{\partial}{\partial v_k} C_{j,k}(v_j, v_k; \phi_{j,k}),$$

$$C_{j,k}(u_j, u_k; \phi_{j,k}) = \int_0^{u_j} \int_0^{u_k} c_{j,k}(v_1, v_2; \phi_{j,k}) dv_1 dv_2 \text{ である.}$$

$f \circ g$ は写像 g と写像 f の合成, すなわち $(f \circ g)(x) = f(g(x))$ である.

また, パラメータを $\phi = (\rho, \nu)$, ただし, 相関係数を $-1 < \rho < 1$, 自由度を $\nu > 0$, 2 変量 t 分布分布関数を T_ν , t 分布密度関数を t_ν とすると, D-vine t コピュラは (2.9) 式である.

$$C(u_j, u_k; \phi_{j,k}) = T_\nu(t_\nu^{-1}(u_j), t_\nu^{-1}(u_k); \rho_{j,k})$$

$$(2.9) \quad h(u_j | u_k) = C_{j|k}(u_j | u_k; \phi_{j,k}) = t_{\nu+1} \left(\left[\frac{\nu_{j,k} + 1}{\nu_{j,k} + (t_\nu^{-1}(u_k))^2} \right]^{1/2} \frac{t_\nu^{-1}(u_j) - \rho_{j,k} t_\nu^{-1}(u_k)}{\sqrt{1 - \rho_{j,k}^2}} \right)$$

$$u_j = h^{-1}(\omega_j | u_k) = C_{j|k}^{-1}(\omega_j | u_k; \phi_{j,k}) = t_\nu \left(\left[\frac{(1 - \rho_{j,k}^2)(\nu_{j,k} + (t_\nu^{-1}(u_k))^2) + 1}{\nu_{j,k} + 1} \right]^{1/2} t_{\nu+1}^{-1}(\omega_j) + \rho_{j,k} t_\nu^{-1}(u_k) \right)$$

2.3. ベイジアン MCMC 法アプローチ

MCMC 法におけるサンプリング手順, そのパラメータの条件付事後分布は以下のとおりである. Smith and Khaled[39]では周辺分布, D-vine t コピュラのパラメータのほか, データ拡大法を提案し潜在変数 \mathbf{u} をサンプリングしている.

$$\text{Step1} \quad f(\Theta | \phi, \mathbf{u}, \mathbf{x})$$

$$\text{Step2} \quad f(\mathbf{u} | \Theta, \phi, \mathbf{x})$$

$$\text{Step3} \quad f(\rho_{j,k} | \mathbf{u})$$

$$\text{Step4} \quad f(\nu | \mathbf{u})$$

ここで, Θ は周辺分布のパラメータ, $\phi = \{\rho_{j,k}, \nu\}$ は D-vine t コピュラの周辺分布 j, k の相関係数と自由度パラメータである.

$\Theta, u_{i,j}, \rho_{j,k}, \nu$ のサンプリングアルゴリズムは, Müller[33]によって提唱された Metropolis within Gibbs アルゴリズムを採用する. これは, Gibbs Sampler(Gibbs法)によって条件付事後分布からサンプリングができないときに,

それを適当なMetropolis Hastings (MH) 法*²³に置き換えた, Gibbs法とMH法とを組合せた方法である. サンプリングの概略のフローチャートは図 2-1のとおりである.

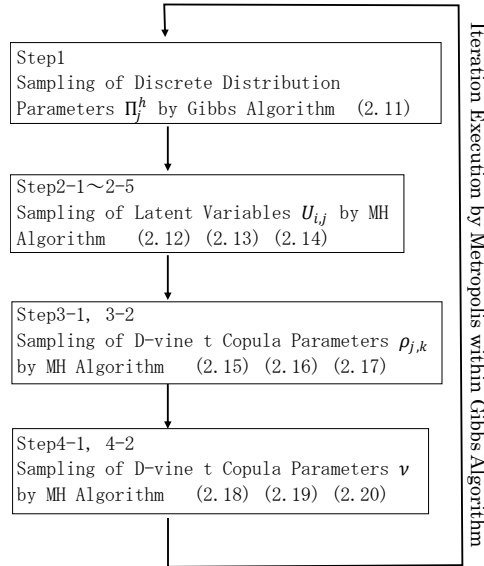


図 2-1. Parameters sampling flowchart by MH within Gibbs.

まず, Step1 について, Θ を条件付事後分布からサンプリングする. Smith and Khaled[39]によれば, 各周辺分布の 1 種類のパラメータ θ_j の事前分布を $f(\theta_j)$ とすると条件付事後分布は (2.10) 式である. *²⁴

$$(2.10) \quad f(\Theta | \phi, \mathbf{u}, \mathbf{x}) \propto f(\mathbf{x} | \Theta, \mathbf{u}) f(\Theta) = \prod_{j=1}^m \prod_{i=1}^n \mathbf{I}(a_{i,j} \leq u_{i,j} < b_{i,j}) f(\theta_j)$$

本研究では, (2.4) 式の累積確率のビン π_h を用いた離散分布により周辺分布を近似する.

ここで, (2.10) 式における各周辺分布の 1 個のパラメータ θ_j を (各周辺分布の全てのビン数 - 1) 個のパラメータ数*²⁵に拡張することから, 記号の表記について変更, 新たに追加設定することとする.

Π_j^h は周辺分布 j の h ビンの累積確率の確率変数, その実現値を π_j^h とする. また, $x_{i,j}^h, u_{i,j}^h$ は周辺分布 j の h ビンに属する i 番目の観測データと潜在変数,

$$\pi_j^h = F_j(x_{i,j}^h), \quad 0 < \pi_j^1 < \pi_j^2 < \dots < \pi_j^h < \dots < \pi_j^{l-1} < \pi_j^l = 1,$$

$i=1, \dots, n, j=1, \dots, m, h=1, \dots, l-1$ である.

*²³ MH 法については Chib and Greenberg[13]に詳しい.

*²⁴ Smith and Khaled[39]p294 左下段に記載されている.

*²⁵ 右端のビンは, その累積確率が 1 と定数でありパラメータの推定対象外である.

例えば，図 2-2 のように，観測データ x_1, x_2 の場合， $\pi^1 = F(x_1) = F(x_2)$ ，観測データ x_3 の場合， $\pi^2 = F(x_3), \pi^1 = F(x_3^-)$ となる．

ここで， x_1, \dots, x_n のビンの帰属は観測されていること，後述の u_i のサンプリング (Step2-1～Step2-5) から明らかのように， x_i と同じデータ番号の u_i は同一のビンに帰属することに注意が必要である．離散分布のビンの例を図 2-2 に示す．

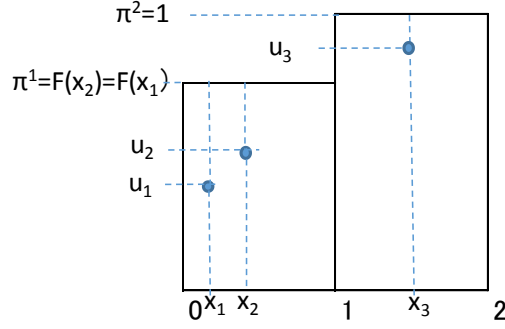


図 2-2. An example is given for the case where the number of bins is two. The suffix j is omitted for the measurement unit, bin number h for π and the data number i for x, u are written. And the X -coordinate is set to 0,1,2 for simplicity.

$\Pi_j^1 < \dots < \Pi_j^h < \dots < \Pi_j^{l-1}$ を $l-1$ 個の独立な一様分布 $(-1, 1)$ に従う順序統計量とすると， $(\Pi_j^1, \dots, \Pi_j^h, \dots, \Pi_j^{l-1})$ の同時確率密度関数は， $g_{\Pi_j^1, \dots, \Pi_j^{l-1}}(\pi_j^1, \dots, \pi_j^{l-1}) = (l-1)!$ であり，これを事前分布とすると，条件付事後分布は (2.11) 式となる．

$b_{i,j} = F_j(x_{i,j}; \theta_j), a_{i,j} = F_j(x_{i,j}^-; \theta_j)$ より

$$\begin{aligned}
& \prod_{j=1}^m \prod_{i=1}^n I(a_{i,j} \leq u_{i,j} < b_{i,j}) g_{\Pi_j^1, \dots, \Pi_j^{l-1}}(\pi_j^1, \dots, \pi_j^{l-1}) \\
&= \prod_{j=1}^m \underbrace{I(\Pi_j^0 \leq u_{1,j}^1 < \Pi_j^1) \times \dots \times I(\Pi_j^0 \leq u_{n_j,j}^1 < \Pi_j^1)}_{\text{第 1 ビンに帰属するデータ } u} \times \underbrace{I(\Pi_j^1 \leq u_{1,j}^2 < \Pi_j^2) \times \dots \times I(\Pi_j^1 \leq u_{n_j,j}^2 < \Pi_j^2)}_{\text{第 2 ビンに帰属するデータ } u} \\
&\times \dots \times \underbrace{I(\Pi_j^{l-2} \leq u_{1,j}^{l-1} < \Pi_j^{l-1}) \times \dots \times I(\Pi_j^{l-2} \leq u_{n_{l-1},j}^{l-1} < \Pi_j^{l-1})}_{\text{第 } l-1 \text{ ビンに帰属するデータ } u} \times \underbrace{I(\Pi_j^{l-1} \leq u_{1,j}^l < \Pi_j^l) \times \dots \times I(\Pi_j^{l-1} \leq u_{n_l,j}^l < \Pi_j^l)}_{\text{第 } l \text{ ビンに帰属するデータ } u} \times (l-1)! \\
&\propto \prod_{j=1}^m I(\Pi_j^0 \leq u_{1,j}^1 < \Pi_j^1) \times \dots \times I(\Pi_j^0 \leq u_{n_j,j}^1 < \Pi_j^1) \times I(\Pi_j^1 \leq u_{1,j}^2 < \Pi_j^2) \times \dots \times I(\Pi_j^1 \leq u_{n_j,j}^2 < \Pi_j^2) \\
&\times \dots \times I(\Pi_j^{l-2} \leq u_{1,j}^{l-1} < \Pi_j^{l-1}) \times \dots \times I(\Pi_j^{l-2} \leq u_{n_{l-1},j}^{l-1} < \Pi_j^{l-1}) \times I(\Pi_j^{l-1} \leq u_{1,j}^l < \Pi_j^l) \times \dots \times I(\Pi_j^{l-1} \leq u_{n_l,j}^l < \Pi_j^l) \\
(2.11) \quad &= \prod_{j=1}^m I(\max(\text{第1ビンに帰属する } u_{1,j}^1, \dots, u_{n_j,j}^1) \leq \Pi_j^1 < \min(\text{第2ビンに帰属する } u_{1,j}^2, \dots, u_{n_j,j}^2)) \\
&\times \dots \times I(\max(\text{第 } l-1 \text{ ビンに帰属する } u_{1,j}^{l-1}, \dots, u_{n_{l-1},j}^{l-1}) \leq \Pi_j^{l-1} < \min(\text{第 } l \text{ ビンに帰属する } u_{1,j}^l, \dots, u_{n_l,j}^l))
\end{aligned}$$

ここで， $u_{n_h,j}^h$ は j 周辺分布の第 h ビンに帰属する第 n_h 番目の潜在変数で， n_h は第 h ビンに帰属するデータ総数である．ただし， $\Pi_j^0 = 0, \Pi_j^l = 1, 0 < \Pi_j^h < 1$ である．例えば， j 周辺分布の第 1 ビンに帰属する潜在変数は， $\{u_{1,j}^1, \dots, u_{n_1,j}^1\}$ となる．

$\Pi_j^1, \dots, \Pi_j^{l-1}$ については、(2.11)式の定義関数が1となる範囲で、一様乱数を各周辺分布 $j = 1, \dots, m$ ごとにサンプリングする。

次に、Step2 について、潜在変数 $U_{i,j}$ のサンプリングアルゴリズムは Smith and Khaled[39]に従い MH 法を選択する。

(\mathbf{X}, \mathbf{U}) の同時密度関数の周辺化により、 \mathbf{X} の同時分布を導出する。

命題2.3(Smith and Khaled[39] p 292 Proposition1)

(\mathbf{X}, \mathbf{U}) の同時密度関数が $f(\mathbf{x}, \mathbf{u}) = f(\mathbf{x} | \mathbf{u})c(\mathbf{u}) = \prod_{j=1}^m I(F_j(x_j^-) \leq u_j < F_j(x_j)) c(\mathbf{u})$ と

記述される時周辺分布 \mathbf{X} の確率関数 (mass function) は (2.1) 式と等しく次のようになる。

$$f(\mathbf{x}) = \Pr(\mathbf{X}_1 = x_1, \dots, \mathbf{X}_m = x_m) = \Delta_{a_1}^{b_1} \Delta_{a_2}^{b_2} \dots \Delta_{a_m}^{b_m} C(\mathbf{v})$$

潜在変数 U_j の密度関数を導出するために、まず、 (U_1, \dots, U_j) の同時分布を考える。

命題2.4(Smith and Khaled[39] p 292 Proposition2)

$j=1, \dots, m-1$ のとき、 $(U_1, \dots, U_j) | \mathbf{X}$ の密度関数は

$$f(u_1, \dots, u_j | \mathbf{x}) = \frac{c(u_1, \dots, u_j)}{f(\mathbf{x})} \left(\Delta_{a_{j+1}}^{b_{j+1}} \dots \Delta_{a_m}^{b_m} C_{j+1, \dots, m | 1, \dots, j}(\mathbf{v}_{j+1}, \dots, \mathbf{v}_m | u_1, \dots, u_j) \right) \prod_{k=1}^j I(a_k \leq u_j < b_k)$$

ここで、 $c(u_1, \dots, u_j) = \int c(\mathbf{u}) du_{j+1} \dots du_m$ 、 $C_{j+1, \dots, m | 1, \dots, j}$ は $U_{j+1}, \dots, U_m | U_1, \dots, U_j$ の分布関数である。

(U_1, \dots, U_j) の同時分布から、 U_j の密度関数を導出する。

系2.5(Smith and Khaled[39] p 292 Corollary1)

$j=2, \dots, m$ のとき、 $U_j | U_1, \dots, U_{j-1}, \mathbf{X}$ の密度関数は

$$f(u_j | u_1, \dots, u_{j-1}, \mathbf{x}) = c_{j | 1, \dots, j-1}(u_j | u_1, \dots, u_{j-1}) I(a_j \leq u_j < b_j) K_j(u_1, \dots, u_j),$$

ここで $j=m$ のとき、

$$K_m(u) = \frac{1}{\Delta_{a_m}^{b_m} C_{m | 1, \dots, m-1}(\mathbf{v}_m | u_1, \dots, u_{m-1})},$$

$j=2, \dots, m-1$ のとき,

$$K_j(u_1, \dots, u_j) = \frac{\Delta_{a_{j+1}}^{b_{j+1}} \cdots \Delta_{a_m}^{b_m} C_{j+1, \dots, m|1, \dots, j}(v_{j+1}, \dots, v_m | u_1, \dots, u_j)}{\Delta_{a_j}^{b_j} \cdots \Delta_{a_m}^{b_m} C_{j, \dots, m|1, \dots, j-1}(v_j, \dots, v_m | u_1, \dots, u_{j-1})}$$

提案分布は系2.5から

$$(2.12) \quad g_j(u_{i,j} | u_{i,1}, \dots, u_{i,j-1}) = \frac{c_{j|1, \dots, j-1}(u_{i,j} | u_{i,1}, \dots, u_{i,j-1}; \phi) I(a_{i,j} \leq u_{i,j} < b_{i,j})}{C_{j|1, \dots, j-1}(b_{i,j} | u_{i,1}, \dots, u_{i,j-1}; \phi) - C_{j|1, \dots, j-1}(a_{i,j} | u_{i,1}, \dots, u_{i,j-1}; \phi)}$$

ただし, $g_1(u_{i,1}) = \frac{I(a_{i,1} \leq u_{i,1} < b_{i,1})}{b_{i,1} - a_{i,1}}$ であり, そのサンプリング手順は次の

Step2-1~Step2-5を $i=1, \dots, n$ について繰り返す.

Step2-1. $u_{i,1} \sim U(a_{i,1}, b_{i,1})$ に従う乱数をサンプリングする.

Step2-2. $w_{i,1} \sim U(A_{i,1}, B_{i,1})$ に従う乱数をサンプリングする.

ただし, $A_{i,j} = C_{j|1, \dots, j-1}(a_{i,j} | u_{i,1}, \dots, u_{i,j-1}; \phi)$, $B_{i,j} = C_{j|1, \dots, j-1}(b_{i,j} | u_{i,1}, \dots, u_{i,j-1}; \phi)$ である.

同様に, $(w_{i,1}, \dots, w_{i,m})$ をサンプリングする. ($j=2, \dots, m$)

ここで, $\phi_{j,k}$ は周辺分布 j, k のD-vine tコピュラのパラメータ $\rho_{j,k}, \nu$ である.

Step2-3. $u_{i,j} = C_{j|1, \dots, j-1}^{-1}(w_{i,j} | u_{i,1}, \dots, u_{i,j-1}; \phi)$ から $(u_{i,1}, \dots, u_{i,m})$ を計算する.

Step2-4. $u_{i,jk}, u_{i,k|j}$ を (2.13) 式で更新する.

$$(2.13) \quad \begin{aligned} u_{i,jk} &= h_{j,k}(u_{i,jk+1} | u_{i,k|j-1}; \phi), & k &= j-1, \dots, 1 \\ u_{i,k|j} &= h_{j,k}(u_{i,k|j-1} | u_{i,jk+1}; \phi), & k &= 1, \dots, j-1 \end{aligned}$$

Step2-5. 採択確率は (2.14) 式である.

$$(2.14) \quad \alpha_i = \min \left(1, \prod_{j=2}^m \frac{C_{j|1, \dots, j-1}(b_{i,j} | u_{i,1}^{new}, \dots, u_{i,j-1}^{new}; \phi) - C_{j|1, \dots, j-1}(a_{i,j} | u_{i,1}^{new}, \dots, u_{i,j-1}^{new}; \phi)}{C_{j|1, \dots, j-1}(b_{i,j} | u_{i,1}^{old}, \dots, u_{i,j-1}^{old}; \phi) - C_{j|1, \dots, j-1}(a_{i,j} | u_{i,1}^{old}, \dots, u_{i,j-1}^{old}; \phi)} \right)$$

ここで $u_{i,j}^{new}$ は現在試行の標本, $u_{i,j}^{old}$ は前回試行の標本である.

