

セミパラメトリックコピュラモデルにおける ダイバージェンスの性質

清 智也[†]・松本 和也[†]

(受付 2019 年 5 月 31 日；改訂 9 月 27 日；採択 10 月 4 日)

要 旨

多次元の量的データに対し、コピュラとしてパラメトリックモデルを仮定し、周辺分布には何も仮定をおかない統計モデルのことをセミパラメトリックコピュラモデルという。本論文ではこのモデルにおけるダイバージェンスの性質を考察する。特に、多変量順位統計量の周辺分布によって定まる順位ダイバージェンスと、Kullback–Leibler ダイバージェンスの局外パラメータに関する最小値として定義されるプロファイルダイバージェンスの関係を調べる。また区分一様コピュラとガウスコピュラの場合について具体的な計算結果を示す。

キーワード：コピュラ、最適輸送理論、情報幾何、ダイバージェンス、複合変換モデル、ホロノミック勾配法。

1. はじめに

d 次元の確率密度関数 $c(x)$ ($x = (x_1, \dots, x_d) \in [0, 1]^d$) がコピュラ密度であるとは、全ての 1 次元周辺密度が $[0, 1]$ 上の一様密度になることと定義する。Sklar の定理から、 \mathbb{R}^d 上の任意の確率密度関数 $p(x) = p(x_1, \dots, x_d)$ は

$$(1.1) \quad p(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{i=1}^d F_i'(x_i)$$

という形で一意的に表すことができる。ここで c はコピュラ密度、 F_i は x_i の周辺分布関数を表す。また F_i' は F_i の導関数である。コピュラ密度 c にパラメトリックモデルを仮定し、周辺分布 F_i には仮定をおかない統計モデルのことをセミパラメトリックコピュラモデル (semiparametric copula model) という。より正確な定義は式 (3.2) で与える。

セミパラメトリックコピュラモデルのパラメータ推定については様々な研究がある。Klaassen and Wellner (1997) は 2 変量ガウスコピュラの場合に擬似最尤推定量および正規スコア順位相関係数が漸近有効推定量を与えることを示した。Genest and Werker (2002) は、ガウスコピュラ以外では擬似最尤推定量が必ずしも漸近有効でないことを指摘した。Chen et al. (2006) は周辺分布の推定に sieve を用いることで漸近有効推定量を構成した。Tsukahara (2005) は順位統計量のみ依存した推定量のクラスを与え、その漸近的性質を明らかにした。順位統計量のみ依存した漸近有効推定量の構成は重要な未解決問題となっている。ガウスコピュラの部分モデルに限れば、Hoff et al. (2014) が漸近有効情報量の特徴付け、Segers et al. (2014) が実際に漸

[†] 東京大学大学院 情報理工学系研究科：〒113-8656 東京都文京区本郷 7-3-1

近有効推定量を構成した。その他、コピュラに関する総合的な解説は塚原 (2011) を参照されたい。

本論文ではセミパラメトリックコピュラモデルにおけるダイバージェンスの性質を調べる。Kullback–Leibler ダイバージェンスを規範として、順位ダイバージェンスとプロファイルダイバージェンスという 2 つの量を定義する。これらはそれぞれ順位尤度 (Hoff, 2007) とプロファイル尤度の母集団特性を表したものである。いずれも一般には陽に計算できないが、区分一様コピュラに限れば具体的に計算を進めることができる。その結果、正則条件のもとで順位ダイバージェンスがプロファイルダイバージェンスに収束することが示される (定理 2)。またガウスコピュラの場合、プロファイルダイバージェンスの計算は有限次元の最適化問題に帰着され、順位ダイバージェンスは多次元正規分布の象限確率で表される。後者はホロノミック勾配法を利用して計算できる。以上の結果は推定量の性能を知るための手がかりを与えるものと考えられる。

論文の構成は以下の通りである。まず 2 節では簡単な例を通してセミパラメトリックコピュラモデルにおけるダイバージェンスを論ずる必要性を確認する。3 節では実際にダイバージェンスを定義し、基本的な定理を与える。4 節と 5 節ではそれぞれ区分一様コピュラとガウスコピュラの場合について具体的な計算結果を示す。最後に 6 節で今後の課題を述べる。

2. 簡単な例

簡単な例として、次のような 2 次元の区分一様分布を考えてみよう。正方形領域 $[0, 1]^2$ を 4 つの小さい正方形に分割し、それぞれの領域で定数となるようなコピュラ密度

$$c(x_1, x_2) = \begin{cases} 1.8 & \text{if } (x_1, x_2) \in [0, \frac{1}{2}]^2 \cup [\frac{1}{2}, 1]^2, \\ 0.2 & \text{otherwise} \end{cases}$$

を考える (図 1(a))。また 1 次元の分布関数として

$$(2.1) \quad F_1(\xi) = F_2(\xi) = \begin{cases} \frac{2}{3}\xi & \text{if } \xi \in [0, \frac{3}{4}), \\ \frac{1}{2} + 2(\xi - \frac{3}{4}) & \text{if } \xi \in [\frac{3}{4}, 1] \end{cases}$$

を選び、式 (1.1) によって密度 $p(x_1, x_2)$ を定義すれば、

$$p(x_1, x_2) = \begin{cases} 0.8 & \text{if } (x_1, x_2) \in [0, \frac{3}{4}]^2, \\ 0.8/3 & \text{if } (x_1, x_2) \in ([0, \frac{3}{4}) \times [\frac{3}{4}, 1]) \cup ([\frac{3}{4}, 1] \times [0, \frac{3}{4})) \\ 7.2 & \text{if } (x_1, x_2) \in [\frac{3}{4}, 1]^2, \end{cases}$$

となる (図 1(b))。 $c(x_1, x_2)$ と $p(x_1, x_2)$ は周辺分布を変えただけであり、式 (1.1) のコピュラ部分は共通であることに注意する。

ところが、いま $u(x_1, x_2) = 1$ を一様密度 (独立コピュラ) とすれば、 u と c の間の Kullback–Leibler ダイバージェンスは

$$\begin{aligned} \text{KL}(u, c) &= \int_{[0,1]^2} u(x) \log \frac{u(x)}{c(x)} dx \\ &= \frac{1}{2} \log \frac{1}{1.8} + \frac{1}{2} \log \frac{1}{0.2} \\ &\approx 0.511 \end{aligned}$$

となるのに対し、 u と p の間のダイバージェンスは

コピュラモデルにおけるダイバージェンスの性質

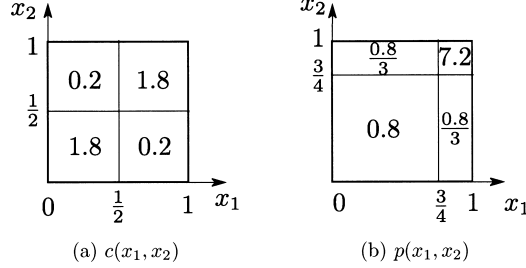


図 1. 周辺密度は異なるがコピュラ部分は等しい 2 つの密度関数. 各領域内の数値は密度関数の値を表す. 一様密度からの Kullback–Leibler ダイバージェンスは $c(x_1, x_2)$ よりも $p(x_1, x_2)$ の方が小さい.

$$\begin{aligned} \text{KL}(u, p) &= \left(\frac{3}{4}\right)^2 \log \frac{1}{0.8} + 2 \left(\frac{3}{4}\right) \left(\frac{1}{4}\right) \log \frac{1}{0.8/3} + \left(\frac{1}{4}\right)^2 \log \frac{1}{7.2} \\ &\approx 0.498 \end{aligned}$$

となる. つまり, u から見て c よりも p の方が少し近いことが分かる.

このように, 周辺分布を動かすことでダイバージェンスの値は変化する. その最小値を本稿ではプロファイルダイバージェンスと呼ぶ. 一方, 次節で詳しく見るように, セミパラメトリックコピュラモデルの情報は順位統計量に集約されていると考えることができる. そこで順位統計量に基づいて定義されるダイバージェンスを順位ダイバージェンスと呼ぶ. これらの関係を明らかにすることが本研究の主題である.

3. 順位ダイバージェンスとプロファイルダイバージェンス

ここではまず, セミパラメトリックコピュラモデルが座標ごとの変数変換に関して不変であることを説明する (Hoff, 2007; Hoff et al., 2014). よって複合変換モデル (付録 A) として扱うことができる. この観点から順位ダイバージェンスとプロファイルダイバージェンスという 2 つのダイバージェンスを定義する. また両者の関係を定理として与える.

