

尺度に基づく対称性からの逸脱度の可視化と 非対称モデルとの関係

東京理科大学大学院・創域理工学研究科 浦崎 航

Wataru Urasaki

Departure of Information Science, Tokyo University of Science

明星大学・データサイエンス学環 中川 智之

Tomoyuki Nakagawa

School of Data Science, Meisei University

東京理科大学・創域理工学部 田畑 耕治

Kouji Tahata

Departure of Information Sciences, Tokyo University of Science

1 導入

カテゴリカル変数は、医学、教育学、社会科学などの様々な分野で現れ、解析の対象となることが多い。特に、 r 個のカテゴリと c 個のカテゴリから構成される 2 種類のカテゴリカル変数を考えたとき、それらの組み合わせから得られる r 行 c 列の表を二元分割表と呼ぶ。分割表を用いた解析では、2 つの変数間に関係性が存在するかどうか、つまり、統計的独立性の評価を目的に、様々な手法が提案されてきた。その中で、Pearson のカイ二乗検定や Fisher の正確確率検定が有名であり、現在でも広く一般的に用いられている。これらの手法により、独立性がないと判断されたとき、2 つの変数間には何らかの連関性が存在すると考えるのが妥当である。そこで、連関性を調べるための様々な手法が提案されてきた。変数間の関係構造を連関モデルとして捉える方法に、Goodman [7] の一様連関モデル、Agresti [1] の線形-線形連関モデルなどが、そのほかにも、独立性からの逸脱度合を連関性の強度と見做し、 $[0, 1]$ 閉区間で強弱を評価するための連関尺度として、Cramér [5] のクラメール係数 V^2 などがある。

次に、表 1 のような二元分割表を考えたい。表 1 は、1, 2 回目のカフェイン抜きコーヒー購入時における 5 つのブランド選択、についてを調査してクロス集計したデータである。このような分割表は、行と列が同一分類であることから、一般的な分割表と異なる特徴を持ち、「正方」分割表という特別な呼ばれ方をしている。表 1 でも観られるように、正方分割表は分割表の対角セル付近に観測値が集中し、対角から離れるにつれて少なくなる傾向がある。そのため、正方分割表の 2 変数間の連関性が極めて高いことは明確であり、統計的独立性は成り立たない。そこで、正方分割表の解析

では、準独立性という主対角セル以外のセルに無関連構造を仮定することで、その性質を評価する Goodman [6] の準独立モデルなどが提案されている。

準関連性とは異なる正方分割表へのアプローチとして、対角成分以外に着目し、上、下三角成分の関係に対称性を見出すことで、変数間の移り変わりを捉える手法がある。対称性の解析は、Bowker [4] が対称モデルを導入したことが歴史的な始まりのようである。また、Bowker の対称モデルからの逸脱を、非対称構造として捉える非対称モデルが様々提案されており、McCullagh [10] の条件付き対称 (CS) モデル、Read [11] のグローバル対称 (GS) モデルなどが例に挙げられる。対称性についても、対称性からの逸脱の強度を評価する尺度が提案されている。その例に、Tomizawa *et al.* [13] の一般化尺度 $\Phi^{(\lambda)}$ があり、2 つの分布間の差異を定量化できる power-divergence [11] という情報量のクラスから提案されている。これらの先行研究は、正方分割表にどのような対称構造があるか、また、対称性からの逸脱の程度を調べる手段として画期的である。しかし、更なる応用に向けた問題として、これらは正方分割表全体にのみ焦点を当てるに留まるため、表 1 の各ブランドごとの対称性の詳細に着目したい場合には不十分である。

本研究の目的は、正方分割表全体の対称性を調べるだけでは考察が不十分なデータに対し、各カテゴリの対称性の数値化と二次元座標上への可視化から、新たな知見と解釈、それらの直感的理解の獲得を目指すことにある。特に、power-divergence 型尺度を取り上げ、power-divergence に包含される情報量のクラスならば、 $[0, 1]$ 閉区間での分割表全体の対称性の隔たりを、各カテゴリの部分的な対称性への分割と可視化が可能であることを示す。また、対称性からの逸脱がいくつかの非対称モデルへの分解が可能であることを踏まえ、逸脱が確認される場合には、どの非対称構造が関係し、影響を及ぼすかを評価する手法についても提案する。

表 1 カフェイン抜きコーヒー購入時のブランド選択データ (Grover and Srinivasan [9])

1 回目の購入	2 回目の購入					計
	High Pt	Taster's	Sanka	Nescafé	Brim	
High Point	93	17	44	7	10	171
Taster's Choice	9	46	11	0	9	75
Sanka	17	11	155	9	12	204
Nescafé	6	4	9	15	2	36
Brim	10	4	12	2	27	55
計	135	82	231	33	60	541