また, Step3 について, $\rho_{j,k}$ のサンプリングアルゴリズムは Min and Czado[25] に

従い提案分布を切断正規分布とする MH with random walk (MH-RW) 法*26を選択する。

Step3-1. $\rho_{j,k}^{new}$ の事前分布を一様分布 $(-1,1)$ のとすると、提案分布は (2.15)

式, D-vine t コピュラの尤度を $\prod_{i=1}^n c(\mathbf{u}_i; \phi_{j,k})$ *27 とすると、条件付事後分布は (2.16) 式となる。

ただし、提案分布に平均 $\rho_{j,k}^{old}$ 、分散 $\tau_{\rho_j}^2$ 、区間 $(-1,1)$ で定義された切断正規分布 $TN_{(-1,1)}(\rho_{j,k}^{old}, \tau_{\rho_j}^2)$ を仮定する。

$$(2.15) \quad \rho_{j,k}^{new} \sim TN_{(-1,1)}(\rho_{j,k}^{old}, \tau_{\rho_j}^2)$$

$$(2.16) \quad p(\rho_{j,k}^{new} | v, u_{i,j}) \propto \frac{1}{2} \prod_{i=1}^n c(\mathbf{u}_i; \phi_{j,k})$$

Step3-2. 採択確率は (2.17) 式である。

$$(2.17) \quad \alpha_j^{\rho} = \min \left\{ \frac{p(\rho_{j,k}^{new})}{p(\rho_{j,k}^{old})} CF_{\rho,j}(\rho_{j,k}^{new}, \rho_{j,k}^{old}), 1 \right\}$$

$$\text{ここで, } CF_{\rho,j}(\rho_{j,k}^{(new)}, \rho_{j,k}^{(old)}) \equiv \frac{\Phi\left\{\frac{1 - \rho_{j,k}^{(old)}}{\tau_{\rho}}\right\} - \Phi\left\{\frac{-1 - \rho_{j,k}^{(old)}}{\tau_{\rho}}\right\}}{\Phi\left\{\frac{1 - \rho_{j,k}^{(new)}}{\tau_{\rho}}\right\} - \Phi\left\{\frac{-1 - \rho_{j,k}^{(new)}}{\tau_{\rho}}\right\}}$$

であり、提案分布に切断正規分布 $TN_{(-1,1)}(\rho_{j,k}^{old}, \tau_{\rho_j}^2)$ を適用したことによる調整項である。

最後に Step4 について、 $\rho_{j,k}$ と同様に、提案分布を切断正規分布とする MH-RW 法を選択する。

提案分布に平均 v^{old} 、分散 τ_v^2 、区間 $(1, \text{Upper})$ で定義された切断正規分布 $TN_{(1, \text{Upper})}(v^{old}, \tau_v^2)$ を仮定する。

Step4-1. v^{new} の事前分布を一様分布 $(1, \text{Upper})$ のとすると、提案分布は (2.18) 式、条件付事後分布は (2.19) 式となる。なお U はパラメータ v^{new} の上限とし、Min and Czado[25], Min and Czado[26] に従い $\text{Upper} = 20$ とする。

$$(2.18) \quad v^{new} \sim TN_{(1, \text{Upper})}(v^{old}, \tau_v^2)$$

$$(2.19) \quad p(v^{new} | \rho_{j,k}, u_{i,j}) \propto \frac{1}{\text{Upper} - 1} \prod_{i=1}^n c(\mathbf{u}_i; \phi_{j,k})$$

Step4-2. 採択確率は (2.20) 式である。

$$(2.20) \quad \alpha_j^v = \min \left\{ \frac{p(v^{new})}{p(v^{old})} CF_{\rho,j}(v^{new}, v^{old}), 1 \right\}$$

*26 MH-RW 法については Chib and Greenberg[13] に詳しい。

*27 D-vine コピュラの定式化、尤度など Min and Czado[25] に詳しい。

$$\text{ここで, } CF_{v,j}(v^{new}, v^{old}) \equiv \frac{\Phi\left\{\frac{(U-v^{old})}{\tau_v}\right\} - \Phi\left\{\frac{(1-v^{old})}{\tau_v}\right\}}{\Phi\left\{\frac{(U-v^{new})}{\tau_v}\right\} - \Phi\left\{\frac{(1-v^{new})}{\tau_v}\right\}}$$

であり提案分布 $TN_{(1, \text{Upper})}(v^{old}, \tau_v^2)$ の調整項である.

2.4. 数値実験

オペリスク損失金額データの分析においては、実際のデータで分析することが望ましいが、その秘匿性が極めて高く、外部には公表されていないことから、データを入手することはできない。一方、森他[30]、BCBS[5]など監督官庁、中央銀行などによる金融機関の提出したオペリスク損失金額データの分析事例、Nakatsuma[34]など研究者と金融機関との共同研究による分析事例などが公開されている。本研究では、これら文献を参考に、シナリオ、低頻度高額損失、高頻度低額損失などオペリスク損失金額データの特徴を反映させた模擬損失金額データを生成する。

数値実験については、5計量単位、1計量単位のビン数は200個の事例について検証する。5計量単位の場合、10種類のD-vine tコピュラから構成され、そのパラメータについて10種類の相関係数 $\rho_{j,k}$ 、1種類の自由度 ν を考える。また、ビンについては、995(=199×5)個のパラメータを推定する。なお、構成される各計量単位のデータ数は等しく偏りが無いことを前提とする。

2.4.1. 数値実験におけるデータ作成

(1) D-vine tコピュラから依存構造を反映した $u_{i,j}$ の生成

5変量のD-vine tコピュラ密度関数は $\mathbf{u} = (u_1, u_2, u_3, u_4, u_5)$,

$P(X_j < x_j | X_k = x_k) \equiv F(x_j | x_k)$ とすると、(2.21)式のように分解できる。

$$\begin{aligned} c(u_1, u_2, u_3, u_4, u_5) &= c_{12}(u_1, u_2)c_{23}(u_2, u_3)c_{34}(u_3, u_4)c_{45}(u_4, u_5)c_{13|2}(u_{1|2}, u_{3|2} | X_2 = x_2) \\ c_{24|3}(u_{2|3}, u_{4|3} | X_3 = x_3)c_{35|4}(u_{3|4}, u_{5|4} | X_4 = x_4)c_{14|23}(u_{1|23}, u_{4|23} | X_2 = x_2, X_3 = x_3) \\ c_{25|34}(u_{2|34}, u_{5|34} | X_3 = x_3, X_4 = x_4)c_{15|234}(u_{1|234}, u_{5|234} | X_2 = x_2, X_4 = x_4, X_3 = x_3) \\ &= c_{12}(F(x_1), F(x_2))c_{23}(F(x_2), F(x_3))c_{45}(F(x_4), F(x_5))c_{13|2}(F(x_1 | x_2), F(x_3 | x_2)) \\ &= c_{24|3}(F(x_2 | x_3), F(x_4 | x_3))c_{35|4}(F(x_3 | x_4), F(x_5 | x_4))c_{14|23}(F(x_1 | x_2, x_3), F(x_4 | x_2, x_3)) \\ &= c_{25|34}(F(x_2 | x_3, x_4), F(x_5 | x_3, x_4))c_{15|234}(F(x_1 | x_2, x_3, x_4), F(x_5 | x_2, x_3, x_4)) \\ &= c_{12}(u_1, u_2)c_{23}(u_2, u_3)c_{34}(u_3, u_4)c_{45}(u_4, u_5)c_{13|2}(u_{1|2}, u_{3|2})c_{24|3}(u_{2|3}, u_{4|3}) \\ &= c_{35|4}(u_{3|4}, u_{5|4})c_{14|23}(u_{1|23}, u_{4|23})c_{25|34}(u_{2|34}, u_{5|34})c_{15|234}(u_{1|234}, u_{5|234}) \end{aligned} \quad *28$$

*28 Haff et al. [19]によれば周辺分布が一様分布や楕円分布(正規・tなど)に従う場合に、条件付コピュラと無条件コピュラにおいて等号が成立する。また、その他の分布に従う場合であっても、条件付コピュラの評価を無条件コピュラで近似することができるとの分析結果がある。

次に、周辺分布の依存関係を反映した一様乱数(0,1)である $u_{i,j}$ の生成について述べる. D-vine t コピュラの引数を省略して $c_{12}, c_{23}, c_{34}, c_{45}, c_{13|2}, c_{24|3}, c_{35|4}, c_{14|23}, c_{25|34}, c_{15|234}$ とその相関係数パラメータを $\rho_{12}, \rho_{23}, \rho_{34}, \rho_{45}, \rho_{13|2}, \rho_{24|3}, \rho_{35|4}, \rho_{14|23}, \rho_{25|34}, \rho_{15|234}$, 自由度パラメータを各 D-vine t コピュラ共通の ν とすると, 以下のように設定し, Aas et al. [2] のアルゴリズムによって, (2.22) 式から, 計量単位ごとに 1000 個の乱数 $u_{i,j}$ をサンプリングする.

$$\begin{aligned}
 \rho_{12} &= 0.3, \rho_{23} = 0.4, \rho_{34} = 0.5, \rho_{45} = 0.6, \rho_{13|2} = 0.7, \rho_{24|3} = 0.8, \\
 \rho_{35|4} &= -0.2, \rho_{14|23} = -0.3, \rho_{25|34} = -0.4, \rho_{15|234} = -0.5, \nu = 7 \\
 u_{i,1} &= w_{i,1} \\
 u_{i,2} &= C_{2|1}^{-1}(w_{i,2} | u_{i,1}) = h_{2,1}^{-1}(w_{i,2}) \\
 u_{i,3} &= C_{3|1,2}^{-1}(w_{i,3} | u_{i,1}, u_{i,2}) = h_{3,2}^{-1} \circ h_{3,1}^{-1}(w_{i,3}) \\
 u_{i,4} &= C_{4|1,2,3}^{-1}(w_{i,4} | u_{i,1}, u_{i,2}, u_{i,3}) = h_{4,3}^{-1} \circ h_{4,2}^{-1} \circ h_{4,1}^{-1}(w_{i,4}) \\
 u_{i,5} &= C_{5|1,2,3,4}^{-1}(w_{i,5} | u_{i,1}, u_{i,2}, u_{i,3}, u_{i,4}) = h_{5,4}^{-1} \circ h_{5,3}^{-1} \circ h_{5,2}^{-1} \circ h_{5,1}^{-1}(w_{i,5})
 \end{aligned}
 \tag{2.22}$$

ここで, $(w_{i,1}, w_{i,2}, w_{i,3}, w_{i,4}, w_{i,5})$ は独立な一様乱数(0,1)である. 各計量単位の乱数 $u_{i,j}$ の散布図は図 2-3 である.

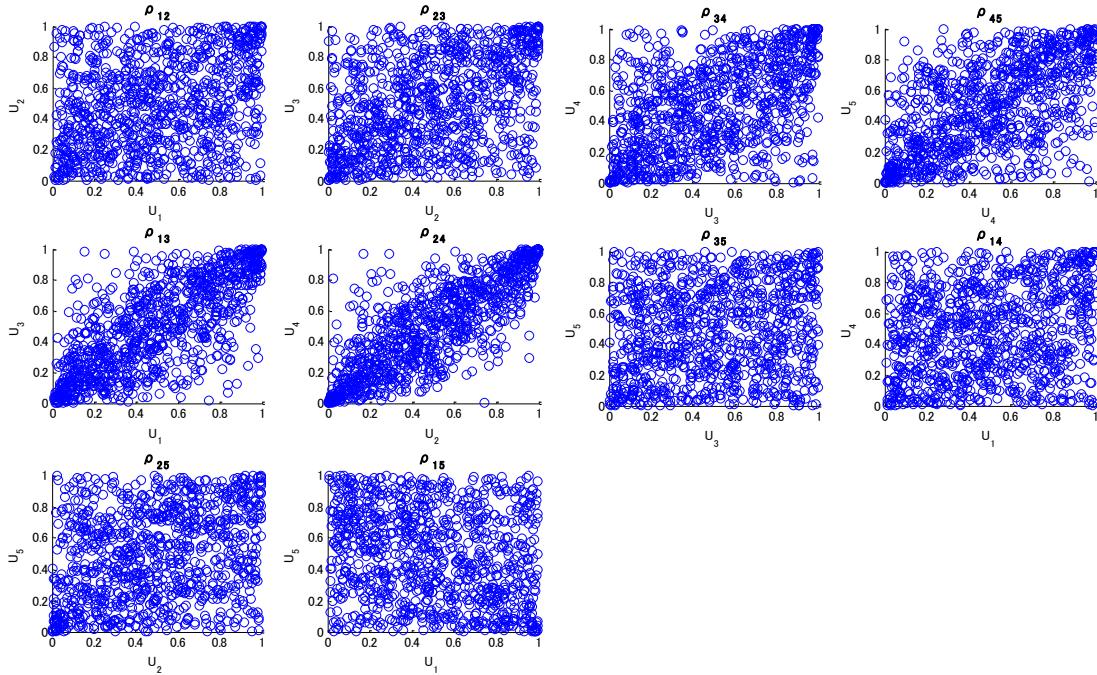


図 2-3. Scatter plot of $u_{i,j}$.

(2) 観測データ $x_{i,j}^h$ の生成

オペリスク損失金額データの特徴である高頻度低額損失, 低頻度高額損失を反映させたビンによって, $x_{i,j}^h$ を生成する.

ここで、 j 周辺分布のビンの x 軸座標を、 $0 < x_j^1 < \dots < x_j^{h-1} < x_j^h < x_j^{h+1} < \dots < x_j^l$ とする。ここで、データ拡大法で生成された $u_{i,j}$ を例にとると、 $0 < u_{i,j} < \pi_j^1$ のとき、 $u_{i,j}$ は第1ビンに帰属して、その観測データ $x_{i,j}$ を、 $x_{i,j}^1 \sim U(0, x_j^1)$ に従う乱数からサンプリングする。同様に、 $\pi_j^{h-1} < u_{i,j} < \pi_j^h$ のとき、 $u_{i,j}$ は第 h ビンに帰属して、 $x_{i,j}$ を $x_{i,j}^h \sim U(x_j^{h-1}, x_j^h)$ に従う乱数からサンプリングする。

数値実験では、観測データ $x_{i,j}^h$ を1000個、5計量単位、各計量単位のビン数200を作成し、そのビン幅は $0.05 (= 10/200)$ とする。

オペリスク損失金額データの特徴である高頻度低額損失、低頻度高額損失のビンへの反映のさせ方は、 $\pi_1^{100} = 0.9, \pi_2^{100} = 0.8, \pi_3^{100} = 0.7, \pi_4^{100} = 0.8, \pi_5^{100} = 0.7$ 、すなわち、第100番目のビンまでの累積確率が第1~5計量単位では0.9, 0.8, 0.7, 0.8, 0.7に設定する。また、第1番目から100番目、101番目から200番目までのビンの累積確率の増分は等確率とする。例えば、 $\pi_1^{100} = 0.9$ の場合は、第1ビンから第100ビンまでの確率の増分は、 $0.9/100$ 、第101ビンから第200ビンまでの確率の増分は $(1-0.9)/100$ である。

表 2-2 Basic statistics of observation data $x_{i,j}$.

	Average	Standard Deviation	Median	Minimum	Maximum	Kurtosis	Skewness
Cell 1	3.015	2.159	2.666	0.010	9.987	4.080	1.074
Cell 2	3.470	2.530	3.047	0.001	9.994	3.052	0.878
Cell 3	3.997	2.769	3.482	0.002	9.993	2.216	0.549
Cell 4	3.415	2.462	3.069	0.016	9.996	3.217	0.881
Cell 5	3.865	2.677	3.478	0.009	9.998	2.328	0.546

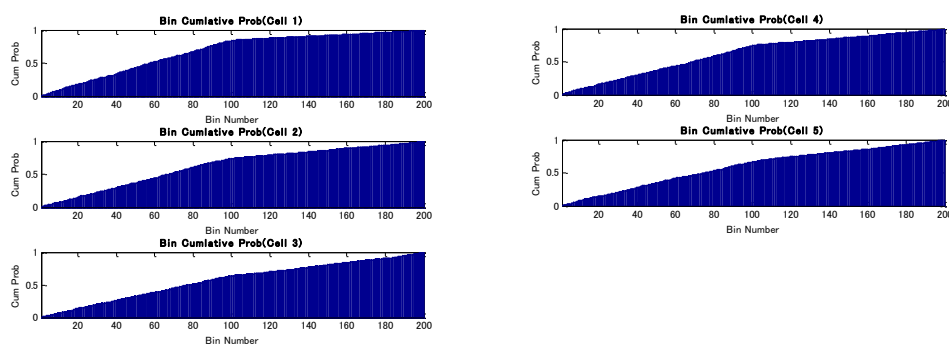


図 2-4. Cumulative probability of bin of observation data $x_{i,j}$.