3.1 セミパラメトリックコピュラモデル

セミパラメトリックコピュラモデルを改めて定義しよう. $[0, 1]^d$ 上の正値確率密度関数の全体を \mathcal{P} と表す. ただし密度関数 $p(x)$ が正値であるとはほとんどいたるところで $p(x) > 0$ という意味とする. また座標ごとの変数変換 $T(x) = (T_1(x_1), \dots, T_d(x_d))$ のうち, 各 $T_i : [0, 1] \rightarrow [0, 1]$ が単調増加, 全単射かつ T_i も T_i^{-1} も絶対連続であるようなものの全体を \mathcal{T} と表す. たとえば式 (2.1) で考えた周辺分布の組 (F_1, F_2) は \mathcal{T} の元である. \mathcal{T} は関数の合成に関して群をなす. 密度 $p \in \mathcal{P}$ と変換 $T \in \mathcal{T}$ に対し, 密度の押し出し $T_*p \in \mathcal{P}$ を

$$(3.1) \quad (T_*p)(x_1, \dots, x_d) = p(T_1^{-1}(x_1), \dots, T_d^{-1}(x_d)) \prod_{i=1}^d (T_i^{-1})'(x_i)$$

によって定義する. これは p に従う確率変数ベクトルを T で変数変換したときに得られる密度関数である. この写像 $(T, p) \mapsto T_*p$ によって \mathcal{T} は \mathcal{P} に作用する. この作用に関する軌道 (同値

類)を $[p] = \{T_*p \mid T \in \mathcal{T}\}$ と表す. Sklar の定理より, 各軌道にはコピュラ密度がただ一つ存在する. つまり, コピュラを指定することは軌道を指定することと同じである.

ここではコピュラと変換群の整合性を考えて $[0, 1]^d$ 上の密度関数に限定したが, \mathbb{R}^d 上の密度関数を扱うことももちろん可能である. その場合, \mathbb{R}^d から $(0, 1)^d$ への変換をあらかじめ一つ決めておけば, 結果として \mathbb{R}^d から \mathbb{R}^d への変数変換を扱うことができる.

以上の記法のもとで, セミパラメトリックコピュラモデルは,

$$(3.2) \quad \mathcal{M} = \{T_*c_\theta \mid \theta \in \Theta, T \in \mathcal{T}\}$$

と定義される. ただし $\{c_\theta \mid \theta \in \Theta\}$ はパラメトリックなコピュラ密度の族である. パラメトリックコピュラのいろいろな例は Nelsen (2006) を参考にされたい. 本稿では 4 節と 5 節でそれぞれ区分一様コピュラとガウスコピュラを扱う.

モデル \mathcal{M} は \mathcal{T} の作用に関して複合変換モデルとなる. 興味のあるパラメータは θ である. つまり, 個々の分布ではなく軌道に興味がある. そこで, 2 つの軌道 $[p], [q]$ の間のダイバージェンスを定義したい.

3.2 順位ダイバージェンス

$p \in \mathcal{P}$ を真の密度とし, p に従うランダム標本 $X = (x_{ti})_{1 \leq t \leq n, 1 \leq i \leq d}$ を考える. ここで n はサンプルサイズである. 連続分布を考えているので, 各 i に対して $\{x_{ti}\}_{t=1}^n$ は相異なる値を取るものと仮定してよい. 次の命題は 1 次元の場合にはよく知られており (たとえば Eaton, 1983; 吉田, 2006), 多次元の場合にも同様に示される. 最大不変量の定義は付録 A を参照せよ.

命題 1 (Hoff, 2007). セミパラメトリックコピュラモデルにおける最大不変量は, 多変量順位統計量

$$r_{ti} = \#\{s \in \{1, \dots, n\} \mid x_{si} \leq x_{ti}\}, \quad 1 \leq t \leq n, \quad 1 \leq i \leq d,$$

である. ただし $\#A$ は集合 A の要素数を表す. また行列表記で $R = (r_{ti})$ と表す.

以下では多変量順位統計量を単に順位統計量と呼ぶ. 真の密度が p のとき, 順位統計量 R の周辺分布を $\bar{p}_n(R)$ と表し, 順位尤度 (rank likelihood) と呼ぶ. R の取り得る値は有限個 (正確には $(n!)^d$ 個) であるから, \bar{p}_n は離散分布であることに注意しよう.

順位尤度は高次元の積分によって記述することができる. 実際, 順位統計量 R と矛盾しない標本 $X \in \mathbb{R}^{n \times d}$ の集合を $\{R(X) = R\}$ と表せば, 密度 p に対する順位尤度は

$$(3.3) \quad \bar{p}_n(R) = \int_{\{R(X)=R\}} \prod_{t=1}^n p(x_{t1}, \dots, x_{td}) dX$$

と書ける. 4 節で述べる通り, 区分一様分布の場合には積分を含まない形で書き下すことができる.

順位尤度を用いて, 順位ダイバージェンスを次のように定義する.

定義 1. 密度関数 $p, q \in \mathcal{P}$ と標本サイズ n に対し,

$$(3.4) \quad \begin{aligned} D_n([p], [q]) &= \frac{1}{n} \text{KL}(\bar{p}_n, \bar{q}_n) \\ &= \frac{1}{n} \sum_R \bar{p}_n(R) \log \frac{\bar{p}_n(R)}{\bar{q}_n(R)} \end{aligned}$$

を順位ダイバージェンス (rank divergence) という.

$D_n([p], [q])$ が well-defined であることは、 $\bar{p}_n(R)$ が $[p]$ の元の選び方に依存しないことから言える。また式(3.4)の右辺において n で割っているのは、順位統計量 R が $O(n)$ の情報を持っていることを考慮したものである。Kullback–Leibler ダイバージェンスの単調性と加法性 (Kullback, 1959) から

$$(3.5) \quad D_n([p], [q]) \leq \frac{1}{n} \text{KL}(p_n, q_n) = \text{KL}(p, q)$$

が成り立つ。ここで p_n, q_n はそれぞれ p, q からの標本 $X \in \mathbb{R}^{n \times d}$ の同時密度関数を表す。

順位ダイバージェンスは正值性(識別性)を満たさない。すなわち $[p] \neq [q]$ であっても $D_n([p], [q]) = 0$ となる場合がある。実際、次のような例が存在する。2次元コピュラ密度 p で対称性 $p(x_1, x_2) = p(1 - x_1, x_2)$ を満たし、かつ一様でないものを選ぶ(たとえば $p(x_1, x_2) = 1 + \cos(2\pi x_1) \cos(2\pi x_2)$ など)。このとき式(3.3)と p の対称性より $\bar{p}_2(R)$ は離散一様分布となる。しかし一様密度 u の順位尤度 $\bar{u}_2(R)$ も離散一様分布となるから $D_2([p], [u]) = 0$ となり、正值性が成り立たない。

一方、「全ての n に対して $D_n([p], [q]) = 0$ 」となるような $[p] \neq [q]$ は(正則条件のもとで)存在しない。これは後の定理1と定理2より示される。

3.3 プロファイルダイバージェンス

複合変換モデル(付録A)の類推により、次のダイバージェンスを考えることは自然である。

定義 2. 密度関数 $p, q \in \mathcal{P}$ に対してプロファイルダイバージェンスを

$$(3.6) \quad \begin{aligned} \tilde{D}([p], [q]) &= \inf_{T, U \in \mathcal{T}} \text{KL}(T_*p, U_*q) \\ &= \inf_{T \in \mathcal{T}} \text{KL}(T_*p, q) \end{aligned}$$

と定義する。2番目の等号は Kullback–Leibler ダイバージェンスの不変性 $\text{KL}(T_*p, T_*q) = \text{KL}(p, q)$ より従う。

p, q がともにコピュラ密度であったとしても一般には $\tilde{D}([p], [q]) < \text{KL}(p, q)$ となる。2節で挙げた例はそのような例である。

$\tilde{D}([p], [q]) = \text{KL}(p, q)$ が成り立つための p, q の条件を調べておこう。計算の便宜上、写像 $T \in \mathcal{T}$ による密度 $q \in \mathcal{P}$ の引き戻しを

$$(3.7) \quad (T^*q)(x) = q(T_1(x_1), \dots, T_d(x_d)) \prod_{i=1}^d T'_i(x_i)$$

と定義する。これは式(3.1)で定義した押し出しの逆演算である。 $\text{KL}(T_*p, q) = \text{KL}(p, T^*q)$ とすることに注意しよう。

命題 2. p, q は $(0, 1)^d$ において連続的微分可能な(コピュラ密度とは限らない)確率密度関数とする。このとき $\tilde{D}([p], [q]) = \text{KL}(p, q)$ が成り立つための必要条件は

$$(3.8) \quad \partial_i \log p_i(x_i) = \mathbb{E}_p[\partial_i \log q(x)|x_i], \quad i = 1, \dots, d,$$

である。ただし p_i は p の周辺密度、 ∂_i は x_i による偏微分、 $\mathbb{E}_p[\cdot]$ は密度 p に関する条件付き期待値を表す。さらに、 q が対数凹(log-concave)であれば式(3.8)は十分条件にもなる。

証明. 変分法による。 $T_i(x_i) = x_i + \delta T_i(x_i)$ とおき、 $\delta T_i(x_i)$ は開区間 $(0, 1)$ 内にコンパクト

なサポートを持つ滑らかな関数とする. $\text{KL}(p, T^*q)$ を δT_i について 1 次まで展開すると

$$(3.9) \quad \text{KL}(p, T^*q) = \int p(x) \log \frac{p(x)}{q(T(x)) \prod_i T'_i(x_i)} dx \\ \simeq \text{KL}(p, q) - \sum_i \int p(x) (\partial_i \log q(x)) \delta T_i(x_i) dx - \sum_i \int p(x) \delta T'_i(x_i) dx$$

$$(3.10) \quad = \text{KL}(p, q) + \sum_i \int \{-p(x) \partial_i \log q(x) + \partial_i p(x)\} \delta T_i(x_i) dx \\ = \text{KL}(p, q) + \sum_i \int p_i(x_i) \{E_p[-\partial_i \log q(x)|x_i] + \partial_i \log p_i(x_i)\} \delta T_i(x_i) dx_i$$

となる. ただし (3.10) の等号は部分積分と境界条件 $\delta T_i(0) = \delta T_i(1) = 0$ による. よって停留条件 (3.8) を得る. また q が対数凹ならば, \mathcal{T} 上の汎関数 $T \mapsto \text{KL}(p, T^*q)$ は式 (3.9) より凸になるので, 停留条件が満たされればそれが最適解となる. \square

上の証明の最後で, 集合 \mathcal{T} の凸性, および汎関数 $T \mapsto \text{KL}(p, T^*q)$ の凸性を用いた. これらは最適輸送理論の文脈では displacement convexity と呼ばれる性質である (McCann, 1997; Villani, 2003). また命題 2 は p, q のサポートが $[0, 1]^d$ でない場合にも同様に成立する.