2 対称性の解析手法と逸脱度の可視化

2.1 対称性の解析手法

対称性についての先行研究として、本研究と密接な Bowker [4] の対称モデルと Tomizawa *et al.* [13] の power-divergence 型尺度 $\Phi^{(\lambda)}$ を紹介する。

$S \times S$ 正方分割表を考え、 p_{ij} ($i = 1, \dots, S; j = 1, \dots, S$) を i 行 j 列目のセル確率とする。また、 $p_{i\cdot}, p_{\cdot j}$ をそれぞれ行と列の周辺確率として、 $p_{i\cdot} = \sum_{j=1}^S p_{ij}, p_{\cdot j} = \sum_{i=1}^S p_{ij}$ とする。各セルごとの観測度数は n_{ij} とし、総観測度数を $n = \sum_{i=1}^S \sum_{j=1}^S n_{ij}$ とする。行と列カテゴリの 2 変数間における対称性の有無を評価する場合、Bowker の対称モデルを用いた検定として、次の帰無仮説 H_0 を考える。

$$H_0 : p_{ij} = p_{ji}; \forall i, j.$$

この帰無仮説 H_0 のもと、Bowker のカイ二乗検定統計量 χ_S^2 は

$$\begin{aligned} \chi_S^2 &= \frac{1}{2} \sum_{i=1}^S \sum_{j=1}^S \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}} \\ &= \sum_{i < j} \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}}, \end{aligned}$$

で与えられ、自由度 $S(S-1)/2$ のカイ二乗分布に従う。

その一方、尺度による対称性の解析方法として、power-divergence 型尺度 $\Phi^{(\lambda)}$ は以下のように定義される。

$$\Phi^{(\lambda)} = \frac{\lambda(\lambda+1)}{2^\lambda - 1} I^{(\lambda)}(\{p_{ij}^*\}; \{p_{ij}^s\})$$

ただし、 $\lambda > -1$ で

$$I^{(\lambda)}(\{p_{ij}^*\}; \{p_{ij}^s\}) = \frac{1}{\lambda(\lambda+1)} \sum_{i \neq j} \sum p_{ij}^* \left[\left(\frac{p_{ij}^*}{p_{ij}^s} \right)^\lambda - 1 \right],$$

である。 p_{ij}^*, p_{ij}^s は、 $\delta = \sum \sum_{i \neq j} p_{ij}$ としたとき、 $p_{ij}^* = p_{ij}/\delta, p_{ij}^s = (p_{ij}^* + p_{ji}^*)/2$ で与えられる。また、尺度 $\Phi^{(\lambda)}$ は以下の性質をもつ。

性質 1. $\lambda > -1$ に対して、尺度 $\Phi^{(\lambda)}$ は以下の 3 つの性質をもつ、

1. $0 \leq \Phi^{(\lambda)} \leq 1$.
2. $\Phi^{(\lambda)} = 0 \Leftrightarrow$ 正方分割表は完全な対称構造を持つ、つまり、全ての i, j について $p_{ij} = p_{ji}$ が成り立つ。
3. $\Phi^{(\lambda)} = 1 \Leftrightarrow$ 正方分割表は完全な非対称構造を持つ、つまり、全ての i, j ($i \neq j$) について対称なセル確率 p_{ij}, p_{ji} に対し、 $p_{ij} > 0$ かつ $p_{ji} = 0$ 、または、 $p_{ij} = 0$ かつ $p_{ji} > 0$ が成り立つ。

実際の正方分割表解析では、これらを含めた様々な手法から、観測されたデータの特徴を多角的に捉えていく。しかし、モデルや尺度によって調べられるのは正方分割表全体についてであり、表 1 のようなデータへの適用で調査できることは、全ブランドを一纏めにした 1, 2 回目の購入歴の移り変わりのみである。表 1 のようなデータの場合は、各ブランドごとの購入記録の遷移、つまり、全体

よりも個々のブランドの対称性と各ブランド間の関係性に関心が向けられる。以上を踏まえ、個々のカテゴリの対称性の評価と同時に可視化をすることで、専門分野に区別なく、一目で理解可能な対称性と関係性の把握を促す手法を提案する。