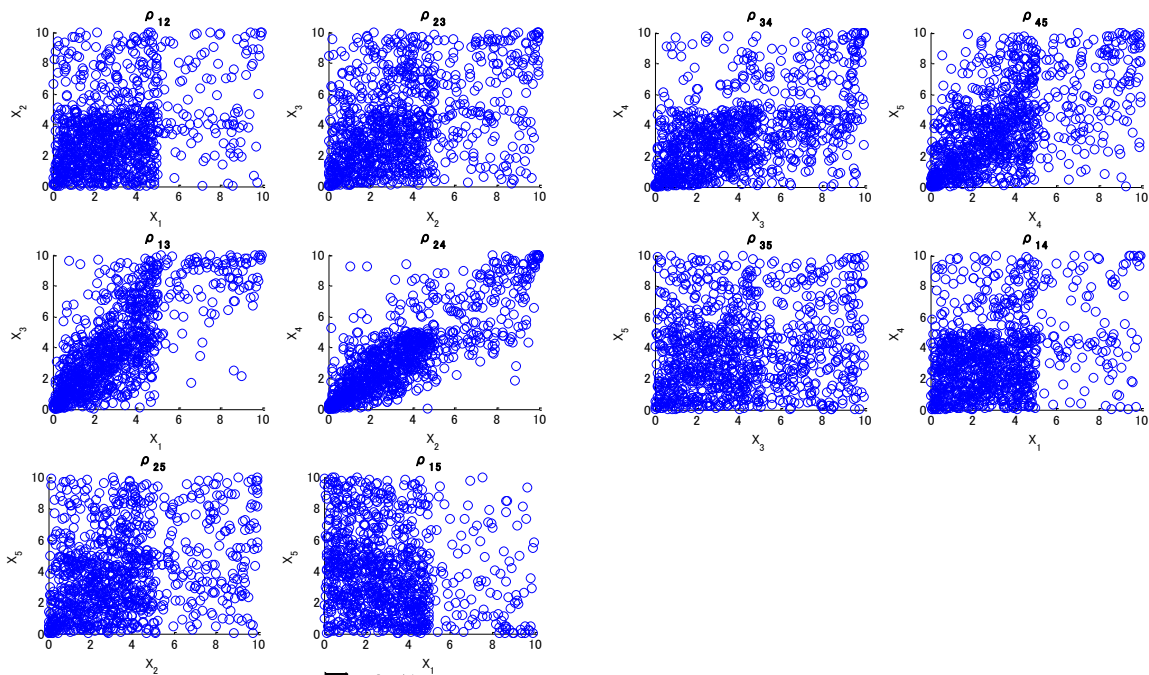


図 2-5. Scatter plot of observed data $x_{i,j}$.

数値実験における模擬データの設定について以下に述べる．損失金額データ数の設定に関しては，金融機関が内部損失やシナリオ損失金額データに関する情報を開示していないことから，その詳細は明らかでない．また，その業態や規模により異なることが想像され，その妥当性について議論することは難しい．本研究では，各計量単位のデータ数は 1000 個に設定している．これは，筆者の経験から相応の規模の銀行の場合に適切な水準である．計量単位数については，2.1 節脚注 16 で記載したように，56 区分の計量単位が当局によって規定されているが，日本銀行[36]によれば，実務においては，データの状況を考慮して分析者の裁量で 5～7 区分に調整されており，計算負荷も考慮して 5 計量単位に設定した．各計量単位のデータ数に偏りがなく設定したことについても，特に問題がないものと考えられる．模擬データの生成については，3 種類のオペリスクの特徴の高頻度低額，低頻度高額損失が反映されていること，10 種類の依存構造が設定されている．この状況は表 2-2 の観測データ $x_{i,j}^h$ の基本統計量，図 2-4 のビンの累積確率のグラフから，高頻度低額損失，低頻度高額損失のオペリスク損失金額データの特徴を反映したデータが作成されていること，図 2-5 の散布図から各計量単位の様々な依存構造が確認できる．以上から，精度検証における模擬データは，オペリスク損失データの特徴，その依存構造について多様な事例が考慮されていると考える．

2.4.2. MCMC 法によるパラメータ推定結果

MCMC 法について単一連鎖法を選択，試行回数は 10 万回で，最初の 7 万回は捨て，その後の 3 万回を標本とする．

ここで，平均と設定値との乖離率を(2.23)式と定義する．

$$(2.23) \quad \text{乖離率} \equiv |(\text{事後平均値} - \text{設定値}) / \text{設定値}|$$

収束診断 Convergence Diagnostics (CD)^{*29}については、Geweke[18]が推奨するように前半と後半の標本数を $T_1=3000, T_2=15000$ とする。

まず、ビンの推定結果について記載する。

全ての 995 ビンのパラメータの推定結果について、事後平均値は設定値よりも、やや小さく推定される傾向が確認され、その乖離率の最小は 0.04%，最大は 62.08%，平均は 4.51%と高い精度が確認された。表 2-3 に乖離率の大きいビン上位 10 について記載する。

表 2-3. Top 10 bins with the largest divergence ratio

Bin Number	π_2^1	π_5^2	π_4^7	π_3^1	π_2^4	π_4^8	π_2^3	π_5^{13}	π_5^{11}	π_4^5
Divergence Ratio (%)	62.1	31.6	21.8	21.1	18.7	18.2	18.0	18.0	17.2	16.9
Posterior Mean	0.013	0.018	0.068	0.008	0.038	0.076	0.028	0.107	0.090	0.047
Set Value	0.008	0.014	0.056	0.007	0.032	0.064	0.024	0.091	0.077	0.040

表 2-3 の乖離率の大きい上位 10 のビンについては、総じてビン番号が小さく累積確率が小さい事例で、特に π_2^1 の事例では、ビンの累積確率の設定値が 0.008 と微小なため、標本の事後平均のわずかな変動で乖離率が 62.1%と大きく計上されたものと解釈される。CD については、有意水準 5%の確率で 100 ビンが受容され、その割合は 10.1%(=100/995)で、棄却されたパラメータの標本時系列については、条件付事後分布へ収束していない可能性がある。非効率性因子 (Numerical inefficiency factor NIF) については、最小が 3.397，最大が 5.616 となり、サンプリング方法はあまり効率的でない可能性がある。

図表による詳細なビンパラメータの推定結果については、紙面の都合上、40 ビンごとを抽出して表 2-4 に「設定値」「事後平均」「標準誤差」「95%信用区間の下限と上限」^{*30}、「CD」「NIF」^{*31}「乖離率」を記載、図 2-6 に「標本経路」「事後確率密度」「標本自己相関」を図示する。

抽出した事例の推定結果については、事後平均値が設定値よりも、やや小さく推定される傾向があり、その乖離率の最小は 0.1%，最大は 7.5%，平均は 3.7%と高い精度が確認された。図 2-6 から、標本時系列をプロットしたビンパラメータは状態空間を十分にサンプリングしており安定的に推移しているように観測されるが、CD については表 2-4 より $\pi_1^{199}, \pi_2^{199}, \pi_3^{40}, \pi_3^{160}, \pi_4^{199}, \pi_5^{80}$ の 6 ビンが受容され、その割合は 24.0%(=6/25)であり、棄却されたパラメータの標本時系列については、条件付事後分布へ収束していない可能性がある。NIF については、表 2-4

*29 収束診断 Convergence Diagnostics (CD) については Appendix 2.(3)を参照のこと。

*30 信用区間については Appendix 2.(1)を参照のこと。

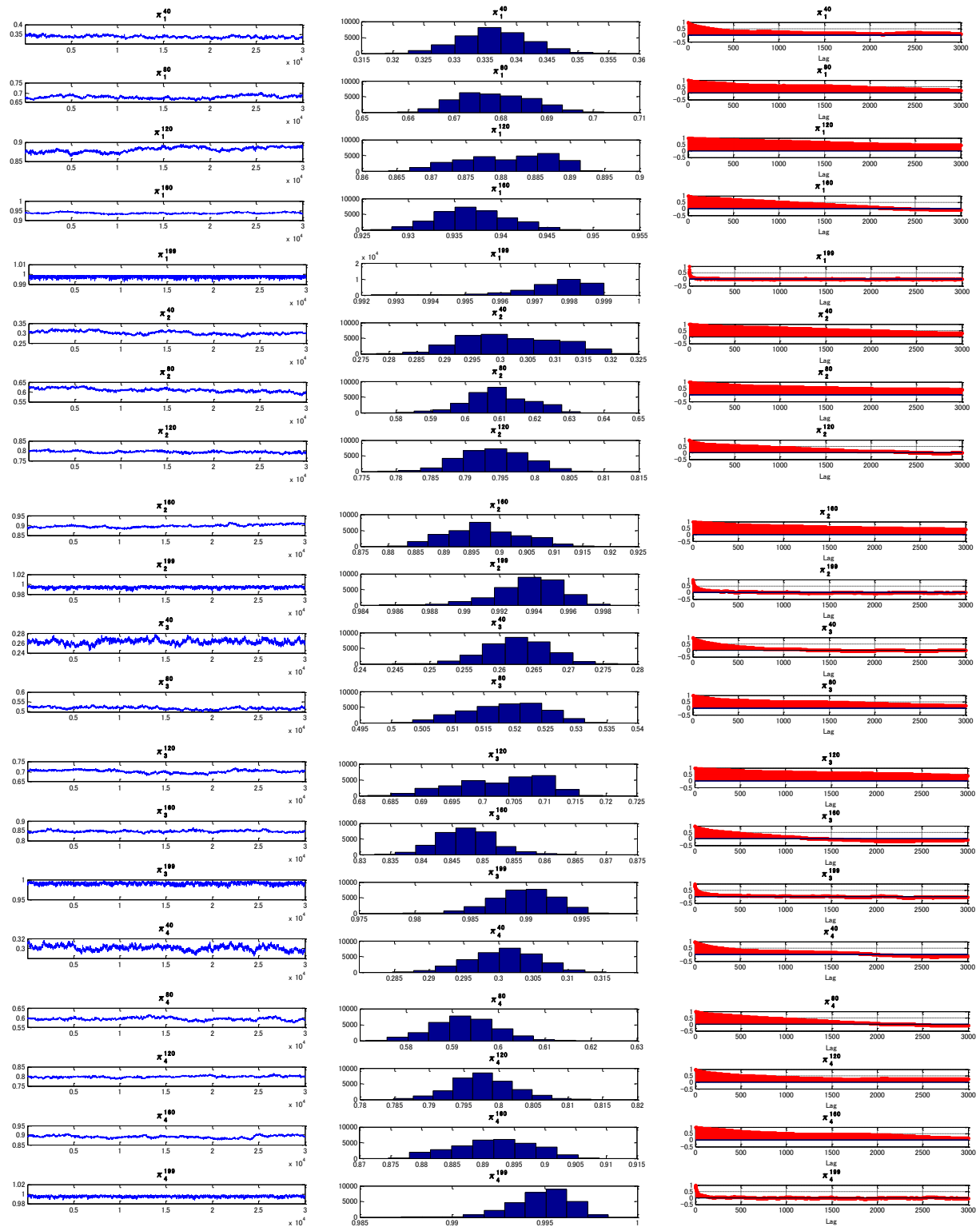
*31 NIF については Appendix 2.(2)を参照のこと。

から最小が 3.397, 最大が 5.601 となり, また, 標本自己相関について, 図 2-6 から総じて, 急速な減衰が観測されず, サンプル方法はあまり効率的でない可能性がある. 事後確率密度の形状については, 特に問題は観察されない.

全てのビンの事例と抽出した事例を比較すると, CD の受容割合に差異があるが, その他は概ね同水準の結果であり, 後者により前者の推定結果の大まかな状況が推測できるものと思われる.

表 2-4. Estimation results of bin parameters

	Bin Number	Set Value	Posterior Mean	Standard Error	2.5%	97.5%	CD	NIF	Divergence Ratio (%)
Cell 1	π_1^{40}	0.360	0.336	0.0002	0.326	0.347	21.674	5.437	6.5
	π_1^{80}	0.720	0.679	0.0002	0.665	0.693	-9.370	5.551	5.8
	π_1^{120}	0.920	0.881	0.0002	0.869	0.890	-48.762	5.601	4.3
	π_1^{160}	0.960	0.937	0.0001	0.930	0.944	-2.552	5.526	2.4
	π_1^{199}	0.999	0.998	0.0001	0.996	0.999	-0.106	3.397	0.1
Cell 2	π_2^{40}	0.320	0.302	0.0003	0.288	0.317	26.559	5.579	5.7
	π_2^{80}	0.640	0.610	0.0003	0.591	0.626	41.878	5.577	4.7
	π_2^{120}	0.840	0.794	0.0001	0.7845	0.803	28.410	5.436	5.5
	π_2^{160}	0.920	0.897	0.0002	0.885	0.910	-37.623	5.580	2.5
	π_2^{199}	0.998	0.994	0.0001	0.990	0.997	-0.995	4.653	0.4
Cell 3	π_3^{40}	0.280	0.263	0.0001	0.254	0.271	0.719	5.409	6.2
	π_3^{80}	0.560	0.518	0.0002	0.506	0.529	36.315	5.521	7.4
	π_3^{120}	0.760	0.703	0.0002	0.689	0.714	24.118	5.565	7.5
	π_3^{160}	0.880	0.847	0.0001	0.839	0.857	1.202	5.428	3.7
	π_3^{199}	0.997	0.990	0.0001	0.984	0.995	2.974	4.666	0.7
Cell 4	π_4^{40}	0.320	0.301	0.0001	0.292	0.310	13.025	5.448	6.0
	π_4^{80}	0.640	0.593	0.0002	0.579	0.606	3.430	5.540	7.4
	π_4^{120}	0.840	0.798	0.0001	0.789	0.806	-19.544	5.459	5.0
	π_4^{160}	0.920	0.892	0.0002	0.880	0.903	12.293	5.552	3.1
	π_4^{199}	0.998	0.995	0.0001	0.992	0.997	-1.343	4.531	0.3
Cell 5	π_5^{40}	0.280	0.282	0.0003	0.264	0.298	-5.566	5.585	0.8
	π_5^{80}	0.560	0.541	0.0002	0.527	0.554	-1.000	5.550	3.4
	π_5^{120}	0.760	0.748	0.0002	0.738	0.758	5.252	5.547	1.5
	π_5^{160}	0.880	0.862	0.0001	0.853	0.870	-10.030	5.494	2.1
	π_5^{199}	0.997	0.995	0.0001	0.991	0.998	-7.147	4.453	0.2



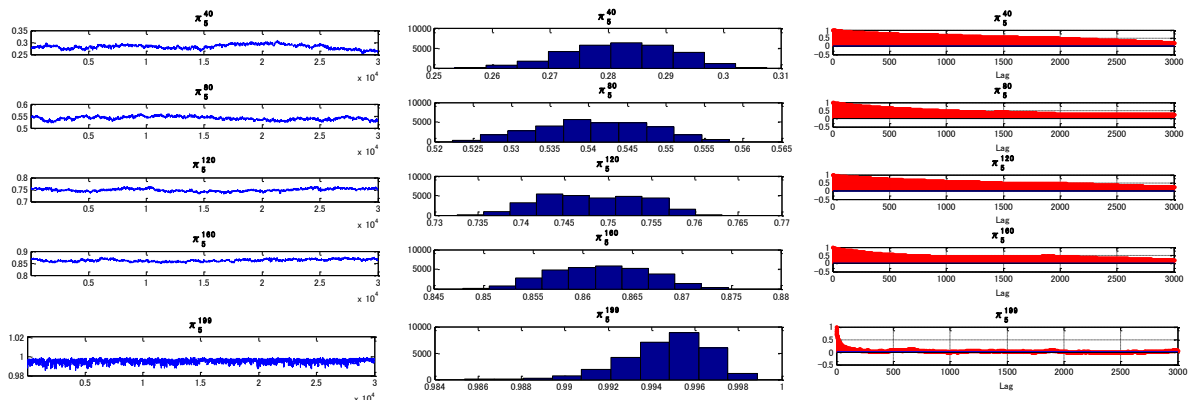


図 2-6. Sample path, posterior probability density, and sample autocorrelation for bin parameters.

次に，MH 法により，1000 個の潜在変数 $U_{i,j}$ をサンプリングする．1 回の試行について，全ての j について，データ番号 i ごとに採択棄却が実行され，その結果は表 2-5 とおりである．全データ番号における全試行について，採択確率は，最小が 48.0%，最大が 89.9%，平均が 78.2% と良好であり，提案分布 (2.12) 式が適切に設定されていると考えられる．

表 2-5. Acceptance rate of latent variable $U_{i,j}$

	Maximum	Minimum	Average
Acceptance Ratio(%)	89.9	48.0	78.2
Data Number	38	372	

最後に，D-vine t コピュラの相関パラメータ，自由度については，Besag et al. [9] に従い 20～80% の採択確率となるように，提案分布 (2.15)，(2.18) 式のハイパーパラメータ $\tau_{\rho_{j,k}}, \tau_v$ を表 2-6 に，「初期値」「設定値」「事後平均」「標準誤差」「95%信用区間の下限と上限」「CD」「NIF」「採択確率」「乖離率」を表 2-7 に記載，「標本経路」「事後確率密度」「標本自己相関」を図 2-7 に図示する．なお，周辺分布 j, k の順位相関(ケンドールの τ)を「初期値」とする．

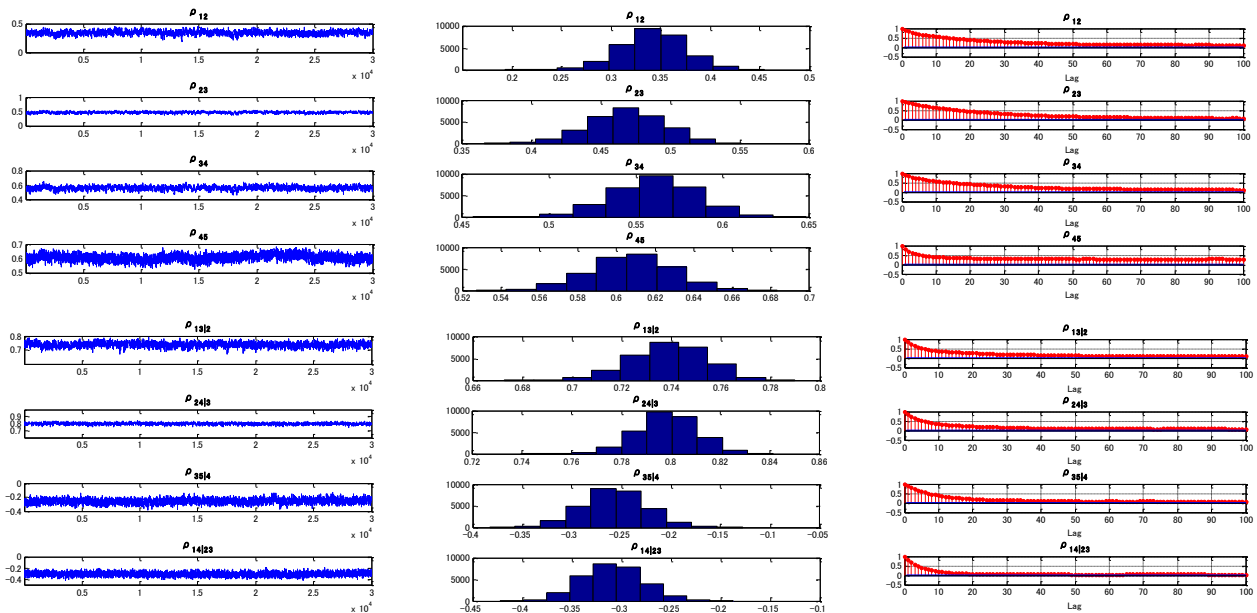
表 2-6. Hyperparameter.