命題 2 において特に q を $[0, 1]^d$ 上の一様密度とおけば, 式 (3.8) は $\partial_i p_i = 0$, つまり p がコピュラ密度であることと同値である. またこのとき $\text{KL}(p, q) = \int p(x) \log p(x) dx$ はエントロピーの -1 倍である. こうして次の結果を得る.

系 1. $p \in \mathcal{P}$ がコピュラ密度であるための必要十分条件は, p が軌道 $[p]$ の中でエントロピー最大となることである.

3.4 主結果

順位ダイバージェンスとプロファイルダイバージェンスに関する主定理を示す. 式 (3.5) より, 両者には一般に $D_n([p], [q]) \leq \tilde{D}([p], [q])$ という関係が成り立つことに注意しよう.

まずプロファイルダイバージェンスの正值性に関して次の定理が成り立つ.

定理 1. p, q は正值のコピュラ密度関数とし, q は上に有界かつ上半連続とする. このとき $\tilde{D}([p], [q]) = \text{KL}(p, T^*q)$ を満たす $T \in \mathcal{T}$ が存在する. 特に $p \neq q$ ならば $\tilde{D}([p], [q]) > 0$ となる.

以下の証明には最適輸送理論の帰結が用いられる. そこでは最小化問題 $\tilde{D}([p], [q]) = \inf_{T \in \mathcal{T}} \text{KL}(T^*p, q)$ が輸送写像 T に関するエネルギー最小化問題と解釈される.

証明. $p = q$ の場合は T を恒等写像とすればよい. 以下 $p \neq q$ とする. プロファイルダイバージェンスの定義より, $\text{KL}(p_m, q)$ が $\tilde{D}([p], [q])$ に収束するような密度関数の列 $\{p_m\}_{m=1}^\infty \subset [p]$ が取れる. $\{p_m\}$ は確率分布列としてタイトであるから, 最初から弱収束先が存在すると仮定してよい. ところで仮定より $q(x)$ は有界だから $q(x) \leq M$ とすれば

$$\int p_m(x) \log p_m(x) dx = \text{KL}(p_m, q) + \int p_m(x) \log q(x) dx \\ \leq \text{KL}(p_m, q) + \log M$$

となる. よって

$$\liminf_{m \rightarrow \infty} \int p_m(x) \log p_m(x) dx \leq \tilde{D}([p], [q]) + \log M < \infty$$

が成り立つ. すると p_m の弱収束先は絶対連続となる (McCann, 1997, Corollary 3.5). この収束先の密度関数を p_∞ とおけば

$$\liminf_{m \rightarrow \infty} \int p_m(x) \log p_m(x) dx \geq \int p_\infty(x) \log p_\infty(x) dx$$

が成り立つ (McCann, 1997, Lemma 3.4). また $q(x)$ の上半連続性より

$$\liminf_{m \rightarrow \infty} \int p_m(x) \log(1/q(x)) dx \geq \int p_\infty(x) \log(1/q(x)) dx$$

も成り立つ (van der Vaart, 2000). 以上から

$$\tilde{D}([p], [q]) = \lim_{m \rightarrow \infty} \int p_m(x) \log \frac{p_m(x)}{q(x)} dx \geq \int p_\infty(x) \log \frac{p_\infty(x)}{q(x)} dx > 0$$

を得る. 最後の不等号は p_∞ と q が異なる密度であることによる. 実際に $p_\infty \neq q$ であることは以下のようにして示される. まず $p_\infty \notin \mathcal{P}$ の場合 (p_∞ のサポートが $[0, 1]^d$ の真部分集合の場合) は明らかに $p_\infty \neq q$ である. よって $p_\infty \in \mathcal{P}$ と仮定してよい. このとき $p_\infty \in [p]$ であることを示せば十分である ($q \notin [p]$ より). さて p_m の弱収束先が p_∞ であったから, p_m の周辺分布も p_∞ の周辺分布に弱収束する. すると, $p_m = (T_m)_* p$ を満たす写像列 $T_m \in \mathcal{T}$ は, ある $T \in \mathcal{T}$ に一様収束する (T は p_∞ の周辺分布から定まる). したがって $p_m = (T_m)_* p$ は $T_* p$ に弱収束する (例えば吉田, 2006 の定理 1.54). よって $p_\infty = T_* p \in [p]$ が従う. 以上で定理 1 が示された. □

定理 1 における q の有界性は証明の都合によるものであり, 必要ではない. 実際, ガウスコピュラの場合はこの仮定が満たされないが, プロファイルダイバージェンスの正值性は直接確認できる (5 節). 定理の条件を緩めることは今後の課題である.

次の結果は有限次元の複合変換モデルに対して成り立つ事実 (付録 A) の類推である.

定理 2. p, q は正値かつ連続なコピュラ密度関数とし, ある正の定数 C_0, C_1 が存在して任意の $x \in [0, 1]^d$ に対して $C_0 \leq p(x), q(x) \leq C_1$ を満たすものとする. 真の密度が p のとき, $n \rightarrow \infty$ のもとで確率 1 で

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{\bar{p}_n(R)}{\bar{q}_n(R)} = \tilde{D}([p], [q])$$

が成立する. また順位ダイバージェンスはプロファイルダイバージェンスに収束する:

$$\lim_{n \rightarrow \infty} D_n([p], [q]) = \tilde{D}([p], [q]).$$

以下の証明には, 4 節で示す区分一様コピュラに関する結果を用いる.

証明. 2次元の場合を証明する. まず2次元の区分一様密度を考える (式 (4.5) 参照). $p = p_\theta$, $q = p_\phi$ とおく. ここで $\theta = (\theta_{ij}), \phi = (\phi_{ij}) \in \mathbb{R}^{I \times J}$ であり, θ_{ij} および ϕ_{ij} は長方形領域 $A_{ij} = [\frac{i-1}{I}, \frac{i}{I}] \times [\frac{j-1}{J}, \frac{j}{J}]$ における定数の確率密度を表す. 真の密度が p であるとして $\bar{q}_n(R)$ の漸近形を求める. 定理 3 より

$$\bar{q}_n(R) = (IJ)^{-n} \sum_{\sigma} \sum_{\tau} \frac{1}{\prod_i \sigma_i! \prod_j \tau_j!} \prod_i \prod_j \phi_{ij}^{n_{ij}(R, \sigma, \tau)}$$

となる. ただし $n_{ij} = n_{ij}(R, \sigma, \tau)$ は長方形領域 A_{ij} に入る観測値の度数を表し, $\sigma = (\sigma_i), \tau = (\tau_j)$ は周辺度数を表す. $\hat{\pi}_{ij} = n_{ij}/n$ とおくと,

$$\log \bar{q}_n(R) = -n \log(IJ) + \log \left(\sum_{\sigma} \sum_{\tau} \frac{\prod_i \prod_j \phi_{ij}^{n\hat{\pi}_{ij}}}{\prod_i (n\hat{\pi}_{i+})! \prod_j (n\hat{\pi}_{+j})!} \right)$$

となる．ここで全ての非負整数 m に対して $m \log m - m \leq \log m! \leq (m+1) \log(m+1) - m$ が成り立つ(たとえば Feller (1968) の II.9 節)．また σ, τ の取り得る値の個数はそれぞれ $\binom{n+I-1}{n}$, $\binom{n+J-1}{n}$ であり，これは n に関して多項式オーダーである．よってラプラス近似を適用でき，

$$\frac{1}{n} \log((n!)^2 \bar{q}_n(R)) = \sup_{\sigma, \tau} \left(\sum_i \sum_j \hat{\pi}_{ij} \log \phi_{ij} - \sum_i \hat{\pi}_{i+} \log(I\hat{\pi}_{i+}) - \sum_j \hat{\pi}_{+j} \log(J\hat{\pi}_{+j}) \right) + o(1)$$

が言える．ここで $o(1)$ は $n \rightarrow \infty$ のもとで R に関して一様に 0 に収束する項である．さて経験測度 $\hat{P} = n^{-1} \sum_{t=1}^n \delta_{(x_{t1}, x_{t2})}$ を用いると，ある $T \in \mathcal{T}$ が存在して $\hat{\pi}_{ij} = \hat{\pi}_{ij}^T := \hat{P}(T^{-1}(A_{ij}))$ と書ける．ただし A_{ij} は上で定義した長方形領域である．さらに T は各 $T^{-1}(A_{ij})$ において線形としてよい．このとき写像 T を動かすことと σ, τ を動かすことは同値となる．よって

$$\frac{1}{n} \log((n!)^2 \bar{q}_n(R)) = \sup_T \left(\sum_i \sum_j \hat{\pi}_{ij}^T \log \phi_{ij} - \sum_i \hat{\pi}_{i+}^T \log(I\hat{\pi}_{i+}^T) - \sum_j \hat{\pi}_{+j}^T \log(J\hat{\pi}_{+j}^T) \right) + o(1)$$