	$\tau_{\rho_{12}}$	$\tau_{\rho_{23}}$	$\tau_{\rho_{34}}$	$\tau_{\rho_{45}}$	$\tau_{\rho_{13 2}}$	$\tau_{\rho_{24 3}}$	$\tau_{\rho_{35 4}}$	$\tau_{\rho_{14 23}}$	$\tau_{\rho_{25 34}}$	$\tau_{\rho_{15 234}}$	τ_v
Standard Deviation	0.1	0.1	0.2	0.2	0.3	0.3	0.1	0.2	0.3	0.3	1

表 2-7. Estimation results of D-vine t copula parameters.

	Initial Value	Set Value	Posterior Mean	Standard Error	2.5%	97.5%	CD	NIF	Acceptance Ratio (%)	Divergence Ratio (%)
ρ_{12}	0.199	0.3	0.343	0.0008	0.279	0.405	-2.193	4.372	53.2	14.2
ρ_{23}	0.287	0.4	0.468	0.0007	0.413	0.518	-1.273	4.581	34.3	17.0
ρ_{34}	0.360	0.5	0.561	0.0006	0.512	0.606	-3.029	4.428	29.6	12.1
ρ_{45}	0.421	0.6	0.608	0.0005	0.565	0.648	0.490	3.882	30.6	1.3
$\rho_{13 2}$	0.554	0.7	0.739	0.0003	0.708	0.767	0.321	3.772	23.2	5.6
$\rho_{24 3}$	0.624	0.8	0.799	0.0003	0.776	0.822	9.177	3.681	18.4	0.2
$\rho_{35 4}$	0.115	-0.2	-0.257	0.0007	-0.318	-0.194	-2.828	3.838	76.4	28.5
$\rho_{14 23}$	0.169	-0.3	-0.309	0.0006	-0.368	-0.247	2.717	3.135	66.9	3.0
$\rho_{25 34}$	0.212	-0.4	-0.409	0.0005	-0.463	-0.350	1.923	2.990	54.5	2.2
$\rho_{15 234}$	-0.143	-0.5	-0.523	0.0004	-0.571	-0.472	2.462	2.625	47.0	4.7
ν	10.0	7.0	6.288	0.0118	5.200	7.698	0.185	3.199	50.8	10.2
Average of Divergence Ratio(%)									9.0	

D-vine t コピュラのパラメータ推定値の乖離率は最小 0.2%，最大 28.5%とばらつきがみられるが，平均は 9.0%と概ね良好な推定結果が確認された．標本時系列は状態空間を十分にサンプリングして安定的に推移しているように観測されるが，CD については，11 個のパラメータにおいて有意水準 5%の確率で $\rho_{23}, \rho_{45}, \rho_{13|2}, \rho_{25|34}, \nu$ が受容され，その割合は 45.5%(=5/11)で，棄却されたパラメータの標本時系列については，条件付事後分布へ収束していない可能性がある．NIF については，表 2-7 から最小が 2.625，最大が 4.581 となり，また，図 2-7 から $\rho_{12}, \rho_{23}, \rho_{34}, \rho_{45}$ の標本自己相関については，急速な減衰が観測されず，サンプリング方法はあまり効率的でない可能性がある．事後確率密度の形状については，単峰形で正規分布に近い形状であり，特に問題は観察されない．



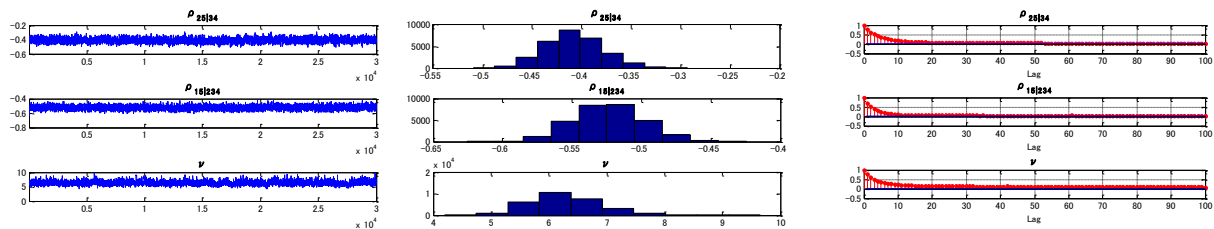


図 2-7. Sample path, posterior probability density, and sample autocorrelation for D-vine t copula parameters.

2.4.3. モデルの選択とリスク量の比較

リスク量の計算については5計量単位で、依存構造は2.4.1項(1)とする。まず、D-vine tコピュラの依存構造に従う一様乱数(0,1)を生成する。次に、生成した一様乱数に対し、各計量単位の損失分布の分位点を計算し、その損失金額を合算したものを計量単位全体の損失金額とし、この試行を1000回繰り返す。そのパーセントタイルをリスク量とする。ここで、リスク量は95, 99, 99.9パーセントタイルの場合を考える。なお、損失金額データについては、200ビンの離散分布を真の分布として、2.4.1項(2)の手法により生成する。

(1) 離散分布の分布型の違いによるリスク量の比較

離散分布における分布型の違いをビン数の違いとし、100ビンと50ビンの事例を考える。まず、200ビンから生成した損失金額データから、本章で提案した手法により100ビンと50ビンの場合のビンのパラメータ(ビンの累積確率)を推定する。なお、試行回数は2万回で、最初の1万回は捨て、その後の1万回を標本とする。次に、各計量単位の損失金額を2.4.1項(1)、2.4.1項(2)の手法により生成する。最後に、各計量単位の合算した損失金額を生成し、リスク量を推定する。

表 2-8 から離散分布において、分布型が異なっている場合でもリスク量は安定的に推定されている。

表 2-8. Comparison of the percentile between different discrete distribution.

Num of Bins /Percentile	95%	99%	99.9%
200(True Value)	34.43	40.71	44.25
100	36.06	42.45	48.84
50	35.84	43.23	46.31

(2) 連続分布のモデルならびに推定法の違いによるリスク量の比較

先行研究で採用されているパラメトリックな連続分布のモデルとして、5計量単位の分布に対数正規、ガンマ分布を選択する。まず、200ビンから生成したデ

ータから、各計量単位の連続分布のパラメータを最尤法、モーメント法、最小2乗法によって推定する。最尤法については戸坂・吉羽[41]を、モーメント法、最小2乗法については三菱信託銀行[29]を参照する。次に、各計量単位の損失金額を生成する。さらに、各計量単位の合算した損失金額を生成し、最後に、リスク量を推定する。

①計量単位の分布のパラメータ推定

表 2-9 のとおり、推定手法が異なることで、各計量単位のパラメータ推定結果において差異が確認できる。

表 2-9. Parameter estimation results.

Dist,Mthd / Cell	Lognormal Dist.					
	MLE		MM		LSM	
Cell	μ	σ	μ	σ	μ	σ
1	0.768	0.958	0.897	0.643	0.910	0.815
2	0.850	1.129	1.031	0.652	1.032	0.841
3	1.023	1.034	1.190	0.626	1.176	0.883
4	0.848	1.066	1.019	0.647	1.029	0.833
5	0.971	1.075	1.156	0.626	1.158	0.866

Dist,Mthd / Cell	Gamma Dist.					
	MLE		MM		LSM	
Cell	α	β	α	β	α	β
1	1.637	1.842	1.952	1.544	1.763	1.775
2	1.411	2.459	1.884	1.842	1.638	2.194
3	1.525	2.622	2.086	1.916	1.513	2.823
4	1.458	2.342	1.926	1.774	1.670	2.135
5	1.458	2.651	2.086	1.852	1.537	2.702

ここで、MLEは最尤法、MMはモーメント法、LSMは最小2乗法である。また、対数正規分布、ガンマ分布の密度関数は(2.24)、(2.25)式である。

$$(2.24) \quad f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right)$$

$$(2.25) \quad f(x|\alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), \quad \Gamma(\alpha) \text{はガンマ関数}$$

②リスク量

表 2-10 から、第1に、計量単位の分布に対数正規分布とガンマ分布を仮定した場合には、その仮定する分布形および選択するパラメータ推定手法によりリスク量は大きく変動して、安定していないことが確認される。例えば、99.9%

分位点において，対数正規分布を仮定した事例で最尤法を用いた場合に 252.88，ガンマ分布を仮定した事例で最小 2 乗法を用いた場合に 85.89 とリスク量が最大となっており，また両者は大きく乖離している．第 2 に，ビン数 200 のリスク量（真値）と対数正規分布，ガンマ分布を仮定した場合のリスク量を比較すると，後者は過大に推定されていると考えられる．

表 2-10. Amount of risk for probability distribution.

Percentile	Est Mthd	Lognormal Dist.	Gamma Dist.	Discrete Dist.(200Bins)
95%	MLE	58.83	38.27	34.43
	MM	35.53	35.94	
	LSM	49.82	39.41	
99%	MLE	129.03	55.98	40.71
	MM	55.57	50.49	
	LSM	90.37	58.25	
99.9%	MLE	252.88	83.98	44.25
	MM	87.29	72.27	
	LSM	165.10	85.89	

ヒストリカル法では過去の実績に基づく経験分布を想定しており，最大値以上のデータについて確率が付与されていないという問題があるが，オペリスクの損失データ分析においては，損失金額分布のボディ部分に相当する内部損失金額データとテール部分に相当するシナリオ損失金額データを網羅している状況であると考えられる．また，分布形や推定手法によるリスク量の大きな変動は，リスクの過大評価，過小評価の原因となる．従って，上記の（1），（2）の結果から，提案した手法は，リスク量が安定的に推定されること，および，その精度から，オペリスク計量手法として有用であると考えられる．

2.5. 2章のまとめと今後の課題

本章では，経営に重大な影響を与える低頻度高額損失のシナリオ損失金額を含め，各計量単位の依存性が反映されたオペリスク損失金額分布を推定するために，周辺分布の離散分布による近似と D-vine t コピュラによるモデル化を提案した．パラメータ推定について，データ数を n ，計量単位を m とすると (2.1)，(2.2) 式からコピュラの尤度関数は $n2^m$ 項数で記述され，データ数が多く，計量単位の次数が高い場合は最尤法による推定は現実的でないと考えられる．従って，Smith and Khaled[39]のデータ拡大法によるベイズ手法によるサンプリングを参照した．離散分布で近似したビンのパラメータについては計量単位の依存性，隣接するビンの累積確率の大小関係の制約を反映させるために，ビンの累積確率に順序統計量の同時分布を事前分布として選択し，条件付事後分布の定義関数の条件を充足する一様乱数のサンプリングを提案した．D-vine t コピュラのパラメータのサンプリングについては，提案分布に切断正規分布を用いた Min and

Czado[25]を参照した。

提案した周辺分布および依存構造のモデル，パラメータ推定手法について，モデルのパラメータの推定精度，ならびに推定されたリスク量の安定性を考慮して実務への応用可能性が示された．パラメータの推定精度については，数値実験によって検証され，推定値は設定したパラメータが高い精度で再現された．従って，実務において本章で示した損失金額データの離散近似方法，パラメータ推定手法を採用することで，シナリオを含む全ての損失金額データを対象に計量単位ごとに，その依存性を考慮したビンが推定され，D-vine t コピュラによってモデルを構築することが期待される．推定されたリスク量の安定性については，損失金額データが損失金額分布のボディ部分に相当する内部損失金額データとテール部分に相当するシナリオ損失金額データを網羅している状況であること，また，一例ではあるが先行研究の手法とのリスク量を比較し，より安定的に高い精度でリスク量が推定されることから，提案した手法が有用であることが確認できた．

今後の課題として，次の4点が挙げられる．第1に，ビン，D-vine t コピュラのパラメータ推定結果のCDからは標本の条件付事後分布へ収束していない可能性がある事例，NIFからはサンプリングが効率的に行われていない可能性がある事例があり，周辺分布のサンプリング手法によるものか，試行回数によるものか，データ数と推定パラメータ数によるものであるかなど，原因究明が必要である．第2に，数値実験として5次元計量単位を取り上げたが，より高次元計量単位を対象とする場合には，Smith and Khaled[39]における計量単位間の独立性の判定，Min and Czado[26]におけるReversible Jump MCMCによるモデル選択など，依存性のある計量単位の次数選択や異なるコピュラ関数の選択問題などに取組む必要がある．第3に条件付きコピュラの評価を無条件コピュラで近似しているが，推定精度向上のため数値計算手法の検討が必要である．最後に，ビンのパラメータ推定において，ビン幅，ビン数を既知としているが，同時推定についても検討が必要である．

Appendix 2. 用語の説明（信用区間，NIF，CD）

（1）信用区間

標本の信用区間はパラメータについて区間を用いて要約する方法で以下のように定義される．

100(1- α)%信用区間とは，パラメータ ϕ が $\Pr\{\phi < a_\alpha\} = \alpha/2$ ， $\Pr\{\phi > b_\alpha\} = \alpha/2$ を満たす区間 (a_α, b_α) である．

（2）NIF

サンプリング方法の効率性の指標による診断する方法として，NIFが知ら

れており (2A. 1) 式で定義される.

$$(2A.1) \quad NIF_L \equiv \sqrt{1 + 2 \sum_{l=1}^L \left(1 - \frac{l}{L+1}\right) r_l}$$

ここで, r_l はラグ l の標本自己相関関数, L はサンプリング回数の 3 乗根を超えない最大の整数値である.

目視による診断方法としては, 標本自己相関関数のプロットが知られている. これは, 標本自己相関の減衰状況からサンプリングの効率性が診断され, 例えば, 急激に減衰している場合, このサンプリングは独立な確率標本に近いことを示唆していると解釈される.

(3) CD

Geweke[18]によって提唱された平均値の差の検定を応用した指標で (2A. 2) 式で定義される.

$$(2A.2) \quad CD \equiv \frac{m_{T_1} - m_{T_2}}{\sqrt{NIF_{T_1}^2 s_{T_1}^2 / T_1 + NIF_{T_2}^2 s_{T_2}^2 / T_2}}$$

ここで, T_1 はサンプリング全体の前半部分, T_2 は後半部分の回数, m_{T_j} は標本平均, $s_{T_j}^2$ は標本分散, NIF_{T_j} は標本回数 T_j の NIF である.

また, MCMC による標本が定常分布に収束している場合, CD 統計量の漸近分布は標準正規分布になることが知られている.

目視による診断方法としては, 標本経路の作成が有用であることが知られている. これは, その変動が初期値に依存せず安定的な動作になっているかどうかで不変分布に収束したかどうかを診断する方法である.

2 章参考文献

- [1] Aas, K. and Berg, D., Models for construction of multivariate dependence: A comparison study, *The European Journal of Finance*, **15** (2009), 639-659.
- [2] Aas, K., Czado, C., Frigessi, A. and Bakken, H., Pair-copula Constructions of Multiple Dependence, *Insurance: Mathematics and Economics*, **44** (2009), 182-198.
- [3] Badescu, A. L., Gong, L., Lin X. S. and Tang D., Modeling correlated frequencies with application in operational risk management, *Journal of Operational Risk*, **10** (2015), 1-43.
- [4] バーゼル銀行監督委員会 (Basel Committee on Banking Supervision), *Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework*, 2004.