が得られる．Glivenko–Cantelli の定理(たとえば van der Vaart, 2000)から，確率 1 で， T について一様に $\hat{\pi}_{ij}^T$ は $\pi_{ij}^T := P(T^{-1}(A_{ij}))$ に収束する．ここで P は真の確率分布，すなわち $P(dx) = p(x)dx$ である．よって確率 1 で

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log((n!)^2 \bar{q}_n(R)) = \sup_T \left(\sum_i \sum_j \pi_{ij}^T \log \phi_{ij} - \sum_i \pi_{i+}^T \log(I\pi_{i+}^T) - \sum_j \pi_{+j}^T \log(J\pi_{+j}^T) \right)$$

となる．一方， T が各 $T^{-1}(A_{ij})$ において線形であるから， $x \in T^{-1}(A_{ij})$ のとき $T'_1(x_1) = 1/(I\pi_{i+}^T)$ などが言えるので，

$$\begin{aligned} \text{KL}(p, T^*q) &= \int p(x) \log p(x) dx + \sum_i \sum_j \int_{T^{-1}(A_{ij})} p(x) \log \frac{1}{\phi_{ij} T'_1(x_1) T'_2(x_2)} dx \\ &= \int p(x) \log p(x) dx - \sum_i \sum_j \pi_{ij}^T \log \phi_{ij} + \sum_i \pi_{i+}^T \log(I\pi_{i+}^T) + \sum_j \pi_{+j}^T \log(J\pi_{+j}^T) \end{aligned}$$

となる．以上から

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log((n!)^2 \bar{q}_n(R)) = \int p(x) \log p(x) dx - \inf_T \text{KL}(p, T^*q)$$

が示された． $\bar{p}_n(R)$ についても同じことが言えるので，結局

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\frac{\bar{p}_n(R)}{\bar{q}_n(R)} \right) &= -\inf_T \text{KL}(p, T^*p) + \inf_T \text{KL}(p, T^*q) \\ &= \inf_T \text{KL}(p, T^*q) \\ &= \check{D}([p], [q]) \end{aligned}$$

が得られる．最後の等号は命題 3 による．

p, q が区分一様とは限らない有界連続密度の場合は，いったん密度を区分一様密度で近似すればよい．つまり，任意の $\varepsilon > 0$ に対して $(1-\varepsilon)p_\varepsilon \leq p \leq (1+\varepsilon)p_\varepsilon$ を満たす区分一様密度 p_ε

が存在するので (q も同様), これに基づいて順位尤度を評価すれば

$$\frac{1}{n} \log \frac{\bar{p}_n(R)}{\bar{q}_n(R)} \leq \frac{1}{n} \log \frac{(1+\varepsilon)^n \bar{p}_{\varepsilon,n}(R)}{(1-\varepsilon)^n \bar{q}_{\varepsilon,n}(R)} \rightarrow \log \frac{1+\varepsilon}{1-\varepsilon} + \tilde{D}([p_\varepsilon], [q_\varepsilon]) \quad (n \rightarrow \infty)$$

が成り立つ (下側の評価も同様). この不等式で $\varepsilon \rightarrow 0$ とすればよい.

順位ダイバージェンスの収束は有界収束定理より従う. 実際, 仮定から $C_0 \leq p(x) \leq C_1$ であるから, $C_0^n \leq \bar{p}_n(R) \leq C_1^n$ が成り立つ. $\bar{q}_n(R)$ についても同様なので $n^{-1} \log(\bar{p}_n(R)/\bar{q}_n(R))$ は有界である. \square

この定理により, セミパラメトリックコピュラモデルの漸近的な性質はプロファイルダイバージェンスが担っていると考えられることができる.

4. 区分一様コピュラ

前節で定義した順位ダイバージェンスやプロファイルダイバージェンスは一般には陽に計算できない. しかし区分一様分布やガウス分布に限れば計算を進めることができる. 本節では区分一様分布を考える. 2次元に限定して議論を進めるが, 多次元でも同様の結果が成り立つ. なお区分一様分布はチェス盤分布 (chessboard distribution) とも呼ばれる (Ghosh and Henderson, 2001).

4.1 プロファイルダイバージェンス

I, J を正の整数とし, $[0, 1]^2$ を $I \times J$ 個の小長方形に分けてそれぞれ $A_{ij} = \left[\frac{i-1}{I}, \frac{i}{I}\right) \times \left[\frac{j-1}{J}, \frac{j}{J}\right)$ ($1 \leq i \leq I, 1 \leq j \leq J$) とおく. 密度関数 p が区分一様 (piecewise uniform) であるとは,

$$p(x_1, x_2) = p_{ij} \quad \text{if } (x_1, x_2) \in A_{ij}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J$$

と書けることとする. ここで p_{ij} は $\sum_{i=1}^I \sum_{j=1}^J p_{ij}/IJ = 1$ を満たす正の実数である. 区分一様密度がコピュラ密度であるための必要十分条件は

$$\sum_{i=1}^I \frac{p_{ij}}{I} = 1 \quad (1 \leq j \leq J), \quad \sum_{j=1}^J \frac{p_{ij}}{J} = 1 \quad (1 \leq i \leq I)$$

となることである. したがって区分一様コピュラ密度全体は $(I-1)(J-1)$ 次元のパラメトリックモデルとなる. 一見すると分割表モデルと同じだが, 2節で既に論じたように, ダイバージェンスの構造は異なるものとなる.

次の命題から, q が区分一様である限り, プロファイルダイバージェンス $\tilde{D}([p], [q])$ の計算は有限次元の最適化問題に帰着されることが分かる.

命題 3. p を任意のコピュラ密度, q を区分一様コピュラ密度とし, 領域 A_{ij} における $q(x)$ の値を q_{ij} とする. このとき $\tilde{D}([p], [q]) = \text{KL}(p, T^*q)$ を達成する $T = (T_1, T_2) \in \mathcal{T}$ は

$$(4.1) \quad \xi_i = T_1^{-1}(i/I), \quad \eta_j = T_2^{-1}(j/J) \quad (1 \leq i \leq I-1, \quad 1 \leq j \leq J-1)$$

を節点とする区分線形変換 T_1, T_2 で与えられる. ここで ξ_i, η_j は次の最小化問題の解である:

$$(4.2) \quad \begin{aligned} \text{Minimize} \quad & \sum_i \sum_j \left(\int_{\xi_{i-1}}^{\xi_i} \int_{\eta_{j-1}}^{\eta_j} p(x) dx \right) \log \frac{1}{q_{ij}} \\ & + \sum_i (\xi_i - \xi_{i-1}) \log(\xi_i - \xi_{i-1}) + \sum_j (\eta_j - \eta_{j-1}) \log(\eta_j - \eta_{j-1}) \end{aligned}$$

subject to $0 = \xi_0 < \xi_1 < \cdots < \xi_I = 1$, $0 = \eta_0 < \eta_1 < \cdots < \eta_J = 1$.

この目的関数は $\text{KL}(p, T^*q)$ と定数差を除いて等しい.

証明. $\{\xi_i\}, \{\eta_j\}$ を固定し, 式(4.1)を満たす T の中で $\text{KL}(p, T^*q)$ が最小となるのは区分線形変換であることを示す. このとき q の区分一様性から $q(T(x))$ は T の選び方によらないので

$$(4.3) \quad \begin{aligned} \text{KL}(p, T^*q) &= \int p(x) \log \frac{p(x)}{q(T(x))T'_1(x_1)T'_2(x_2)} dx \\ &= (\text{const.}) - \int_0^1 \log T'_1(x_1) dx_1 - \int_0^1 \log T'_2(x_2) dx_2 \end{aligned}$$

となる. ただしコピュラ密度の条件 $\int p(x) dx_2 = \int p(x) dx_1 = 1$ を用いた. さて, 一般に $t(0) < t(1)$ を固定したもとの $-\int_0^1 \log t'(x) dx$ を最小化する単調増加関数 $t(x)$ は一次式である. 実際, $-\log$ の凸性から $-\int_0^1 \log t'(x) dx \geq -\log \int_0^1 t'(x) dx = -\log(t(1) - t(0))$ となり, 等号は $t''(x) = 0$ のとき成立する. 以上から, T_1, T_2 は区分線形である. 式(4.2)の目的関数は, $x_1 \in [\xi_{i-1}, \xi_i]$ のとき $T'_1(x_1) = 1/(I(\xi_i - \xi_{i-1}))$ などに注意すれば, 式(4.3)より得られる. \square

式(4.2)の最適化問題は一般に非凸であり, 解の一意性は成り立たない. しかし定理 1 より解は必ず存在する.

以下, $I = J = 2$ の場合を詳しく見てみよう. すなわち次の形のコピュラ密度を考える:

$$c_\theta(x) = \begin{cases} 1 + \theta & \text{if } x \in [0, 1/2]^2 \cup [1/2, 1]^2, \\ 1 - \theta & \text{otherwise.} \end{cases}$$

ただし $-1 < \theta < 1$ とする. 2 節で考えたコピュラ密度もこの形であった. $\theta = 0$ の場合は一様密度となる.