<https://www.bis.org/publ/bcbs107.htm>

- [5]バーゼル銀行監督委員会 (Basel Committee on Banking Supervision). Results from the 2008 Loss Data Collection Exercise for Operational Risk, 2009.
<http://www.bis.org/publ/bcbs160.htm>
- [6]Bedford, T. and Cooke, R. M., Vines — A new graphical model for dependent random variables, *The Annals of Statistics*, **30** (2002), 1031–1068.
- [7]Behrens, C. N. Lopes, H.F. and Gamerman D., Bayesian analysis of extreme events with threshold estimation, *Statistical Modelling*, **4** (2004), 227-244.
- [8]Berg D., Copula goodness-of-fit testing: An overview and power comparison, *The European Journal of Finance*,15(2009),675-701.
- [9]Besag, J., Green, P., Higdon, D. and Mengersen, K. Bayesian Computation and Stochastic Systems, *Statistics Science*, **10**(1995), 3-66.
- [10]Brechmann, E.C., Czado, C. and Paterlini, S., Modeling dependence of operational loss frequencies, *Journal of Operational Risk*, **8** (2013), 105-126.
- [11]Brechmann, E.C., Czado, C. and Paterlini, S., Flexible Dependence Modeling of Operational Risk Losses and Its Impact on Total Capital Requirements, *Journal of Banking & Finance* **40** (2014), 271-285
- [12]Chapelle, A., Crama, Y., Hübner, G. and Peters, J.-P., Basel II and Operational Risk: Implications for risk measurement and management in the financial sector, Working Paper No.51, National Bank of Belgium, 2004.
<http://www.nbb.be/doc/ts/publications/WP/WP51En.pdf>
- [13]Chib, S. and Greenberg, E., Understanding the Metropolis-Hastings Algorithm, *The American Statistician*, **49** (1995), 327-335.
- [14]Dutta, K., and Perry, J., A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital, Working Paper No. 06-13, Federal Reserve Bank of Boston, 2006.
<http://www.bos.frb.org/economic/wp/wp2006/wp0613.pdf>
- [15]Fischer, M., Köck, C., Schlüter, S. and Weigert, F., An empirical analysis of multivariate copula models, *Quantitative Finance*, **9** (2009), 839-854.
- [16]de Fontnouvelle, P., Jordan, J. and Rosengren E., Implications of Alternative Operational Risk Modeling Techniques, Working Paper No. W11103, National Bureau of Economic Research, 2005.
<https://www.nber.org/papers/w11103>
- [17]Genest, C. and Nešlehová, J., A primer on copulas for count data, *Astin Bulletin*, **37** (2007), 475-515.
- [18]Geweke, J., Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, In: Bernardo, J. M. Berger, J. O. Dawid, A. P. and

- Smith, A. F. M. (eds.), *Bayesian Statistics 4*, Oxford University Press, 1992, 169-193.
- [19]Haff, I. H., Aas, K. and Frigessi, A., On the simplified pair-copula construction Simply useful or toosimplistic?, *Journal of Multivariate Analysis*, **101** (2010),1296-1310.
- [20]岩永育子, ヴァイン・コピュラを用いた CAPM の非正規・非線形への拡張: 日本株式市場における実証分析, In: 日本金融・証券計量・工学学会(編), *ファイナンスとデータ分析, ジャフィ・ジャーナル: 金融工学と市場計量分析*, 2015, 69-113.
- [21]小林孝明, 清水真一郎, 西口健二, 森永聡, *オペレーショナル・リスク管理 高度化への挑戦*, きんざい, 2009.
- [22]金融庁, 銀行法第十四条の二の規定に基づき, 銀行がその保有する資産等に照らし自己資本の充実の状況が適当であるかどうかを判断するための基準, 告示第十九号, 2006.
http://www.fsa.go.jp/policy/basel_ii/01.pdf
- [23]金融庁, 最終化されたバーゼルⅢの国内実施に関する告示改正案の概要, 2021
<https://www.fsa.go.jp/news/r3/ginkou/20210928-2/000.pdf>
https://www.boj.or.jp/announcements/release_2016/data/rel160307a1.pdf
- [24]Kurowicka, D. and Joe, H. (eds.), *Dependence Modeling: Vine Copula Handbook*, World Scientific, Singapore, 2011.
- [25]Min, A. and Czado, C., Bayesian Inference for multivariate copulas using pair-copulas constructions, *Journal of Financial Econometrics*, **8** (2010), 511-546.
- [26]Min, A. and Czado, C., Bayesian model selection for D-vine pair-copula constructions, *The Canadian Journal of Statistics*, **39** (2011), 239-258.
- [27]Mittnik, S., Paterlini, S. and Yener, T., Operational risk dependencies and the determination of risk capital, *Journal of Operational Risk*, **8** (2013), 83-104.
- [28]Mittnik, Stefan and Yener, Tina, Estimating operational risk capital for correlated, rare events, *Journal of Operational Risk*, **4** (2009), 29-51.
- [29]三菱信託銀行オペレーショナルリスク研究会, *オペレーショナル・リスクのすべて*, 東洋経済新報社, 2002.
- [30]森淳稔, 木全友則, 長藤剛, 損失額分布やパラメータ推定手法の選択がオペレーショナルリスク計量結果に与える影響について, 日本銀行金融機構局, 2007.
http://www.boj.or.jp/research/brp/ron_2007/data/ron0706a.pdf
- [31]森本祐司, 金融と保険の融合について, *金融研究*, **19**, (2000), 289-342.
<https://www.imes.boj.or.jp/research/papers/japanese/kk19-b1-8.pdf>
- [32]Moscadelli, M., The modeling of operational risk: experience with the analysis of

- the data collected by the Basel Committee, Working Paper (Temi di discussione) No. 517, Bank of Italy, 2004.
<https://www.bancaditalia.it/pubblicazioni/temi-discussione/2004/2004-0517/index.html>
- [33]Müller, P., A generic approach to posterior integration and Gibbs sampling, Technical report No. 91-09, Purdue University, 1991.
<https://www.stat.purdue.edu/docs/research/tech-reports/1991/tr91-09.pdf>
- [34]Nakatsuma, T., Bayesian Risk Assessment with Threshold Mixture Extreme Value Models, JAFEE 2010 冬季大会 予稿集, 2010.
- [35]Nelsen, R. B., An Introduction to Copulas, Second Edition, Springer, 2006.
- [36]日本銀行, オペレーショナル・リスク管理の高度化, 2005.
https://www.boj.or.jp/research/brp/ron_2005/fsk0507a2.pdf
- [37]オペレーショナルリスク管理高度化勉強会, オペレーショナルリスク管理の更なる高度化に関する議論(1), (2), オペレーショナルリスク管理高度化勉強会, 2006.
http://www.boj.or.jp/research/other_release/fsc0606a.pdf
http://www.boj.or.jp/research/other_release/fsc0608e.pdf
- [38] Panagiotelis, A., C. Czado and H. Joe, Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association*, **107** (2012), 1063-1072.
- [39]Smith, M. S. and Khaled M.A. , Estimation of Copula Models with Discrete Margins via Bayesian Data Augmentation, *Journal of the American Statistical Association*, **107** (2012), 290-303.
- [40]塚原英敦, 接合分布関数(コピュラ)の理論と応用, In: 国友直人, 山本拓(監修), 北川源四郎, 竹村彰通(編), 21世紀の統計科学Ⅲ, 2011, 111-139.
- [41]戸坂凡展, 吉羽要直, コピュラの金融実務での具体的な活用方法の解説, *金融研究*, **24** (2005), 115-162.
- [42]Valle, L., Bayesian Copulae Distributions, with application to Operational Risk Management, *Methodology and Computing in Applied Probability*, **11** (2009), 95-115.
- [43]吉羽要直, ストレス状況を勘案した相関構造とリスク合算, *日本銀行ワーキングペーパーシリーズ*, **13-J-4** (2013).
https://www.boj.or.jp/research/wps_rev/wps_2013/wp13j04.htm/

3. 離散型データにおける D-vine t コピュラで構成されたモデルのモデル選択

3.1. はじめに

コピュラは多くの研究で、多変量金融資産ポートフォリオの評価、リスク管理などにおいて周辺分布間の依存構造のモデル化に用いられている。そのほとんどの場合、パラメトリックなコピュラは連続型データにおいて用いられており、コピュラおよび周辺分布のパラメータは最尤法を用いて推定されている事例が多い。^{*32}一方で、離散型データの場合、Genest and Nešlehová[7]は、事例を挙げた解説で、周辺分布の依存性をコピュラでモデル化する際に、連続型データとして疑似尤度によりパラメータを推定すると、両者の推定値は大きく乖離する場合があることを指摘している。Smith and Khaled[20]は、Nelsen[18]のコピュラの差分による確率関数の定義を引用して以下の点を指摘した。まず、離散型データにおいて、その依存構造をコピュラでモデル化し、 n をデータ数、 m をデータの次元とすると、その同時分布の尤度関数の項数は、 $n2^m$ ^{*33}と指数的に増加することから、データ数が多く、高次元の場合には膨大となることである。次に、高次元でない場合でも、最尤法において最適化の計算が複雑となり、周辺分布とコピュラのパラメータの同時推定は困難であることである。そこで、潜在変数をデータ拡大法で生成するマルコフ連鎖モンテカルロ (MCMC) 法を提案している。

コピュラでは、多変量コピュラとそれをペアコピュラに分解した Vine コピュラ^{*34}が知られている。Bedford and Cooke[3]によって提案された Vine コピュラは、多変量コピュラの依存構造をペアコピュラに分解したもので、結合形態の違いで R-vine, その特殊ケースの C-vine, D-vine に分類される。C-vine, D-vine はコピュラの構造が簡単に記述でき、その特徴について、岩永[10]では「C-vine は特定の変数が各変数間の相互作用の中心となることが明らかな場合にメリットがあり」、塚原[22]では「D-vine から構成される分布は各変数が同等に扱われている」と紹介されている。

Vine コピュラの課題は、その分解されたペアコピュラの項数が、 d 次元の場合、 $d(d-1)/2$ と 2 次的に増加することである。その取り組みとして、連続型データの場合、Min and Czado[17]では、RJMC MC 法により、Smith et al.[21]では、ペアコピュラが独立、独立でない場合のダミー変数を導入し、MCMC 法によ

*32 コピュラ理論全般については、Joe[11], Nelsen[18], パラメータの最尤法による推定については、戸坂・吉羽[23], Mashal and Naldi[15]が詳しい。

*33 詳細は 2.2 節(1)を参照のこと。

*34 Vine コピュラについては、Kurowicka and Joe[14], Aas et al.[2]が詳しい。

り、独立コピュラ*³⁵を検出するモデル選択*³⁶手法を提案した。離散型データの場合では、Smith and Khaled[20]が潜在変数をデータ拡大法により生成し、Smith et al. [21]と同様に、ダミー変数を用いる手法を提案した。ここで、「モデル選択」とは、独立または条件付独立のコピュラ(独立コピュラ)を検出し、独立コピュラでないペアコピュラで構成される各モデルの事後確率から最適なモデルを選択することである。これらの研究により、分解されたペアコピュラの項数が減少することで、計算負荷の軽減が期待できる。

RJMCMC 法は、Green[9]によって考案された手法で、異なる複数のモデルに対応する異なった次元のパラメータ部分空間をサンプラーがジャンプするようにMCMC 法を拡張し、モデルのパラメータおよびモデルの同時事後確率密度からサンプリングを行うことが可能となり、モデルの不確実性を分析することができる方法である。その応用として、Min and Czado[17], 粕谷・高川[12]などが存在し、前者は多変量スワップレートの依存構造を D-vine t コピュラでモデル化し独立コピュラを検出、後者は 17 種類の円ドル実質為替レート予測がモデル化され、いずれの場合も、連続型データの場合に、RJMCMC 法によりモデル選択とパラメータを推定した実証分析事例である。

代表的なペアコピュラとして、正規コピュラと t コピュラが知られている。前者の依存性は上側、下側のいずれも漸近独立であり、分布の裾の依存を考える場合、同時的な大幅変動の事象を十分に表現できないことを示唆している。後者の裾依存性については、上側、下側のいずれも漸近依存しており、分布の裾でも強い依存関係を表現することができる。

コピュラの選択について、従来の多変量コピュラと D-vine コピュラの比較に関する研究では、川口・山中・田代[13]によれば、前者は全てのペアの順位相関関係を 1 種類のパラメータで調整することになり、複雑な依存構造を表現できないが、後者はペアごとに異なる依存構造を設定することができ、周辺分布のペアの順位相関関係をより柔軟に表現できること、また、Aas and Berg[1], Fischer et al. [6] では、D-vine t コピュラは、例えば、階層構造多変量アルキメディアンや多変量 t コピュラよりも多変量金融データへの適合度*³⁷が高いことを報告している。

研究目的は、離散型の裾の厚い多変量金融データの分析において、コピュラの構造が簡単で、特定の変数が相互作用の中心ではない、複雑な依存構造、裾依存性を反映できる Vine コピュラを採用し、ペアコピュラの組合せから独立コピュラを探索することによるモデル選択およびパラメータ推定手法を提案するこ

*³⁵ 独立コピュラの定義については 3.3.2 項(1)を参照のこと。

*³⁶ RJMCMC 法によるモデル選択の詳細については 3.3.2 項(6)を参照のこと。

*³⁷ コピュラに関する適合度検定手法(goodness-of-fit testing)については Berg[4]に詳しい。

とにある。

本研究では、周辺分布に経験分布を仮定し、Min and Czado[17]が提案した連続型データにおけるRJCMC法によるモデル選択手法とSmith and Khaled[20]が用いたデータ拡大法による潜在変数の生成手法を併用し、離散型データにおけるモデル選択およびパラメータの推定手法を提案する。また、提案した手法の精度検証のための数値実験では、まず、3種類のVineコピュラからD-vine tコピュラを選択し、設定したペアコピュラの組合せとパラメータから、Aas et al.[2]のアルゴリズムにより、周辺分布の依存構造を反映した一様乱数(0,1)のを生成する。次に、RJCMC法によりモデルを選択およびパラメータを推定し、当初設定したD-vine tコピュラの組合せとそのパラメータが再現されることを示すことにより、提案した手法の精度を検証する。

2章と本章との成果の違いと関連性について以下に整理する。

まず、成果の違いについて述べる。2章では、第1に、オペリスクの損失金額データの周辺分布を離散分布で近似した。第2に、Smith and Khaled[20]において周辺分布に1個のパラメータの分布を前提としていたが、離散分布の構成するビン数にパラメータ数を拡張した。第3に、ビンのパラメータサンプリングにGibbs法を提案し、D-vine tコピュラのパラメータとの同時推定を行った。一方、本研究では、離散型データについて、RJCMC法により独立コピュラを検出することによるモデル選択ならびにD-vine tコピュラのパラメータを推定した。次に、その関係性について述べる。2章では2番目の課題として『RJCMC法によるモデル選択など、依存性のある計量単位の次数選択や異なるコピュラ関数の選択問題などに取組む必要がある。』が挙げられており、本章は、2章の課題に対する取り組みであり、オペリスク計測を想定している。

本章の構成は以下のとおりである。3.2節では、同時分布、周辺分布およびその依存性を表現するD-vine tコピュラを実装したモデル説明、3.3節では、D-vine tコピュラで構成されたモデルのベイズ手法によるモデル選択とパラメータ推定手法の説明、3.4節では、模擬データ生成方法、数値実験の説明、および、提案した手法が有効であることを示し、3.5節では、まとめと今後の課題を述べる。

3.2. モデル

(1) \mathbf{X} の同時分布

m 次元離散型データ \mathbf{X} の同時分布とコピュラ関数の関係を説明する。

2章と同一のモデルを設定しており、定義2.1 (2.1)式を参照のこと。

(2) \mathbf{X} の周辺分布

第 j 変量の離散型確率変数を \mathbf{X}_j , その実現値を $\mathbf{x}_j = (x_{i,j}, \dots, x_{n,j})$ とすると,

経験分布関数は(3.1)式のように表せる．ここで，第 j 変量の第 i 番目のデータを $x_{i,j}$ とする．

$$(3.1) \quad F(x_j) = \frac{1}{n} \sum_{i=1}^n I(x_j \leq x_{i,j}), \quad -\infty < x_{i,j} < \infty$$

$$\text{ただし, } I(x_j \leq x_{i,j}) = \begin{cases} 1, & (x_j \leq x_{i,j}) \\ 0, & (x_j > x_{i,j}) \end{cases} \text{である.}$$

ここで， $I(A)$ は定義関数であり， A が真の場合 $I(A)=1$ ，その他は $I(A)=0$ である．

(3) (\mathbf{X}, \mathbf{U}) の同時分布

データ拡大法を導入するために潜在変数 \mathbf{U} と離散型データ \mathbf{X} の (\mathbf{X}, \mathbf{U}) の同時分布を考える．

2章と同一のモデルを設定しており，定義2.2 (2.7)式を参照のこと．

(4) D-vine t コピュラ

2章と同一のモデルを設定しており，(2.8)式を参照のこと．

3.3. ベイジアン MCMC 法，RJCMC 法アプローチ

3.3.1. 潜在変数 \mathbf{U} の事後分布，提案分布，採択確率の導出

Smith and Khaled[20]では，D-vine t コピュラのパラメータのほか，データ拡大法により，条件付潜在変数 $U_j | U_1, \dots, U_{j-1}, \mathbf{X}$ をMH法によりサンプリングしており，事後分布，提案分布，採択確率を導出する．

まず， (U_1, \dots, U_j) の同時分布から， U_j の条件付事後分布を導出する．2章と同一のモデルを設定しており，提案分布は(2.12)式を参照のこと．また，同様に，採択確率は(2.14)式を参照のこと．

D-vine t コピュラパラメータ ρ, ν のサンプリングアルゴリズムについては，3.3.2項の(3)，(4)，(5)のSTAYおよびBIRTHを参照のこと．

3.3.2. RJCMC 法による独立コピュラの検出とパラメータ推定

本節では，Min and Czado[17]で提案された，RJCMC法による独立なD-vine t コピュラの検出方法の概略を述べる．そのアルゴリズムについては，Appendix 3を参照のこと．

(1) D-vine t コピュラによるモデル定式化

$$c_{i,j|i+li+j-1}(u_{i+li+j-1}, u_{i+j|i+li+j-1}) \equiv c_{i,j|i+li+j-1}(F_{i+li+j-1}(x_{i+li+j-1}), F_{i+j|i+li+j-1}(x_{i+j|i+li+j-1}))$$

2変量 $i, i+j$ の条件付コピュラ密度関数であり, そのインデックスは $i, i+j|i+1:i+j-1$ となる.

Min and Czado[17]によれば, すべての m 次元の依存構造が存在する場合に, $n_c \equiv m(m-1)/2$ 個のD-vine tコピュラの積で構成され, モデルベクトル M はインデックスを使って次式のように記述できる.

$$M \equiv \underbrace{(12, 23, \dots, (m-1)m)}_{(m-1)\text{個のペアコピュラ}}, \underbrace{13|2, \dots, (m-2)m|(m-1)}_{(m-2)\text{個のペアコピュラ}}, \dots, \underbrace{1m|2:(m-1)}_{1\text{個のペアコピュラ}} = (M_1, M_2, \dots, M_s, \dots, M_{n_c})$$

独立コピュラとは, Smith et al. [21], Min and Czado[17]によれば, D-vine tコピュラ密度関数が $c_{i, i+j|i+1:i+j-1}(u, v) = 1, \forall u, v \in (0, 1)$ のときであり, その分布関数は, $C(u, v) = uv, C(u|v) = h(u|v) = u$ となる.

モデルベクトル M の第 s 番目の要素である D-vine tコピュラが $c_{i, i+j|i+1:i+j-1}(u, v) = 1$ のとき, $M_s = i, i+j|i+1:i+j-1 \equiv 0, c_{i, i+j|i+1:i+j-1}(u, v) \neq 1$ のとき $M_s = i, i+j|i+1:i+j-1 \equiv 1$ と定義する.

ここで, $u \equiv F_{i|i+1:i+j-1}(x_{i|i+1:i+j-1}), v \equiv F_{i+j|i+1:i+j-1}(x_{i+j|i+1:i+j-1})$ である.

(2) 遷移種類とその確率

Min and Czado[17]によれば, モデル遷移は, まず, モデルベクトル M が D-vine tコピュラの構成に変化がない場合 (STAY), 1組のD-vine tコピュラが追加される場合 (BIRTH), 1組のD-vine tコピュラが削除される場合 (DEATH) のいずれかから無作為に選択される.

$p_M \equiv \sum_{s=1}^{n_c} I_{\{M_s \neq 0\}}$ とすると, モデル M から新たなモデル $M_s^{(new)}$ への遷移は,

$p_M = n_c$ のとき, STAY, DEATH, $1 < p_M < n_c$ のとき STAY, BIRTH, DEATH, $p_M = 1$ のとき STAY, BIRTH である.

遷移確率など詳細は, Appendix 3を参照のこと.

ここで, 現在のモデル, 遷移後のモデルを

$$M \equiv (M_1, \dots, M_s, \dots, M_{n_c}), \quad M^{(new)} \equiv (M_1^{(new)}, \dots, M_s^{(new)}, \dots, M_{n_c}^{(new)}) \text{ とする.}$$

① STAY

遷移前後のモデルベクトル $M, M^{(new)}$ のD-vine tコピュラ密度関数の構成要素は同じであるが, 全てのパラメータが更新される.

② BIRTH

$M_s = 0, M_s^{(new)} \neq 0, M_t = M_t^{(new)}, t \neq s$ のとき, モデルベクトル M の構成要素 M_s が $M_s^{(new)}$ に更新される. すなわち, 候補となる構成要素 M_s から無作為に選択され, 1組のD-vine tコピュラが追加される.

③ DEATH

$M_s \neq 0, M_s^{(new)} = 0, M_t = M_t^{(new)}, t \neq s$ のとき, モデルベクトル M の構成要素 M_s が $M_s^{(new)}$ に更新される. すなわち, 候補となる構成要素 M_s から無作為に選択され, 1組のD-vine tコピュラが削除される.

(3) ρ_{M_s}, ν_{M_s} の事前分布

Min and Czado[16]に従い，構成要素 M_s の D-vine t コピュラのパラメータである相関 ρ_{M_s} の事前分布は一様分布 $(-1,1)$ に従う．自由度 ν_{M_s} の事前分布は一様分布 $(1, \text{Upper})$ に従い，両者は独立であるとする．また，**Upper** はパラメータ ν_{M_s} の上限として，Min and Czado[16,17]に従い，**Upper** = 20 に設定した．

なお，上記の設定は，2章のStep3，Step4と同様である．

(4) ρ_{M_s}, ν_{M_s} の提案分布

STAY，BIRTHにおけるD-vine t コピュラのパラメータ ρ_{M_s}, ν_{M_s} の提案分布については，Min and Czado[17]に従い，切断正規分布を採用する．ハイパーパラメータについては，Besag, et Al. [5] に従い，1千回の先行サンプリングにおいて， $p_M = n_c$ (フルモデル) の場合，D-vine t コピュラのパラメータの採択確率が概ね20～80%になるように調整する．

① STAY

D-vine t コピュラパラメータはMin and Czado[16]に従い，MH-RW法によってサンプリングする．

構成要素 M_s のD-vine t コピュラパラメータの現時点，次回試行の標本を

$(\rho_{M_s}^{(old)}, \nu_{M_s}^{(old)}), (\rho_{M_s}^{(new)}, \nu_{M_s}^{(new)})$ として，各々切断正規分布TNを仮定すると， ρ_{M_s} の提案分布は平均 $\rho_{M_s}^{(old)}$ ，分散 τ_ρ^2 ，区間 $(-1,1)$ で定義された切断正規分布 (3.2) 式となる．STAYにおけるD-vine t コピュラパラメータ ρ_{M_s} のサンプリング方法は，2章のStep3と同様である．

$$(3.2) \quad \rho_{M_s}^{(new)} \sim TN_{(-1,1)}(\rho_{M_s}^{(old)}, \tau_\rho^2)$$

ν_{M_s} の提案分布は平均 $\nu_{M_s}^{(old)}$ ，分散 τ_ν^2 ，区間 $(1, \text{Upper})$ で定義された切断正規分布 (3.3) 式となる．STAYにおけるD-vine t コピュラパラメータ ν_{M_s} のサンプリング方法は，2章のStep4と同様である．

$$(3.3) \quad \nu_{M_s}^{(new)} \sim TN_{(1, \text{Upper})}(\nu_{M_s}^{(old)}, \tau_\nu^2)$$

なお，MH-RW法によるサンプリングアルゴリズムについては，2章Step3，Step4を参照のこと．

② BIRTH

D-vine t コピュラパラメータはMin and Czado[16]に従い，MH-RW法によってサンプリングする．

提案分布については，フルモデルのときD-vine t コピュラ M_s のデータ拡大法とMCMC法によるパラメータ推定値を $\hat{\rho}_{M_s}, \hat{\nu}_{M_s}$ とし，各々切断正規分布TNを仮定すると， ρ_{M_s} の提案分布は平均 $\hat{\rho}_{M_s}$ ，分散 τ_ρ^2 ，区間 $(-1,1)$ で定義された切断正規分布 (3.4) 式となる．

$$(3.4) \quad \rho_{M_s}^{(new)} \sim TN_{(-1,1)}(\hat{\rho}_{M_s}, \tau_\rho^2)$$

v_{M_s} の提案分布は平均 \hat{v}_{M_s} , 分散 τ_v^2 , 区間(1, Upper)で定義された切断正規分布(3.5)式となる. *³⁸ BIRTHにおけるD-vine tコピュラパラメータ ρ_{M_s} のサンプリング方法は, 切断正規分布の平均パラメータの設定が異なるが, 2章のStep3と同様である.

$$(3.5) \quad v_{M_s}^{(new)} \sim TN_{(1, \text{Upper})}(\hat{v}_{M_s}, \tau_v^2)$$

BIRTHにおけるD-vine tコピュラパラメータ v_{M_s} のサンプリング方法は, 切断正規分布の平均パラメータの設定が異なるが, 2章のStep4と同様である. なお, MH-RW法によるサンプリングアルゴリズムについては, 2章のStep3, Step4を参照のこと.

(5) STAYにおける ρ_{M_s}, v_{M_s} の採択確率, BIRTHおよびDEATHの採択確率

①STAY

MH-RW法によってパラメータ($\rho_{M_s}^{(new)}, v_{M_s}^{(new)}$)がサンプリングされる. ρ_{M_s}, v_{M_s} の採択確率は, 2章のStep3. (2.17)式, Step4. (2.20)式を参照のこと.

②BIRTH

1種類のモデルベクトル構成要素が無作為に追加される. BIRTHの採択確率は Appendix 3. (3A.3)式を参照のこと.

③DEATH

1種類のモデルベクトル構成要素が無作為に削除される. DEATHの採択確率は Appendix 3. (3A.6)式を参照のこと.

(6) モデル選択

RJMCMC法により, 独立コピュラを検出し, 独立コピュラでないペアコピュラの組合せを識別し, 試行を繰り返すことで, モデル候補が選出され, その不確実性である事後確率が算出される. 本研究では, Min and Czado[17]に従い, 事後確率が最も高い, すなわち, 最も選出回数が多いモデルを最適なモデルと考える.

*³⁸ Min and Czado[17]においては連続型データを仮定しており, 各提案分布の切断正規分布の平均値を最尤法により推定している. 一方, 本研究では離散型データを仮定しており, 2.1節で述べたように尤度関数の項数や最尤法による最適化の困難さを考慮して Smith and Khaled[20]が提案したデータ拡大法とMCMC法のサンプリング手法を採用した.

(7) D-vine tコピュラのパラメータ推定

Min and Czado[17]によれば、D-vine tコピュラのパラメータ推定は、RJMCMC法の試行から、事後確率が最も高い最適なモデルを選択後に、そのモデルが選択された試行のパラメータの標本を抽出する。本研究では、その標本の平均値を推定値と考える。

3.4. 数値実験

まず、設定したD-vine tコピュラの組合せ、およびパラメータから Aas, et al. [2]のアルゴリズムにより、周辺分布の依存構造を内包した一様乱数(0,1)を生成する。次に、データ拡大法と RJMCMC 法によって、最適なモデルを選択する。さらに、最適なモデルが選択された試行のパラメータの標本を抽出し、その平均値をパラメータの推定値とする。最後に、これらの再現結果から提案した手法の精度を検証する。

RJMCMC法では、単一連鎖法を選択し、試行回数は1万回で、最初の1千回は捨て、9千回を標本とする。

3.4.1.5 次元 D-vine t コピュラとモデルベクトル

(1) 5次元 D-vine t コピュラ

2章と同一のモデルを設定しており、(2.21)式を参照のこと。

(2) モデルベクトル

5次元のD-vine tコピュラの場合、その組合せによるモデル候補数は $2^{10}-1=1023$ 通り存在する。そのモデル選択については、初期モデルをフルモデル($p_M=n_c=10$)として、RJMCMC法によって、異なる複数のモデルに対応する異なった次元のパラメータ部分空間をサンプラーがジャンプすることで、その同時事後確率密度から、適当なモデルを選択する。

5次元のD-vine コピュラは、(2.21)式のとおり分解され、モデルベクトル M の構成要素の順序もこれに従うことで、モデルの識別が可能となる。

①モデルベクトル M の表示

3.3.2項(1)の定義より、例えば、フルモデルのとき $M=(1,1,1,1,1,1,1,1,1)$ 、モデルベクトル M の構成要素で第1番目のD-vine tコピュラ $c_{12}=1$ のとき、 $M=(0,1,1,1,1,1,1,1,1)$ と記述される。

②モデル番号

モデルベクトル M の第 s 列を M_s と記述して、モデル番号を次式で定義する。

$$\text{モデル番号} = \sum_{i=1}^{n_c} 2^{(n_c-i)} M_i \quad \text{ただし } n_c = 10$$

例えば、フルモデルのとき、モデル番号 = 1023となる。

3.4.2. 数値実験事例とその結果

(1) シミュレーションデータの生成

数値実験の対象データ数は 1000 個，5 計量単位とする。

周辺分布の依存関係を反映した一様乱数(0,1)の生成について，各 D-vine t コピュラのパラメータ $\rho_{j,k}, \nu_{j,k}$ は 3.4.2 項 (2) のとおり設定した。周辺分布のシミュレーションデータを作成方法は，2 章と同様に，Aas, et al. [2] のアルゴリズムを採用した。なお，Aas, et al. [2] のアルゴリズムについては，2.4.1 項 (2.22) 式を参照のこと。各周辺分布の u_{ij} の散布図は図 3-1 である。

(2) D-vine t コピュラの構成

3 種類の依存構造，そのパラメータ設定値から生成したシミュレーションデータから，モデル構造，パラメータが再現することを確認する。

モデル構造については，「依存構造が無条件の D-vine t コピュラで構成される場合」「依存構造が無条件と条件付 D-vine t コピュラで構成される場合」「依存構造が条件付 D-vine t コピュラで構成される場合」を考える。

また，D-vine t コピュラのパラメータ $\rho_{i,j}, \nu$ については，2 章と同じ値を設定した。詳細は，2.4.1 項を参照のこと。

① 依存構造が無条件の D-vine t コピュラで構成される場合

$$c(u_1, u_2, u_3, u_4, u_5) = c_{12}(u_1, u_2)c_{23}(u_2, u_3)c_{34}(u_3, u_4)c_{45}(u_4, u_5) \equiv c_{12}c_{23}c_{34}c_{45}$$

$$M = (1,1,1,1,0,0,0,0,0,0), \text{ モデル番号}=960$$

$$\rho_{12} = 0.3, \rho_{23} = 0.4, \rho_{34} = 0.5, \rho_{45} = 0.6, \nu_{12} = \nu_{23} = \nu_{34} = \nu_{45} = 7$$

② 依存構造が無条件と条件付 D-vine t コピュラで構成される場合

$$c(u_1, u_2, u_3, u_4, u_5) = c_{12}(u_1, u_2)c_{23}(u_2, u_3)c_{45}(u_4, u_5)$$

$$c_{35|4}(u_{3|4}, u_{5|4})c_{25|34}(u_{2|3,4}, u_{5|3,4}) \equiv c_{12}c_{23}c_{45}c_{35|4}c_{25|34}$$

$$M = (1,1,0,1,0,0,1,0,1,0), \text{ モデル番号}=842$$

$$\rho_{12} = 0.3, \rho_{23} = 0.4, \rho_{45} = 0.6, \rho_{35|4} = -0.2, \rho_{25|34} = -0.4, \nu_{12} = \nu_{23} = \nu_{45} = \nu_{35|4} = \nu_{25|34} = 7$$

③ 依存構造が条件付 D-vine t コピュラで構成される場合

$$c(u_1, u_2, u_3, u_4, u_5) = c_{13|2}(u_{1|2}, u_{3|2})c_{24|3}(u_{2|3}, u_{4|3})c_{35|4}(u_{3|4}, u_{5|4})$$

$$c_{14|23}(u_{1|2,3}, u_{4|2,3})c_{25|34}(u_{2|3,4}, u_{5|3,4})c_{15|234}(u_{1|2,3,4}, u_{5|2,3,4}) \equiv c_{13|2}c_{24|3}c_{35|4}c_{14|23}c_{25|34}c_{15|234}$$

$$M = (0,0,0,0,1,1,1,1,1,1), \text{ モデル番号}=63$$

$$\rho_{13|2} = 0.7, \rho_{24|3} = 0.8, \rho_{35|4} = -0.2, \rho_{14|23} = -0.3, \rho_{25|34} = -0.4, \rho_{15|234} = -0.5$$

$$\nu_{13|2} = \nu_{24|3} = \nu_{35|4} = \nu_{14|23} = \nu_{25|34} = \nu_{15|234} = 7$$

数値実験における模擬データの設定について以下に述べる。データ数ならびに計量単位数の設定理由については，2 章と同様である。10 種類の依存構造が設定されており，図 3-1 の散布図から各計量単位の様々な依存構造が確認できる。ま

た，依存構造の無条件と条件付 D-vine t コピュラの組合せについて 3 種類を設定している．以上から，精度検証における模擬データは，無条件と条件付 D-vine t コピュラの構成，依存構造など多様な事例が考慮されていると考える．

(3) フルモデルにおけるパラメータの推定

BIRTH において新たにサンプリングされる D-vine t コピュラのパラメータ $\rho_{M_s}^{(new)}, \nu_{M_s}^{(new)}$ は 3.3.2 項 (4) ② で述べたように，フルモデルのときのパラメータ推定値 $\hat{\rho}_{M_s}, \hat{\nu}_{M_s}$ を提案分布における切断正規分布の平均値に設定している．その推定手法は，Smith and Khaled[20]が提案したデータ拡大法を拡張した MCMC 法のサンプリング手法を採用し，初期値には，周辺分布 j, k の順位相関(ケンドールの τ)を設定した．MCMC 法について単一連鎖法を選択，試行回数は 2 万回で最初の 5 千回を捨て，その後の 1 万 5 千回を標本とする．CD については，Geweke[8]が推奨するように前半と後半の標本数を $T_1=1500, T_2=7500$ とする．なお，MH-RW 法によるサンプリングアルゴリズムについては，2 章と同様であり，2.3 節 Step3, Step4 を参照のこと．

① データ拡大法による潜在変数

潜在変数の MH サンプリングアルゴリズムでは，2 章と同様で，1 回の試行につきデータ番号 i と同数の採択棄却が実行される．*39 全データ番号の全試行について，採択確率の最小は 86.5%，最大は 99.04%と安定しており，表 3-1 のとおり平均は 97.28%と高く提案分布 (2.12) 式の近似が良好であることを示している．

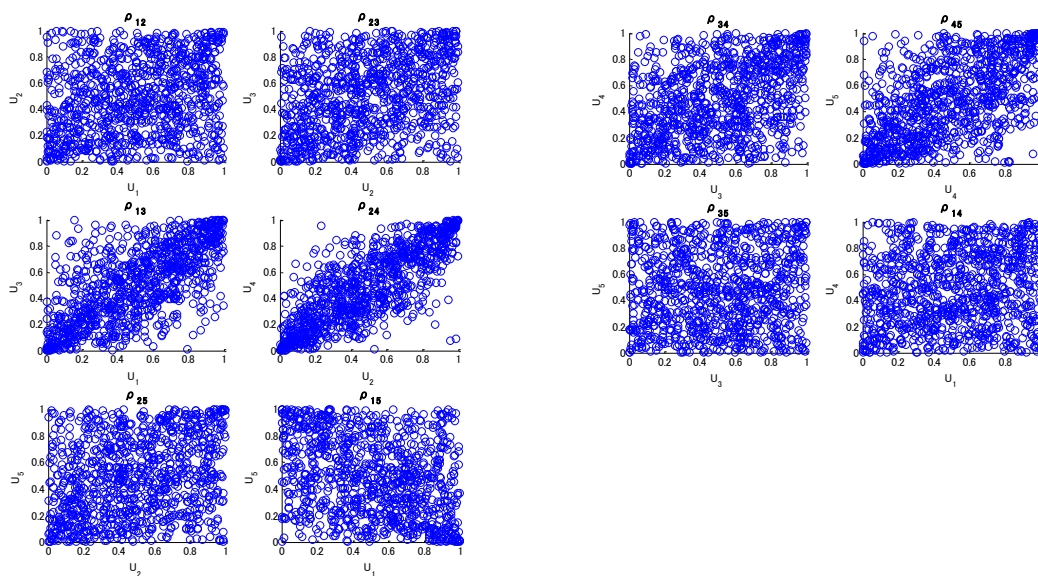


図 3-1. Scatter plot of $u_{i,j}$.