実は, 次の命題に示すような「分岐現象」が成り立つ. ξ_1, η_1 の定義は上と同様である.

命題 4. $\theta > 0$ とする. このとき一様密度 u と c_θ の間のプロファイルダイバージェンスは

$$\tilde{D}([u], [c_\theta]) = \begin{cases} \text{KL}(u, c_\theta) & \text{if } 0 < \theta \leq \tanh(1), \\ \text{KL}(u, T^*c_\theta) & \text{if } \tanh(1) < \theta < 1 \end{cases}$$

となる. ここで $T = (T_1, T_2)$ は区分線形変換であり, その節点は $\xi_1 = \eta_1 = \xi$ または $1 - \xi$ で与えられる. ただし ξ は次の方程式の一意的な解である:

$$(4.4) \quad \xi = \frac{1}{2} \left(1 + \frac{\log \frac{\xi}{1-\xi}}{\log \frac{1+\theta}{1-\theta}} \right), \quad \frac{1}{2} < \xi < 1.$$

証明. 式(4.2)の目的関数を $f(\xi_1, \eta_1)$ とおく. f の停留点の方程式は

$$\eta_1 = \frac{1}{2} \left(1 + \frac{\log \frac{\xi_1}{1-\xi_1}}{\log \frac{1+\theta}{1-\theta}} \right), \quad \xi_1 = \frac{1}{2} \left(1 + \frac{\log \frac{\eta_1}{1-\eta_1}}{\log \frac{1+\theta}{1-\theta}} \right)$$

となる. この方程式は $|\log \frac{1+\theta}{1-\theta}| \leq 2$ すなわち $|\theta| \leq \tanh(1)$ のときは $(\xi_1, \eta_1) = (1/2, 1/2)$ という一意解を持ち, $|\theta| > \tanh(1)$ のときは $(1/2, 1/2)$ の他に対称な 2 つの解 $(\xi, \xi), (1 - \xi, 1 - \xi)$ を持つことが示される. ただし ξ は式(4.4)の解である. また f のヘッセ行列は

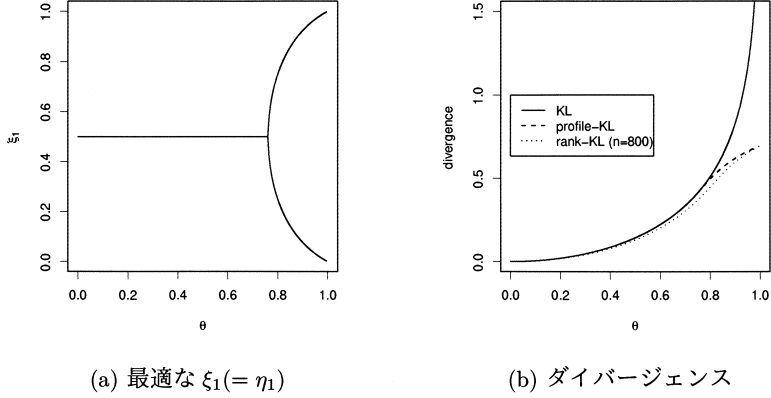


図 2. 分岐現象. (a) $\theta = \tanh(1)$ を境にして, 最適な写像 T は切り替わる. (b) ダイバージェンス $KL(u, c_\theta)$, 順位ダイバージェンス $D_n([u], [c_\theta])$ ($n = 800$), プロファイルダイバージェンス $\bar{D}([u], [c_\theta])$ を θ の関数としてプロットしたもの.

$$\begin{pmatrix} \frac{1}{\xi_1(1-\xi_1)} & 2 \log \frac{1-\theta}{1+\theta} \\ 2 \log \frac{1-\theta}{1+\theta} & \frac{1}{\eta_1(1-\eta_1)} \end{pmatrix}$$

となる. 特に $\theta > \tanh(1)$ ならば $\xi_1 = \eta_1 = 1/2$ は極小点とならず, $\xi_1 = \eta_1 = \xi$ および $\xi_1 = \eta_1 = 1 - \xi$ が極小点となる. \square

式(4.4)を θ について解くと

$$\theta = \frac{\left(\frac{\xi}{1-\xi}\right)^{1/(2\xi-1)} - 1}{\left(\frac{\xi}{1-\xi}\right)^{1/(2\xi-1)} + 1}, \quad \frac{1}{2} < \xi < 1$$

となる. たとえば $\xi = \frac{3}{4}$ とすると, $\theta = 0.8$ となる. これが 2 節で与えた例である.

命題 4 から, 一様密度からのダイバージェンスに関して, $\theta = \tanh(1)$ を境として分岐現象が生ずる. この様子を図示したものが図 2 である. なお, $\theta \rightarrow 1$ のとき $\bar{D}([u], [c_\theta])$ は $\log 2$ に近づくことも示せる. 特にプロファイルダイバージェンスは有界となる. これは, $\theta \rightarrow 1$ のとき $KL(u, c_\theta) \rightarrow \infty$ となることに比べると対照的である.

4.2 区分一様コピュラの順位尤度

区分一様コピュラのもとで, 順位尤度を求める. 区分一様コピュラモデル

$$(4.5) \quad p_\theta(x_1, x_2) = \theta_{ij} \quad \text{if } (x_1, x_2) \in A_{ij}$$

を考える. ただし A_{ij} は前項で定義した小領域であり, θ_{ij} は $\sum_i \theta_{ij}/I = 1, \sum_j \theta_{ij}/J = 1$ を満たすものとする. このモデルに従うランダム標本 $X = \{(x_{t1}, x_{t2})\}_{t=1}^n$ の同時密度関数は

$$(4.6) \quad (IJ)^{-n} \prod_{i=1}^I \prod_{j=1}^J \theta_{ij}^{n_{ij}}$$

となる. ここで n_{ij} は各 A_{ij} に入った観測値の度数である. 度数分布 (n_{ij}) は (θ_{ij}) の十分統計量となる. ただし実際には X は観測されず順位統計量 R だけが観測されるので, (n_{ij}) は潜在変数となる.

度数分布 (n_{ij}) の周辺度数を $\sigma_i = n_{i+} = \sum_j n_{ij}$, $\tau_j = n_{+j} = \sum_i n_{ij}$ とおく. $\sigma = (\sigma_i)$, $\tau = (\tau_j)$ が与えられたもとでは, R が決まれば (n_{ij}) が確定する. これを $n_{ij} = n_{ij}(R, \sigma, \tau)$ と表す.

定理 3. 区分一様コピュラ密度に対し, 順位尤度は

$$(4.7) \quad \bar{p}_n(R) = (IJ)^{-n} \sum_{\sigma} \sum_{\tau} \frac{1}{\prod_i \sigma_i! \prod_j \tau_j!} \prod_i \prod_j \theta_{ij}^{n_{ij}(R, \sigma, \tau)}$$

となる. ここで σ, τ は周辺度数全体を動くものとする.

証明. $\bar{p}_n(R)$ を単に $p(R)$ と表す. 式(4.6)より, $N := (n_{ij})$ の周辺分布は

$$p(N) = \frac{n!}{\prod_i \prod_j n_{ij}!} (IJ)^{-n} \prod_i \prod_j \theta_{ij}^{n_{ij}},$$

で与えられる. 次に N を固定したもとの R の条件付き分布 $p(R|N)$ を求める. N は θ に関して十分統計量であるから, $p(R|N)$ は θ には依存しない. したがって, X が一様分布から生成されたと仮定して $p(R|N)$ を計算すればよい. さて σ, τ は N だけで決まるから,

$$\begin{aligned} p(R|N) &= p(R|N, \sigma, \tau) \\ &= \frac{p(R, N|\sigma, \tau)}{p(N|\sigma, \tau)} \\ &= \begin{cases} \frac{p(R|\sigma, \tau)}{p(N|\sigma, \tau)} & \text{if } N = N(R, \sigma, \tau), \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

となる. いま X が一様分布からのサンプルであるとしているから, $p(N|\sigma, \tau)$ は超幾何分布となり, また $p(R|\sigma, \tau) = p(R) = 1/(n!)^2$ となる. 以上から

$$\begin{aligned} p(R) &= \sum_N p(R|N)p(N) \\ &= \sum_{\sigma} \sum_{\tau} p(R|\sigma, \tau) \frac{1}{p(N|\sigma, \tau)} p(N) \Bigg|_{N=N(R, \sigma, \tau)} \\ &= \sum_{\sigma} \sum_{\tau} \frac{1}{(n!)^2} \frac{n! \prod_i \prod_j n_{ij}!}{\prod_i \sigma_i! \prod_j \tau_j!} \frac{n!}{\prod_i \prod_j n_{ij}!} (IJ)^{-n} \prod_i \prod_j \theta_{ij}^{n_{ij}} \Bigg|_{N=N(R, \sigma, \tau)} \end{aligned}$$

となり, これを整理すれば式(4.7)が得られる. \square

図 2(b)にある順位ダイバージェンスのグラフは定理 3 の結果とモンテカルロ法を利用して計算したものである.