*39 潜在変数のサンプリングアルゴリズムについては，2.3 節 Step2 を参照のこと．

表 3-1. Acceptance rate of latent variable $U_{i,j}$.

	Maximum	Minimum	Average
Acceptance Ratio (%)	99.04	86.50	97.28
Data Number	256	2	

②BIRTHにおける提案分布のパラメータ $\hat{\rho}_{M_s}, \hat{\nu}_{M_s}$ の推定

ハイパーパラメータは表 3-2 に、推定結果は表 3-3 に記載、標本パス、事後ヒストグラム、標本自己相関は図 3-2 に図示した。

表 3-3 から、相関 $\hat{\rho}_{M_s}$ 自由度 $\hat{\nu}_{M_s}$ について、設定値（真値）は、95%信用区間の範囲内であり、設定値と推定値は概ね同水準となり、パラメータ推定における精度は問題ないと言える。CD については有意水準 5% の確率で全てのパラメータが受容されている。また、NIF については 2~4 の範囲である。図 3-2 の標本パスは状態空間を十分にサンプリングして安定的に推移しているように観測され、CD の結果からも、収束していると考えられる。図 3-2 から、標本自己相関については、 $\hat{\rho}_{12}, \hat{\rho}_{23}, \hat{\rho}_{34}$ を除き急速な減衰が観測され、NIF の結果からも効率的にサンプリングされていると考えられる。また、標本事後確率密度の形状についても、単峰形で正規分布に近い形状であり、特に問題は観察されない。

表 3-2. Hyperparameter.

	$\tau_{\rho_{12}}$	$\tau_{\rho_{23}}$	$\tau_{\rho_{34}}$	$\tau_{\rho_{45}}$	$\tau_{\rho_{13 2}}$	$\tau_{\rho_{24 3}}$	$\tau_{\rho_{35 4}}$	$\tau_{\rho_{14 23}}$	$\tau_{\rho_{25 34}}$	$\tau_{\rho_{15 234}}$	τ_{ν}
Standard Deviation	0.1	0.1	0.2	0.2	0.3	0.3	0.1	0.2	0.3	0.3	1

表 3-3. Estimation results of D-vine t copula parameters.

	Set Value	Initial Value	Posterior Mean	Standard Error	2.5%	97.5%	CD	NIF	Acceptance Ratio (%)	Divergence Ratio (%)
$\hat{\rho}_{12}$	0.3	0.199	0.276	0.001	0.218	0.335	0.382	3.900	23.5	8.0
$\hat{\rho}_{23}$	0.4	0.287	0.372	0.001	0.320	0.425	0.864	4.288	16.8	7.0
$\hat{\rho}_{34}$	0.5	0.36	0.456	0.0008	0.403	0.504	0.574	4.124	18.2	8.8
$\hat{\rho}_{45}$	0.6	0.421	0.585	0.0004	0.547	0.6186	1.028	2.942	20.0	2.5
$\hat{\rho}_{13 2}$	0.7	0.554	0.703	0.0004	0.671	0.733	-0.184	3.265	15.9	0.4
$\hat{\rho}_{24 3}$	0.8	0.624	0.803	0.0003	0.782	0.8224	-0.984	3.315	11.7	0.4
$\hat{\rho}_{35 4}$	-0.2	0.115	-0.238	0.0007	-0.295	-0.182	0.351	2.760	29.3	19.0
$\hat{\rho}_{14 23}$	-0.3	0.169	-0.305	0.0006	-0.364	-0.247	-0.695	2.571	31.5	1.7
$\hat{\rho}_{25 34}$	-0.4	0.212	-0.417	0.0006	-0.465	-0.365	0.404	2.796	27.6	4.3
$\hat{\rho}_{15 234}$	-0.5	-0.143	-0.482	0.0005	-0.533	-0.432	1.158	2.410	28.2	3.6
ν	7	10.0	6.289	0.0096	5.372	7.445	1.257	2.220	49.2	10.2

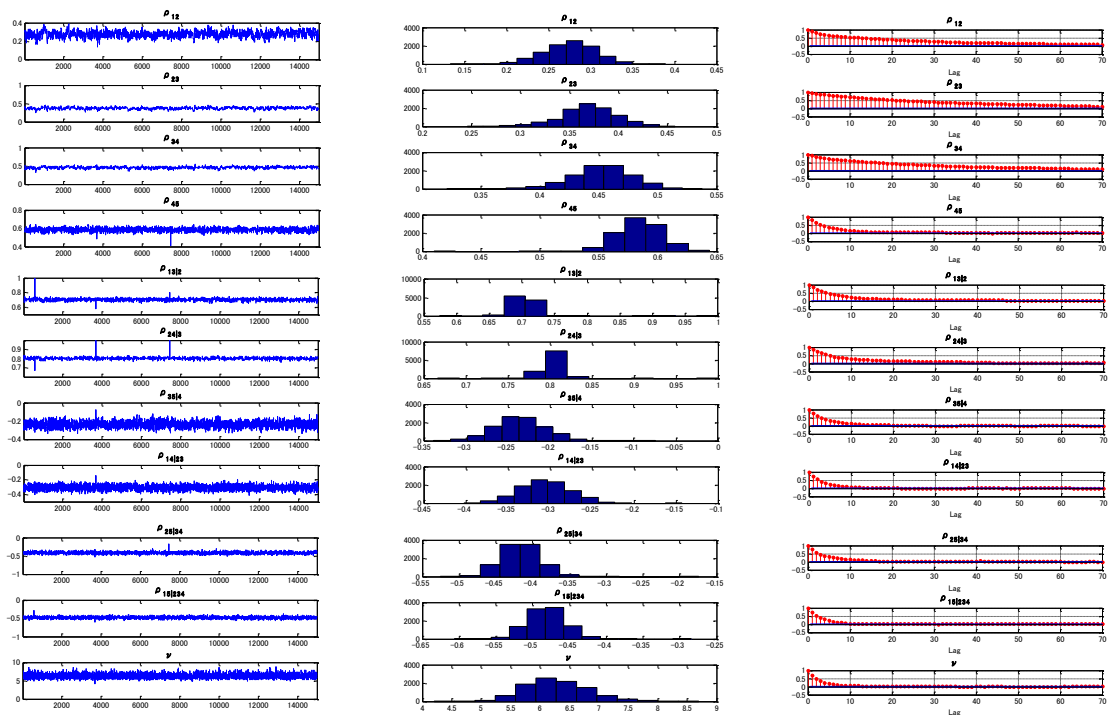


図 3-2. Sample path, posterior probability density, and sample autocorrelation for D-vine t copula parameters.

(4) データ拡大法と RJMCMC 法によるモデル選択

データ拡大法における潜在変数 U の採択確率については、表 3-4, 表 3-6 表 3-8 のとおり、モデル番号 960, 842, 63 の採択確率の平均が 98.69%, 98.78%, 98.48% と高水準で、(2.12) 式の近似が良好であることを示している。

モデル選択における事後確率は、表 3-5, 表 3-7, 表 3-9 のとおり、モデル番号 960 の無条件コピュラのみ構成の場合は 95.3%, モデル番号 842 の無条件と条件付コピュラの構成の場合は 100.0%, モデル番号 63 の条件付コピュラのみ構成の場合が 100.0% といずれも良好であった。^{*40} なお、表 3-5 において、事後確率が 1% 未満のモデルについて省略した。

図 3-3 のモデル遷移のパスについては、最適なモデルが選択される過程を確認するため、第 1 回から 1 万回の試行を掲載した。いずれも安定的に推移していることが観察される。また、Min and Czado[17] によれば「Acceptance rate of birth-death moves」は独立コピュラが多く含まれるモデルで高くなる傾向があり、「約 7~9% 水準、その他の場合で約 1~6% 水準」と報告されているが、本研究では、独立コピュラ数が 6 項のモデル番号 960 では 0.57%, 同様に 5 項のモデル番号 842, 4 項のモデル番号 63 では 0% となり、独立コピュラが多く含まれるモデルで高くなる傾向があるものの、Min and Czado[17] と比較しその水準は低位となった。

⁴⁰ 図 3-3 から確認できるように、標本として採用していない最初の 5000 回までの試行で設定したモデルに遷移して、それ以降の試行では他のモデルに遷移しなかった。

①モデル番号 960, $c_{12}c_{23}c_{34}c_{45}$, $M = (1,1,1,1,0,0,0,0,0)$

(i)データ拡大法における採択確率

表 3-4. Acceptance rate of latent variable $U_{i,j}$.

Model Number 960	Maximum	Minimum	Average
Acceptance Ratio(%)	99.89	87.08	98.69
Data Number	262	877	

(ii)モデル選択の事後確率

表 3-5. Posterior model probability.

Model Number	Model	Model Vector	Posterior Probability(%)
960	$c_{12}c_{23}c_{34}c_{45}$	(1,1,1,1,0,0,0,0,0)	95.30
964	$c_{12}c_{23}c_{34}c_{45}c_{14 23}$	(1,1,1,1,0,0,0,1,0)	2.89
962	$c_{12}c_{23}c_{34}c_{45}c_{25 34}$	(1,1,1,1,0,0,0,0,1)	1.18
Acceptance rate of birth-death moves			0.57

②モデル番号 842, $c_{12}c_{23}c_{45}c_{35|4}c_{25|34}$, $M = (1,1,0,1,0,0,1,0,1)$

(i)データ拡大法における採択確率

表 3-6. Acceptance rate of latent variable $U_{i,j}$.

Model Number 842	Maximum	Minimum	Average
Acceptance Ratio(%)	99.94	77.57	98.78
Data Number	740	2	

(ii)モデル選択の事後確率

表 3-7. Posterior model probability.

Model Number	Model	Model Vector	Posterior Probability(%)
842	$c_{12}c_{23}c_{45}c_{35 4}c_{25 34}$	(1,1,0,1,0,0,1,0,1)	100.00
Acceptance rate of birth-death moves			0.00

③モデル番号 63, $c_{13|2}c_{24|3}c_{35|4}c_{14|23}c_{25|34}c_{15|234}$, $M = (0,0,0,0,1,1,1,1,1)$

(i)データ拡大法における採択確率

表 3-8. Acceptance rate of latent variable $U_{i,j}$.

Model Number 63	Maximum	Minimum	Average
Acceptance Ratio(%)	99.90	78.69	98.48
Data Number	555	932	

(ii)モデル選択の事後確率

表 3-9. Posterior model probability.

Model Number	Model	Model Vector	Posterior Probability(%)
63	$c_{13 2}c_{24 3}c_{35 4}c_{14 23}c_{25 34}c_{15 234}$	(0,0,0,0,1,1,1,1,1,1)	100.00
Acceptance rate of birth-death moves			0.00

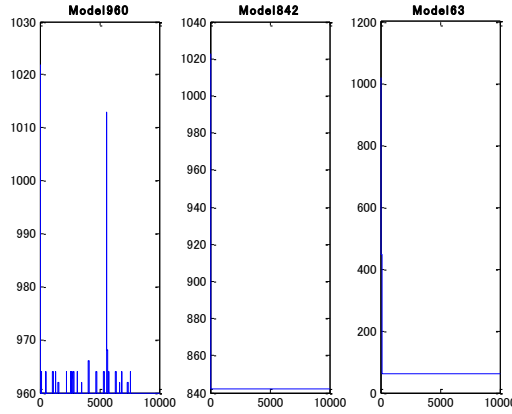


図 3-3. Model switching path.

(5) RJMCMC 法による最適なモデル選択後のパラメータ推定

最適なモデル選択後、最適なモデルが選択された試行から抽出されたパラメータの標本の平均を「事後平均」とする。また、参考までに、信用区間を記載する。

相関 ρ , 自由度 ν について、設定値（真値）は、表 3-10, 表 3-11, 表 3-12 のとおり、95%信用区間の範囲内である。モデル番号 960, 842, 63 の各々のパラメータの設定値と推定値の乖離率の平均は、4.6%, 7.6%, 6.2%と良好な推定精度である。事後確率密度の形状については、図 3-4, 図 3-5, 図 3-6 のとおり、単峰形で正規分布に近い形状であり、特に問題は観察されない。

①モデル番号 960 $c_{12}c_{23}c_{34}c_{45}$, $M = (1,1,1,1,0,0,0,0,0,0)$

表 3-10. No. 960 Estimation results of D-vine t copula parameters.

最適モデルが選択され、パラメータ推定に使用された標本数は 8,577 である。

	Set Value	Posterior Mean	Standard Error	2.5%	97.5%	Divergence Ratio (%)
ρ_{12}	0.3	0.333	0.0014	0.270	0.394	11.0
ρ_{23}	0.4	0.404	0.0012	0.350	0.453	1.0
ρ_{34}	0.5	0.494	0.0011	0.446	0.544	1.2
ρ_{45}	0.6	0.630	0.0009	0.599	0.686	5.0
ν	7.0	7.338	0.045	5.708	9.586	4.8
Average of Divergence Ratio(%)						4.6

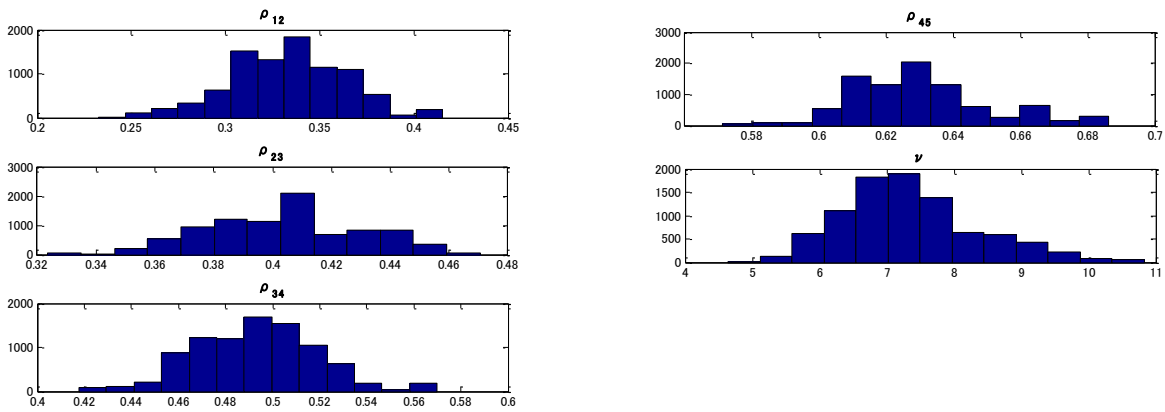


図 3-4. No. 960 Posterior probability density.

②モデル番号 842 $c_{12}c_{23}c_{45}c_{35|4}c_{25|34}$, $M = (1,1,0,1,0,0,1,0,1,0)$

表 3-11. No. 842 Estimation results of D-vine t copula parameters.

最適モデルが選択され，パラメータ推定に使用された標本数は 9000 である．

	Set Value	Posterior Mean	Standard Error	2.5%	97.5%	Divergence Ratio (%)
ρ_{12}	0.3	0.278	0.0013	0.233	0.331	7.3
ρ_{23}	0.4	0.390	0.0013	0.333	0.436	2.5
ρ_{45}	0.6	0.601	0.0008	0.564	0.629	0.2
$\rho_{35 4}$	-0.2	-0.139	0.0017	-0.209	-0.058	30.5
$\rho_{25 34}$	-0.4	-0.386	0.0012	-0.441	-0.325	3.5
ν	7.0	6.901	0.0432	5.361	8.977	1.4
Average of Divergence Ratio(%)						7.6

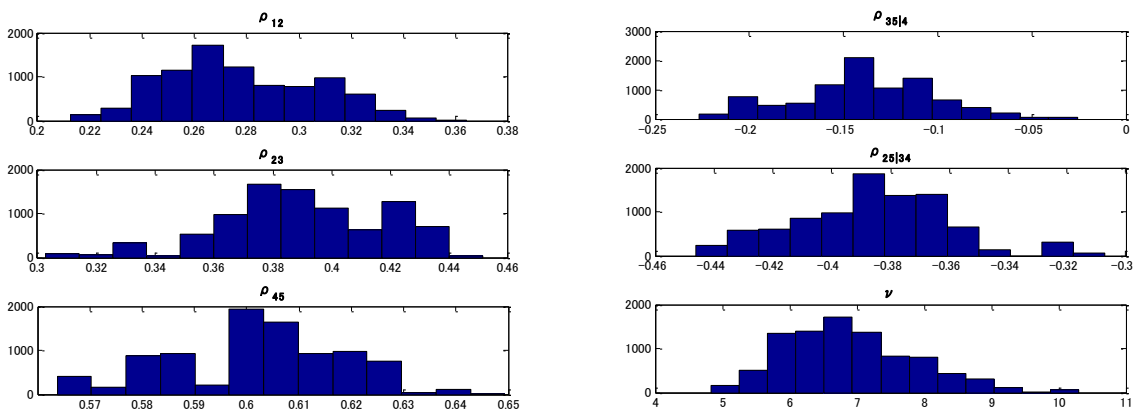


図 3-5. No. 842 Posterior probability density.

③モデル番号 63 $c_{13|2}c_{24|3}c_{35|4}c_{14|23}c_{25|34}c_{15|234}$, $M = (0,0,0,0,1,1,1,1,1)$

表 3-12. No. 63 Estimation results of D-vine t copula parameters.

最適モデルが選択され，パラメータ推定に使用された標本数は 9000 である．

	Set Value	Posterior Mean	Standard Error	2.5%	97.5%	Divergence Ratio (%)
$\rho_{13 2}$	0.7	0.712	0.0005	0.689	0.737	1.7
$\rho_{24 3}$	0.8	0.791	0.0003	0.776	0.803	1.1
$\rho_{35 4}$	-0.2	-0.197	0.0012	-	-0.143	1.5
$\rho_{14 23}$	-0.3	-0.296	0.0015	-	-0.244	1.3
$\rho_{25 34}$	-0.4	-0.416	0.0012	-	-0.367	4.0
$\rho_{15 234}$	-0.5	-0.558	0.0008	-	-0.519	11.6
ν	7.0	5.454	0.0297	4.374	7.038	22.1
Average of Divergence Ratio(%)						6.2

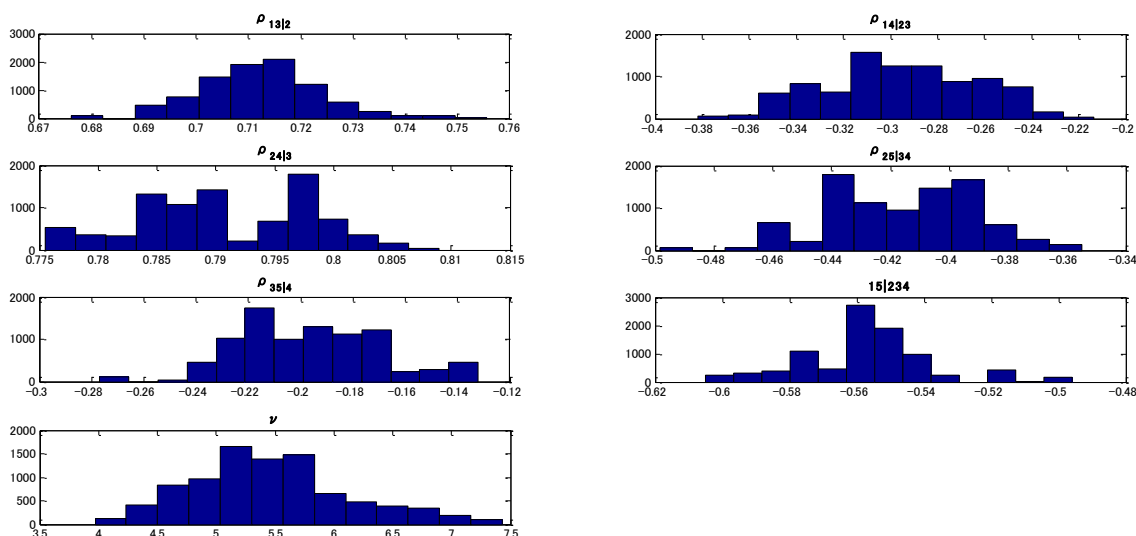


図 3-6. No. 63 Posterior probability density.

3.5. 3章のまとめと今後の課題

これまで，RJCMC 法による独立コピュラを検出する手法は，連続型データの場合は存在したが，離散型データの場合には存在しなかった．そこで，Min and Czado[17]による RJCMC 法を応用した独立コピュラを検出するモデル選択手法とデータ拡大法を併用することで，離散型データの場合に拡張することを提案した．その手法の精度を検証するための数値実験において，独立コピュラの探索によるモデル選択の事後確率は，モデル番号 960 については 95.3%，842 と 63 については 100%であり，モデル識別精度および，パラメータ推定精度は高かった．これにより，本研究で提案した手法により離散型データにおいても，独立コピュラの検出によりペアコピュラの項数が減少することで，計算負荷の軽減が期待できる．

実務への適用例であるが，金融分野で，周辺分布をパラメトリック，ノン

パラメトリックの離散分布でモデル化，依存構造をVineコピュラでモデル化する場合，信用，市場，オペレーションナルリスク分野における用途は幅広く，当該手法は有益であると考える．

今後の課題は以下の2点である．まず，条件付コピュラの評価を無条件コピュラで近似しているが，推定精度向上のため数値計算手法の開発が望まれる．次に実用性に対する説得性や完成度を高めるために金融データを使った実証分析が望まれる．

Appendix 3. RJMCMC法による遷移アルゴリズム

Step3A.1 遷移確率と遷移先(STAY, BIRTH, DEATHとペアコピュラの組合せの選択)を決定する．

$p_M \equiv \sum_{s=1}^{n_c} I_{\{M_s \neq 0\}}$ とすると，モデル M から新たなモデル $M_s^{(new)}$ への遷移確率 $g(M_s \rightarrow M_s^{(new)})$ は次式のとおりである．

- $p_M = n_c$ のとき，STAY, DEATHの遷移確率は， $\frac{1}{n_c+1}, \frac{n_c}{n_c+1}$
- $1 < p_M < n_c$ のとき，STAY, BIRTH, DEATHの遷移確率は， $\frac{1}{n_c+1}, \frac{n_c - p_M}{n_c+1}, \frac{p_M}{n_c+1}$
- $p_M = 1$ のとき，STAY, BIRTHの遷移確率は， $\frac{1}{n_c}, \frac{n_c - 1}{n_c}$

Step3A.2 次に，モデルが無作為(1組のペアコピュラの追加または削除)に選択される．

(i) STAY の場合

提案分布，採択確率について，2章と同様であり，2.3節 Step3, Step4を参照のこと．

(ii) BIRTH の場合

Step3A.2-1 提案分布

フルモデルのときD-vine tコピュラ M_s のデータ拡大法とMCMC法によるパラメータ推定値を $\hat{\rho}_{M_s}, \hat{v}_{M_s}$ とし，各々切断正規分布を仮定すると， $\rho_{M_s}^{(new)}$ の提案分布は平均 $\hat{\rho}_{M_s}$ ，分散 τ_ρ^2 ，区間 $(-1,1)$ で定義された切断正規分布(3A.1)式である．

$$(3A.1) \quad \rho_{M_s}^{(new)} \sim TN_{(-1,1)}(\hat{\rho}_{M_s}, \tau_\rho^2)$$

$v_{M_s}^{(new)}$ の提案分布は平均 \hat{v}_{M_s} ，分散 τ_v^2 ，区間 $(1, \text{Upper})$ で定義された切断正規分布(3A.2)式である．

$$(3A.2) \quad v_{M_s}^{(new)} \sim TN_{(1, \text{Upper})}(\hat{v}_{M_s}, \tau_v^2)$$

Step3A.2-2 採択確率

Min and Czado[17]より，BIRTHでは，モデルベクトル構成要素が追加される．コ

ピュラパラメータの採択確率は(3A.3)式である.

$$(3A.3) \quad \alpha_B = \min \left\{ 1, \frac{c(u | \theta_{M^{(new)}}) \pi(\theta_{M^{(new)}}) g(M^{(new)} \rightarrow M)}{c(u | \theta_M) g(M \rightarrow M^{(new)}) \varphi_2(\theta_{M_s}^{(new)})} * J \right\}$$

ここで, $c(u | \theta_M)$ は現在のモデルベクトル M , $c(u | \theta_{M^{(new)}})$ は遷移後のモデルベクトル $M^{(new)}$ のD-vine tコピュラとすると, 事前分布 $\pi(\theta_{M^{(new)}})$, $\theta_{M_s}^{(new)}$ の確率密度 $\varphi_2(\theta_{M_s}^{(new)})$ は(3A.4)式, (3A.5)式となる.

ただし, Min and Czado[16]に従い, 相関 ρ_{M_s} の事前分布は一様分布 $(-1,1)$ の, 自由度 ν_{M_s} の事前分布は一様分布 $(1, \text{Upper})$ に従い, 独立であるとする.

$$(3A.4) \quad \pi(\theta_{M_s}) = \prod_{M_s \in M, M_s \neq 0} \left(\frac{1}{2} I_{(-1,1)}(\rho_{M_s}) \frac{1}{\text{Upper} - 1} I_{(1, \text{Upper})}(\nu_{M_s}) \right)$$

$$(3A.5) \quad \varphi_2(\theta_{\theta_s}^{(new)}) = tn_{(-1,1)}(\theta_{\theta_s}^{(new)} | \hat{\rho}_{M_s}, \tau_\rho^2) * tn_{(1, \text{Upper})}(\theta_{M_s}^{(new)} | \hat{\nu}_{M_s}, \tau_\nu^2)$$

ここで, $tn_{(-1,1)}(\hat{\rho}_{M_s}, \tau_\rho^2)$, $tn_{(1, \text{Upper})}(\hat{\nu}_{M_s}, \tau_\nu^2)$ は切断分布 TN の密度関数であり, ヤコビアン $J = 1$ である.

(iii) DEATH の場合

Step3A.2-1 採択確率

Min and Czado[17]より, モデルベクトル構成要素の削除の採択確率は(3A.6)式となる.

$$(3A.6) \quad \alpha_D = \min \left\{ 1, \frac{c(u | \theta_M) g(M \rightarrow M^{(new)}) \varphi_2(\theta_{M_s}^{(new)})}{c(u | \theta_{M^{(new)}}) \pi(\theta_{M^{(new)}}) g(M^{(new)} \rightarrow M)} * J \right\}$$

ここで, $\theta_{M_s}^{(new)}$ は削除される t ペアコピュラのパラメータ, $\varphi_2(\theta_{M_s}^{(new)})$ は BIRTH を参照のこと.

Step3A.1 から Step3A.2 を試行回数繰り返す.

3章参考文献

- [1] Aas, K. and Berg, D., Models for construction of multivariate dependence: A comparison study, *The European Journal of Finance*, **15**(2009), 639-659.
- [2] Aas, K., Czado, C., Frigessi, A. and Bakken, H., Pair-copula Constructions of Multiple Dependence, *Insurance: Mathematics and Economics*, **44** (2009), 182-198.
- [3] Bedford, T. and Cooke, R. M., Vines - A new graphical model for dependent random variables, *The Annals of Statistics*, **30** (2002), 1031-1068.
- [4] Berg, D., Copula goodness-of-fit testing: An overview and power comparison, *The European Journal of Finance*, **15** (2009), 675-701.
- [5] Besag, J., Green, P., Higdon, D. and Mengersen, K., *Bayesian Computation and*

- Stochastic Systems, *Statistics Science*, **10** (1995), 3-66.
- [6] Fischer, M., Köck, C., Schlüter, S. and Weigert, F. , An empirical analysis of multivariate copula models, *Quantitative Finance*, **9** (2009), 839-854.
- [7] Genest, C. and Nešlehová, J., A primer on copulas for count data, *Astin Bulletin*, **37** (2007),475-515.
- [8] Geweke, J. ,Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, In: Bernardo, J. M. Berger , J. O. Dawid, A. P. and Smith , A. F. M. (eds.), *Bayesian Statistics 4*, Oxford University Press, 1992, 169-193.
- [9] Green, P. J., Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, **82** (1995), 711-732.
- [10] 岩永育子, ヴァイン・コピュラを用いた CAPM の非正規・非線形への拡張 : 日本株式市場における実証分析, In:日本金融・証券計量・工学学会(編), *ファイナンスとデータ解析*, ジャファイ・ジャーナル: 金融工学と市場計量分析, 2015, 69-113.
- [11] Joe, H., *Multivariate Models and Dependence Concepts*, Chapman & Hall, UK, 1997.
- [12] 粕谷宗久, 高川泉, 実質為替レートの中期予測におけるモデルの不確実性, In: 和合肇 (編) , *ベイズ計量経済分析—マルコフ連鎖モンテカルロ法とその応用*, 東洋経済新報社,東京, 2005, 213-233.
- [13] 川口宗紀,山中卓,田代雄介, コピュラの実務への適用例-与信集中リスク評価への応用, *証券アナリストジャーナル*, **52** (2014), 43-51.
- [14] Kurowicka, D. and Joe, H. (eds.), *Dependence Modeling: Vine Copula Handbook*, World Scientific, Singapore, 2011.
- [15] Mashal, R., and Naldi, M., *Pricing Multiname Credit Derivatives: Heavy Tailed Hybrid Approach*,Working Paper, Columbia Business School, 2002.
<http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=F3B3F699A8AD1D2BD841CAD5EE30D033?doi=10.1.1.139.4460&rep=rep1&type=pdf>
- [16] Min, A. and Czado, C., Bayesian Inference for multivariate copulas using pair-copulas constructions, *Journal of Financial Econometrics*, **8** (2010), 511-546.
- [17] Min, A. and Czado, C., Bayesian model selection for D-vine pair-copula constructions, *The Canadian Journal of Statistics*, **39** (2011), 239-258.
- [18] Nelsen, R. B., *An Introduction to Copulas*, Second Edition, Springer, 2006.
- [19]日本銀行, *オペレーショナル・リスク管理の高度化*, 2005.
https://www.boj.or.jp/research/brp/ron_2005/fsk0507a2.pdf
- [20] Smith, M. and Khaled M.A, Estimation of Copula Models with Discrete Margins via Bayesian Data Augmentation, *Journal of the American Statistical Association*,

107 (2012), 290-303.

- [21] Smith, M., Min, A., Almeida, C., and Czado, C. , Modeling Longitudinal Data Using a Pair-Copula Decomposition of Serial Dependence, *Journal of the American Statistical Association*, **105** (2010), 1467–1479.
- [22] 塚原英敦, 接合分布関数 (コピュラ) の理論と応用, In; 国友直人, 山本拓 (監修), 北川源四郎, 竹村彰通(編), 21 世紀の統計科学 III, 2011, 111-139.
- [23] 戸坂凡展, 吉羽要直, コピュラの金融実務での具体的な活用方法の解説, *金融研究*, **24**(2005), 115-162.

4. 結論

4.1. 本研究のまとめ

本研究では、オペリスク損失金額データとオペリスクのリスク量の特徴を踏まえ、周辺分布の離散分布による近似、D-vine t コピュラによる依存構造のモデル化ならびに MCMC によるパラメータの推定手法について提案した。また、Vine コピュラによって依存構造をモデル化する場合には、パラメータ推定の計算負荷が高いことから、RJCMC を利用した独立コピュラの探索による負荷軽減手法についても提案した。

まず、第1章では、オペレーショナルリスク分析の概要、リスク計量化の課題など本研究の背景、それに取組むための目的、本論文の構成を示した。第2章では、提案した周辺分布および依存構造のモデルおよびパラメータ推定手法について、その推定精度ならびに推定されたリスク量の安定性から実務への応用可能性が示された。パラメータの推定精度については、数値実験によって検証され、推定値は設定したパラメータが高い精度で再現された。従って、実務において、本研究で示した損失金額データの離散近似、D-vine t コピュラによる依存構造のモデル化および MCMC によるパラメータ推定手法を採用することで、内部損失、シナリオ損失の全ての損失金額データを対象とした離散分布、D-vine t コピュラのパラメータを同時推定できることが示された。推定されたリスク量の安定性については、損失金額データが損失金額分布のボディ部分に相当する内部損失金額データとテール部分に相当するシナリオ損失金額データを網羅していること、また、一例ではあるが、先行研究の手法とのリスク量を比較し、より安定的に高い精度でリスク量が推定されることから、提案した手法が有用であることが確認できた。第3章では、離散型データの場合に、RJCMC 法とデータ拡大法を併用することで、独立コピュラを検出するモデル選択手法を新たに提案した。その手法の精度を検証するための数値実験において、モデル識別精度および、パラメータ推定精度が高いことが確認できた。これにより、本研究で提案した手法により離散型データにおいて、独立コピュラの検出によりペアコピュラの項数が減少することで、計算負荷の軽減が期待できる。

以上の分析により、まず、オペリスクのリスク量の推定において、依存構造を D-vine t コピュラでモデル化し、RJCMC 法で独立コピュラを検出する。次に、内部損失金額データとシナリオ損失金額データの1件あたりの損失金額分布を離散分布で近似する。最後に、離散分布ならびに D-vine t コピュラのパラメータを MCMC 法で推定する。これにより、連続型データと離散型データが混在した内部損失金額データとシナリオ損失金額データによる1件当たりの損失分布の推定、リスク量 (VaR) の過小評価の問題を回避した、社内全体のオペリス

クのリスク量が効率的に推定できるものと思慮する。

4.2. 今後の研究課題

第1に、ビン、D-vine t コピュラの MCMC 法によるパラメータ推定結果の CD から、標本が条件付事後分布へ収束していない可能性がある事例、NIF から、サンプリングが効率的に行われていない可能性がある事例があった。これは、周辺分布のサンプリング手法によるものか、試行回数によるものか、データ数と推定パラメータ数によるものであるかなど、原因究明が必要である。第2に、ビンのパラメータ推定において、ビン幅、ビン数を既知としているが、同時推定についても検討が必要である。第3に、条件付きコピュラの評価を無条件コピュラで近似しているが、推定精度向上のため数値計算手法の開発が課題としてあげられる。第4に、数値実験におけるデータの種類について、高次元の計量単位数、極端な裾の厚い分布、D-vine t コピュラにおける極端な相関係数や自由度の設定、およびその他の Vine コピュラの適用などにより提案した手法の有効性のより多くの事例での検証が課題としてあげられる。第5に、高頻度低額損失金額分布は連続分布で、低頻度高額損失金額分布は離散分布の連続分布と離散分布の混合分布を想定しパラメータ推定手法を検討する。最後に、実用性に対する説得性や完成度を高めるために金融データを使った実証分析が望まれる。

謝辞

本論文の作成にあたり、研究全般についてご指導ならびにご助言いただきました中央大学理工学部藤田岳彦教授はじめ、東京理科大学経営学部の吉田直弘助教、慶應大学経済学部の中妻照雄教授の皆様方に感謝の意を表します。また、仕事と研究との両立を精神的に支えてくれた家族に心より感謝します。