5. ガウスコピュラの場合

ガウスコピュラとは, 座標ごとの変数変換によって多変量正規分布となるようなコピュラのことである. 2つのガウスコピュラの間のプロファイルダイバージェンスは簡単な形になる. また順位尤度は多次元正規分布の象限確率の計算に帰着される. これらの結果を述べる.

なお, 3 節で注意した通り, \mathbb{R}^d 上の 2つの密度関数の間のダイバージェンスは $(0, 1)^d$ 上の密度関数に帰着させて定義する. しかし結果として \mathbb{R}^d 上の変数変換を考えればよいことになるので $(0, 1)^d$ を意識する必要はない.

5.1 プロファイルダイバージェンス

ガウスコピュラに対してはプロファイルダイバージェンスが有限次元の凸最適化によって計算できる。これを示そう。平均ベクトル 0 、共分散行列 Σ の正規分布の密度関数を ϕ_Σ と記す。またベクトル u を対角成分とする対角行列を $\text{diag}(u)$ と表す。

命題 5. P, Q を d 次正定値対称行列とし、 $p = \phi_P, q = \phi_Q$ とおく。このとき $\tilde{D}([p], [q]) = \text{KL}(p, T^*q)$ を満たす T は線形変換 $T(x) = \text{diag}(u)x$ で与えられる。ここで $u = (u_i)$ は次の凸最適化問題の解である：

$$(5.1) \quad \begin{aligned} \text{Minimize} \quad & - \sum_i \log u_i + \frac{1}{2} \text{tr}(Q^{-1} \text{diag}(u) P \text{diag}(u)) \\ \text{subject to} \quad & u_1, \dots, u_d > 0. \end{aligned}$$

この目的関数は $\text{KL}(p, T^*q)$ と定数差を除いて等しい。

証明. q は対数凹である。よって、 $T(x) = \text{diag}(u)x$ が命題 2 の停留条件

$$\partial_i \log p_i(x_i) = \mathbb{E}_p[(\partial_i \log(T^*q))(x)|x_i]$$

を満たすことをチェックすれば十分である。左辺は $\partial_i \log p_i(x_i) = -P_{ii}^{-1}x_i$ である。右辺は

$$\begin{aligned} \mathbb{E}_p[(\partial_i \log(T^*q))(x)|x_i] &= - \sum_j u_j (Q^{-1})_{ij} u_j \mathbb{E}_p[x_j|x_i] \\ &= - \sum_j u_j (Q^{-1})_{ij} u_j P_{ji} P_{ii}^{-1} x_i \\ &= -P_{ii}^{-1} x_i \end{aligned}$$

となる。最後の等号は式(5.1)の停留条件による。式(5.1)が実際に一意解を持つことは Marshall and Olkin (1968) で示されている。□

$d = 2$ の場合は、密度 p, q に対応する相関係数をそれぞれ ρ_p, ρ_q とおけば、

$$\tilde{D}([p], [q]) = \log \frac{1 - \rho_p \rho_q}{\sqrt{(1 - \rho_p^2)(1 - \rho_q^2)}}$$

となる。これは対称なダイバージェンスである。また、Fisher の Z 変換 $\rho = \tanh z$ を施すと、

$$\tilde{D}([p], [q]) = \log \cosh(z_p - z_q)$$

となり、平行移動に関する不変性が成り立つ。

ガウスコピュラに限れば、プロファイルダイバージェンスから導かれる計量は、Segers et al. (2014) によって求められている有効情報量に一致する。これを確かめよう。ただし一般にモデル $\{p_\theta \mid \theta \in \mathbb{R}^m\}$ に対し、ダイバージェンスから導かれる計量を

$$\tilde{g}_{ij} = - \left. \frac{\partial^2 \tilde{D}([p_\theta], [p_\phi])}{\partial \theta_i \partial \phi_j} \right|_{\theta=\phi}, \quad i, j \in \{1, \dots, m\},$$

と定義する (Eguchi, 1983; Amari, 1985)。

定理 4. 共分散行列のパラメトリックモデル $\{P = P_\theta \mid \theta \in \mathbb{R}^m\}$ に対し、計量は

$$(5.2) \quad \tilde{g}_{ij} = \frac{1}{2} \text{tr}\{P^{-1}(\partial_i P - \Pi(\partial_i P))P^{-1}(\partial_j P - \Pi(\partial_j P))\}$$

で与えられる。ただし $\partial_i = \partial/\partial\theta_i$ とし、射影 $\Pi: \mathbb{R}^{d \times d} \rightarrow \mathbb{R}^{d \times d}$ は以下で定義される：

$$(5.3) \quad \Pi(A) = P \operatorname{diag}(b(A)) + \operatorname{diag}(b(A))P,$$

$$(5.4) \quad b(A) = (P^{-1} \circ P + I)^{-1}(P^{-1} \circ A)1_d, \quad 1_d = (1, \dots, 1)^\top \in \mathbb{R}^d.$$

$A \circ B$ は行列 A, B の要素ごとの積(アダマール積)を表す。

証明. 多変量正規分布のフィッシャー情報行列が

$$g_{ij} = \frac{1}{2} \operatorname{tr}(P^{-1}(\partial_i P)P^{-1}(\partial_j P))$$

であることはよく知られている。また一般に式(A.2)の意味でのプロファイルダイバージェンスに対して、計量は

$$\tilde{g}_{ij} = E[\{\partial_i \ell - \Pi(\partial_i \ell)\}\{\partial_j \ell - \Pi(\partial_j \ell)\}]$$

となる(Barndorff-Nielsen and Jupp (1988), Theorem 7.2)。ここで $\ell = \log p_{\theta, \nu}$ であり、 Π は局外接空間 $\operatorname{span}\{\frac{\partial \ell}{\partial \nu}\}$ への直交射影である。接ベクトルの対応 $\partial_i P \mapsto \partial_i \ell$ は線形同型であるから、式(5.2)が得られる。あとは直交射影 Π を求めればよい。いま局外パラメータ u の入り方は $\operatorname{diag}(u)P \operatorname{diag}(u)$ であり、 $u = 1_d$ における接ベクトルは

$$\partial_{u_k}(\operatorname{diag}(u)P \operatorname{diag}(u)) = \operatorname{diag}(e_k)P + P \operatorname{diag}(e_k), \quad k = 1, \dots, d,$$

である。ただし e_k は第 k 単位ベクトルである。よって直交射影 Π は式(5.3)の形となる。また $\Pi(A) = PB + BP, B = \operatorname{diag}(b)$ とおくと、任意の対角行列 $C = \operatorname{diag}(c)$ に対して

$$\begin{aligned} 0 &= \frac{1}{2} \operatorname{tr}\{P^{-1}(PC + CP)P^{-1}(A - (PB + BP))\} \\ &= \sum_k c_k (P^{-1}A)_{kk} - \sum_k c_k b_k - \sum_k \sum_l c_k b_l (P^{-1})_{kl} P_{kl} \end{aligned}$$

となる。よって $(P^{-1} \circ A)1_d - (I + P^{-1} \circ P)b = 0$ となり、式(5.4)を得る。□

5.2 ガウスコピュラの順位尤度

ガウスコピュラの場合、順位尤度の式(3.3)が正規分布の象限確率で表される。これは積分領域 $\{R(X) = R\}$ が

$$x_{t(s,i),i} < x_{t(s+1,i),i}, \quad s \in [n-1], \quad i \in [d]$$

と表されることによる。ただし $t(s,i)$ は順位が $r_{ti} = s$ となる観測値番号 t を表す。また $[n] = \{1, \dots, n\}$ と略記した。結果として次の定理が得られる。

定理 5. 共分散行列 Σ の d 次元正規分布に対する順位尤度は

$$\bar{p}_n(R|\Sigma) = \frac{1}{n^{d/2} |\Sigma|^{(n-1)/2} |B|^{1/2}} \int_{\mathbb{R}_+^{(n-1)d}} \phi_{B^{-1}}(w) dw$$

で与えられる。ただし $\mathbb{R}_+ = (0, \infty)$ であり、 $\phi_{B^{-1}}$ は平均ベクトル 0 、共分散行列 B^{-1} の多次元正規分布の密度関数、 B は次で定義される $(n-1)d$ 次の正定値対称行列である：

$$B_{(r-1)d+i, (s-1)d+j} = (\Sigma^{-1})_{ij} \left(\sum_{t=1}^n I_{\{r_{ti} \leq r, r_{tj} \leq s\}} - \frac{rs}{n} \right), \quad r, s \in [n-1], \quad i, j \in [d].$$

証明. $S = \Sigma^{-1}$ と記す。変数変換 $v_i = x_{t(n,i),i}$, $w_{ri} = x_{t(r+1,i),i} - x_{t(r,i),i}$ を考える。それぞれ

れの積分範囲は $v_i \in \mathbb{R}$, $w_{ri} > 0$ となる. 逆変換は $x_{ti} = v_i - \sum_{r=1}^{n-1} I_{\{r_{ti} \leq r\}} w_{ri}$ で与えられる. I は定義関数である. よって X の同時密度関数の指数部は,

$$\begin{aligned}
 \sum_{t=1}^n \sum_{i=1}^d \sum_{j=1}^d S_{ij} x_{ti} x_{tj} &= \sum_{t=1}^n \sum_{i=1}^d \sum_{j=1}^d S_{ij} \left(v_i - \sum_{r=1}^n I_{\{r_{ti} \leq r\}} w_{ri} \right) \left(v_j - \sum_{s=1}^n I_{\{r_{tj} \leq s\}} w_{sj} \right) \\
 &= \sum_i \sum_j \left(n S_{ij} v_i v_j - 2n S_{ij} m_i v_j + \sum_r \sum_s S_{ij} I_{\{r_{ti} \leq r\}} I_{\{r_{tj} \leq s\}} w_{ri} w_{sj} \right)
 \end{aligned}$$

となる. ただし

$$m_i = \frac{1}{n} \sum_{t=1}^n \sum_{r=1}^{n-1} I_{\{r_{ti} \leq r\}} w_{ri} = \frac{1}{n} \sum_{r=1}^{n-1} r w_{ri}$$

とおいた. よって $(v_i), (w_{ri})$ の同時密度関数の指数部は

$$n \sum_i \sum_j S_{ij} (v_i - m_i)(v_j - m_j) + \sum_i \sum_j \sum_r \sum_s B_{(r-1)d+i, (s-1)d+j} w_{ri} w_{sj}$$

となる. v_i について積分をすれば結果が得られる. \square

多次元正規分布の象限確率はホロノミック勾配法によって高精度に計算できることが知られており (Koyama and Takemura, 2015), R のパッケージとして公開されている (Koyama et al., 2014). よって定理 5 を用いて順位尤度および順位ダイバージェンスを計算することができる. ホロノミック勾配法の計算量はホロノミックランクと呼ばれる量が目安となる. m 次元正規分布の象限確率におけるホロノミックランクは 2^m であり, サンプルサイズが大きくなると急激に計算量は大きくなる.

表 1 および表 2 は共分散行列が

$$(5.5) \quad \Sigma = \Sigma(\theta) = \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}, \quad \theta \in \{0.0, 0.1, \dots, 0.9, 0.95\},$$

で与えられる 2 次元ガウスコピュラについて, それぞれ $n = 2$ と $n = 3$ の場合の順位尤度を求めたものである. ただし対称性から順位統計量 R の第 1 列 (第 1 成分の順位統計量) を

表 1. $n = 2$ の場合の順位尤度. ただし $r_{t1} = t$ ($\forall t$) とし, 結果を $n!$ 倍している.

θ	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
$(r_{t2}) = (1, 2)$.5000	.5319	.5641	.5970	.6310	.6667	.7048	.7468	.7952	.8564	.8989
$(2, 1)$.5000	.4681	.4359	.4030	.3690	.3333	.2952	.2532	.2048	.1436	.1011

表 2. $n = 3$ の場合の順位尤度. ただし $r_{t1} = t$ ($\forall t$) とし, 結果を $n!$ 倍している.

θ	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
$(r_{t2}) = (1, 2, 3)$.1667	.1918	.2196	.2509	.2866	.3280	.3773	.4381	.5179	.6359	.7302
$(1, 3, 2)$.1667	.1780	.1880	.1964	.2028	.2067	.2068	.2017	.1875	.1549	.1212
$(2, 1, 3)$.1667	.1780	.1880	.1964	.2028	.2067	.2068	.2017	.1875	.1549	.1212
$(2, 3, 1)$.1667	.1542	.1407	.1262	.1109	.0947	.0777	.0597	.0408	.0210	.0106
$(3, 1, 2)$.1667	.1542	.1407	.1262	.1109	.0947	.0777	.0597	.0408	.0210	.0106
$(3, 2, 1)$.1667	.1439	.1230	.1037	.0859	.0693	.0537	.0391	.0253	.0123	.0061

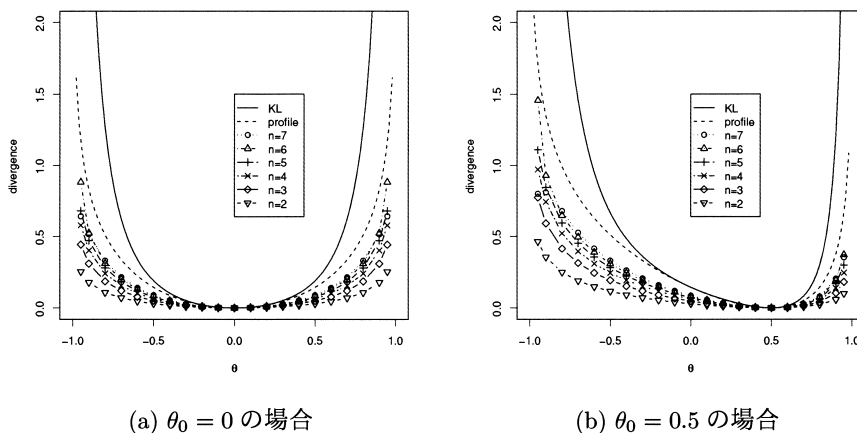


図 3. ガウスコピュラの順位ダイバージェンス $D_n([p_{\theta_0}], [p_{\theta}])$ を θ の関数としてプロットしたものの。

$(1, \dots, n)$ に固定し、代わりに確率の値を $n!$ 倍してある。結果として表の各列の和は 1 となる。

図 3 は同じく 2 次元のガウスコピュラで $\theta_0 = 0$ および $\theta_0 = 0.5$ の場合について、順位ダイバージェンス $D_n([p_{\theta_0}], [p_{\theta}])$ を θ の関数としてプロットしたものである。ここで p_{θ} は式 (5.5) の共分散行列を持つ 2 次元正規分布の密度関数、すなわち $\phi_{\Sigma(\theta)}$ を表す。サンプルサイズは $n = 2$ から $n = 7$ までとした。同図にはプロファイルダイバージェンス $\hat{D}([p_{\theta_0}], [p_{\theta}])$ および通常の Kullback-Leibler ダイバージェンス $KL(p_{\theta_0}, p_{\theta})$ も示した。理論上、 $KL \geq \hat{D} \geq D_n$ という関係があり、これは図においても確かに成立している。

θ を固定したとき、 $n = 2$ から $n = 7$ に増加するにしたがって、順位ダイバージェンスも増加する傾向がある。ただし $n = 7$ のときだけ $n \leq 6$ の曲線と交差している。これは $n = 7$ における計算精度が十分でない可能性が考えられる。今後の検討課題としたい。なお $n = 7$ の場合に要した計算時間は 3 時間程度であった。また計算機環境は Amazon Web Services (32 コア, インスタンスタイプ r3.8xlarge) であり、ソフトウェアは R, ホロノミック勾配法のパッケージとして hgm (Koyama et al., 2014) を用いた。

6. 今後の課題

本稿では順位ダイバージェンスとプロファイルダイバージェンスを定義し、前者は後者に収束することを述べた (定理 2)。しかしここではコピュラ密度関数が有界であることが仮定されており、多くの実用的なコピュラモデルには適用できない。そのような場合に定理が成り立つかどうかを調べる必要がある。この点について査読者より、Huber (1967) の 4 節にある領域分割の方法が役立つのではないかと指摘があった。すなわち、定理 2 の証明で使われている区分様コピュラの領域分割を等分割にしないという方法である。いまの目的では $[0, 1]^d$ の境界に近付くにつれて分割を細かくすることが自然である。そのためには無限個に領域分割した区分様コピュラの挙動を明らかにせねばならない。後の論理は定理 2 の証明と同様であると考えられる。

定理 4 ではガウスコピュラの場合に、プロファイルダイバージェンスから有効情報量が求められることを示した。このことは一般のコピュラモデルに拡張できるものと考えられるが、まだ示せていない。コピュラモデルの有効情報量は Sturm-Liouville 型の方程式で表され、普通

は陽に求められない(Bickel et al., 1993).

また本稿では最初から周辺分布 F_1, \dots, F_d に関する情報は全く得られない状況を仮定した. 現実問題としては周辺分布自体にも多少の興味はあるだろうし, 何らかの事前情報を入れる方が自然な場合も想定される. Chen et al. (2006)の推定量はその一例と解釈できる. このような場合に適切なダイバージェンスを導くことも一つの課題である.

確率分布の間には Kullback–Leibler ダイバージェンス以外にも様々なダイバージェンスが定義される. それらに対応する順位ダイバージェンスやプロファイルダイバージェンスを形式的に定義することは可能である. ただし変数変換不変性や加法性が成り立たない場合には注意深く検討する必要があるだろう.

謝 辞

本稿に対して多くの有益なコメントをくださった編集委員および査読者の皆様に深く感謝いたします. 本研究は JSPS 科研費 JP17K00044 および JP26108003 の助成を受けたものです. また, 研究の発端となる話題を提供してくれた慶應義塾大学(当時)の橋本研一氏に感謝します.

付 録

A. 複合変換モデル

複合変換モデル(composite transformation model)とは, 変数変換に対して不変な統計モデルのことであり, 本稿では次のように定義する. 標本空間 \mathcal{X} 上の確率分布族 $\{P_{\theta, \nu} \mid \theta \in \Theta, \nu \in N\}$ を考え, θ は興味のあるパラメータ, ν は局外パラメータを表すものとする. いま群 G が存在して, G は \mathcal{X} と N にそれぞれ作用し, かつ

$$x \sim P_{\theta, \nu}, \quad g \in G \quad \Rightarrow \quad gx \sim P_{\theta, g\nu}$$

という性質を満たすものとする. また G の N への作用は推移的であると仮定する. このときこの確率分布族 $\{P_{\theta, \nu}\}$ を複合変換モデルという(Barndorff-Nielsen and Jupp, 1988). 不変確率モデルとも呼ばれる(Eaton, 1983; 紙屋, 1999). 群 G の作用によって \mathcal{X} は互いに排反な軌道(orbit)に分解される. この軌道と一対一に対応する統計量を最大不変量(maximal invariant)という. 最大不変量の分布は θ のみに依存する.

例 1. 例として正規分布 $N(\mu, \sigma^2)$ に従うランダム標本 $x = (x_1, \dots, x_n)$ を考える. 標本をスケール変換して $x_i \mapsto ax_i$ ($a > 0$) とすると, パラメータは (μ, σ) から $(a\mu, a\sigma)$ に変換される. よって $N(\mu, \sigma^2)$ はスケール変換に関して複合変換モデルになっている. ここで興味のあるパラメータは $\theta = \mu/\sigma$ であり, 局外パラメータは $\nu = \sigma$ である. また観測値 $(x_1, \dots, x_n) \in \mathbb{R}^n$ を含む軌道は $\{(ax_1, \dots, ax_n) \mid a > 0\}$ となる. 最大不変量はたとえば $w(x) = (x_1/\hat{\sigma}, \dots, x_n/\hat{\sigma})$ が挙げられる. ただし $\hat{\sigma}^2$ は標本分散を表す. $w(x)$ の分布は θ のみに依存する. なお, 最小十分統計量 $(\bar{x}, \hat{\sigma})$ に対する最大不変量は $\bar{x}/\hat{\sigma}$ であり, これは Student の t 統計量の定数倍である(Cox and Hinkley, 1974, Example 5.16).

分布 $P_{\theta, \nu}$ (密度 $p_{\theta, \nu}$) に従う標本サイズ n のランダム標本 $x = (x_1, \dots, x_n)$ を考える. パラメータ θ に関する推測は最大不変量 $w = w(x)$ に基づいて行うのが自然である. そこで w の周辺密度関数を $\bar{p}_\theta(w)$ とおき, 2つの密度関数 $\bar{p}_{\theta_1}, \bar{p}_{\theta_2}$ の間のダイバージェンスを考える:

$$(A.1) \quad D_n(\theta_1, \theta_2) = \frac{1}{n} \int \bar{p}_{\theta_1}(w) \log \frac{\bar{p}_{\theta_1}(w)}{\bar{p}_{\theta_2}(w)} dw.$$

これを周辺ダイバージェンスと呼ぶことにしよう．右辺を n で割っているのは w が $O(n)$ の情報を持っているためである．一方で，局外パラメータに依存しないダイバージェンスとして

$$(A.2) \quad \tilde{D}(\theta_1, \theta_2) := \inf_{\nu_2 \in N} \text{KL}(p_{\theta_1, \nu_1}, p_{\theta_2, \nu_2}) \quad (\forall \nu_1)$$

が考えられる．右辺が ν_1 に依存しないのは Kullback–Leibler ダイバージェンスの不変性による．この \tilde{D} をプロファイルダイバージェンスと呼ぶことにする．Barndorff-Nielsen and Jupp (1988) の Section 4 では profile discrimination information と呼ばれている．次の命題は周辺分布をラプラス近似することによって得られ，実質的には同論文の Section 6 で示されている．

命題 6. 真のパラメータが θ のとき，標準的な仮定のもとで，

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{\bar{p}_\theta(w)}{\bar{p}_\phi(w)} = \lim_{n \rightarrow \infty} D_n(\theta, \phi) = \tilde{D}(\theta, \phi)$$

が成り立つ．

例 2 (つづき)．先の正規分布の例では，周辺ダイバージェンスは非心 t 分布の間のダイバージェンスとなり，陽には求まらない．しかしプロファイルダイバージェンスは

$$\begin{aligned} \tilde{D}(\theta_1, \theta_2) &= \inf_{\sigma_2 > 0} \left(-\frac{1}{2} \log \frac{\sigma_1^2}{\sigma_2^2} - \frac{1}{2} + \frac{(\theta_1 \sigma_1 - \theta_2 \sigma_2)^2 + \sigma_1^2}{2\sigma_2^2} \right) \\ &= -\log \left(\frac{\theta_1 \theta_2 + \sqrt{\theta_1^2 \theta_2^2 + 4(\theta_1^2 + 1)}}{2(\theta_1^2 + 1)} \right) - \frac{\theta_1 \theta_2}{2} \left(\frac{\theta_1 \theta_2 + \sqrt{\theta_1^2 \theta_2^2 + 4(\theta_1^2 + 1)}}{2(\theta_1^2 + 1)} \right) + \frac{\theta_2^2}{2} \end{aligned}$$

と陽に求められる．特に $\theta_1 = 0$ のときは $\tilde{D}(0, \theta_2) = \theta_2^2/2$ ，また $\theta_2 = 0$ のときは $\tilde{D}(\theta_1, 0) = (1/2) \log(\theta_1^2 + 1)$ となり，著しく非対称性を持ったダイバージェンスになっていることが分かる．これは t 検定における水準と検出力の関係の非対称性に対応する．このようにプロファイルダイバージェンスを調べれば，複合変換モデルの構造をある程度把握することができる．

参 考 文 献

- Amari, S. (1985). *Differential-Geometrical Methods in Statistics*, Springer, New York.
- Barndorff-Nielsen, O. E. and Jupp, P. E. (1988). Differential geometry, profile likelihood, L-sufficiency and composite transformation models, *The Annals of Statistics*, **16**(3), 1009–1043.
- Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, Baltimore.
- Chen, X., Fan, Y. and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models, *Journal of the American Statistical Association*, **101**(475), 1228–1240.
- Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*, Chapman and Hall, London.
- Eaton, M. L. (1983). *Multivariate Statistics — A Vector Space Approach*, Wiley, New York.
- Eguchi, S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family, *The Annals of Statistics*, **11**(3), 793–803.
- Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, 3rd ed., John Wiley & Sons, New York.
- Genest, C. and Werker, B. J. M. (2002). Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models, *Distributions with Given*

- Marginals and Statistical Modelling* (eds. C. M. Cuadras and J. A. R. Lallena), 103–112, Kluwer Academic, Dordrecht.
- Ghosh, S. and Henderson, S. G. (2001). Chessboard distributions and random vectors with specified marginals and covariance matrix, *Operations Research*, **50**(5), 820–834.
- Hoff, P. D. (2007). Extending the rank likelihood for semiparametric copula estimation, *The Annals of Applied Statistics*, **1**(1), 265–283.
- Hoff, P. D., Niu, X. and Wellner, J. A. (2014). Information bounds for Gaussian copulas, *Bernoulli*, **20**(2), 604–622.
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 221–233.
- 紙屋英彦 (1999). 不変確率モデルの特徴付け, *統計数理*, **47**(1), 63–69.
- Klaassen, C. A. J. and Wellner, J. A. (1997). Efficient estimation in the bivariate normal copula model: Normal margins are least favourable, *Bernoulli*, **3**(1), 55–77.
- Koyama, T. and Takemura, A. (2015). Calculation of orthant probabilities by the holonomic gradient method, *Japan Journal of Industrial and Applied Mathematics*, **32**, 187–204.
- Koyama, T., Nakayama, H., Ohara, K., Sei, T. and Takayama, N. (2014). Software packages for holonomic gradient method, *Mathematical Software—ICMS 2014* (eds. H. Hong and C. Yap), Lecture Notes in Computer Science, Vol. 8592, 706–712, Springer, Berlin.
- Kullback, S. (1959). *Information Theory and Statistics*, John Wiley and Sons, New York.
- Marshall, A. W. and Olkin, I. (1968). Scaling of matrices to achieve specified row and column sums, *Numerische Mathematik*, **12**, 83–90.
- McCann, R. J. (1997). A convexity principle for interacting gases, *Advances in Mathematics*, **128**, 153–179.
- Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd ed., Springer, New York.
- Segers, J., van den Akker, R. and Werker, B. J. M. (2014). Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation, *The Annals of Statistics*, **42**(5), 1911–1940.
- Tsukahara, H. (2005). Semiparametric estimation in copula models, *The Canadian Journal of Statistics*, **33**(3), 357–375.
- 塚原英敦 (2011). 接合分布関数 (コピュラ) の理論と応用, 『21 世紀の統計科学』 (北川源四郎, 竹村彰通編), 第 3 巻, 101–140, 日本統計学会, <http://park.itc.u-tokyo.ac.jp/atstat/iss75shunen/>.
- van der Vaart, A. W. (2000). *Asymptotic Statistics*, Cambridge University Press, Cambridge.
- Villani, C. (2003). *Topics in Optimal Transportation*, American Mathematical Society, Providence.
- 吉田朋広 (2006). 『数理統計学』, 朝倉書店, 東京.

Properties of Divergence for Semiparametric Copula Models

Tomonari Sei and Kazuya Matsumoto

Graduate School of Information Science and Technology, The University of Tokyo

A semiparametric copula model is a statistical model in which the copula is assumed to be parametric and the marginal distribution is arbitrary. In this paper, we investigate the divergence of the model. In particular, we establish a relationship between the rank divergence induced from the marginal distribution of the multivariate rank statistic and the profile divergence defined by infimum of the Kullback–Leibler divergence with respect to the nuisance parameter. We also obtain formulas for piecewise uniform and Gaussian copulas.