2.2 逸脱度の可視化

近年の分割表における研究では、独立、対称構造を問わず、様々な構造を個々のカテゴリで評価し、二次元座標上へ可視化することで、容易な直感的理解の獲得を目指す傾向がある。Beh and Lombardo [3] は、Bowker のカイ二乗統計量 χ^2_S の行列表現による可視化を提案したが、Pearson 距離の提案にのみ留まる。そこで、power-divergence 型尺度 $\Phi^{(\lambda)}$ を用いて、 $\lambda = 1$ におけるピアソン距離の調整版を含めた Beh and Lombardo [3] の一般化から、可視化の保証と統一的なスケールによる各カテゴリの対称性の逸脱度の可視化を提案する。

power-divergence 型尺度 $\Phi^{(\lambda)}$ は、以下のようにも表現される。

$$\begin{aligned}\Phi^{(\lambda)} &= \sum_{i \neq j} \sum \frac{p_{ij}^* + p_{ji}^*}{2} \left[1 - \frac{\lambda 2^\lambda}{2^\lambda - 1} H_{ij}^{(\lambda)}(\{p_{ij}^c, p_{ji}^c\}) \right] \\ &= \sum_{i \neq j} \sum \phi_{ij}^{(\lambda)},\end{aligned}$$

ただし、

$$\begin{aligned}H_{ij}^{(\lambda)}(\{p_{ij}^c, p_{ji}^c\}) &= \frac{1}{\lambda} [1 - (p_{ij}^c)^{\lambda+1} - (p_{ji}^c)^{\lambda+1}], \\ \phi_{ij}^{(\lambda)} &= \frac{p_{ij}^* + p_{ji}^*}{2} \left[1 - \frac{\lambda 2^\lambda}{2^\lambda - 1} H_{ij}^{(\lambda)}(\{p_{ij}^c, p_{ji}^c\}) \right],\end{aligned}$$

であり、 $p_{ij}^c = p_{ij}/(p_{ij} + p_{ji})$ とする。このとき、各 (i, j) セルごとの対称性からの逸脱は、 $\phi_{ij}^{(\lambda)}$ の非負値で表される。このとき、 $i < j$ で $\sqrt{\phi_{ij}^{(\lambda)}}$ を、 $i > j$ で $-\sqrt{\phi_{ij}^{(\lambda)}}$ を (i, j) 要素にもつ $S \times S$ 歪対称行列を $S_{skew(\lambda)}$ とすると、

$$\begin{aligned}\Phi^{(\lambda)} &= \sum_{i \neq j} \sum \phi_{ij}^{(\lambda)} \\ &= \text{trace}(S_{skew(\lambda)}^T S_{skew(\lambda)}),\end{aligned}$$

となることから、 $S_{skew(\lambda)}$ は部分的な対称性の情報を持ちながらも、尺度 $\Phi^{(\lambda)}$ を再構成できることがわかる。本研究では、 $S_{skew(\lambda)}$ からの二次元座標上へ可視化に必要な座標を導出する。そのために、歪対称行列 $S_{skew(\lambda)}$ への特異値分解を行うと、

$$S_{skew(\lambda)} = AD_\mu B^T,$$

として、3つに分解された行列を得ることができる。ここで、 A, B はそれぞれ $S_{skew(\lambda)}$ の左、右特異ベクトルを持つ $S \times M$ 直交行列、 D_μ は対角成分に降順で特異値 μ_m ($m = 1, \dots, M$) を持つ $M \times M$ 対角行列である。また、 M は行数 (または列数) S で変化し、 S が偶数ならば $M = S$ 、奇数

ならば $M = S - 1$ となる．これらの行列を用いて座標を得るが，同一カテゴリに関する座標の重み付けは Greenacre [8] の提案により，

$$\begin{cases} F = D^{-1/2} A D_\mu, \\ G = D^{-1/2} B D_\mu, \end{cases}$$

で与えられ，各行，列カテゴリごとの対称性からの隔たりを示す座標が得られる．ここで， $D = (D_R + D_C)/2$ であり， D_R, D_C はそれぞれ行と列の周辺確率 $p_{i\cdot}, p_{\cdot j}$ を対角成分に持つ $S \times S$ 対角行列である．また，尺度の値が 0 ならば完全な対称性が成り立つ性質を踏まえると，原点へプロットされるカテゴリは完全な対称性を持つことを示している．その一方で，導出された M 次元の座標における m 番目の主軸が，尺度の情報をどの程度反映しているかを示す寄与率を求める場合には，

$$\begin{aligned} \Phi^{(\lambda)} &= \text{trace}(S_{skew(\lambda)}^T S_{skew(\lambda)}) \\ &= \text{trace}((A D_\mu B^T)^T (A D_\mu B^T)) \\ &= \text{trace}(D_\mu^2), \end{aligned}$$

であることを踏まえ，

$$100 \times \frac{\mu_m^2}{\Phi^{(\lambda)}},$$

により与えられる．

ここまで座標や寄与率の導出を行ってきたが，各カテゴリの対称性の逸脱度を一点のみでしか表せていないため，精度面に不安が残る．そこで，各プロットの信頼領域についてを Beh [2] の方法から提案する．特異値分解から得られる直交行列と座標の関係から，二次元座標上での s 番目の行カテゴリの $100(1 - \alpha)\%$ 信頼領域は，行カテゴリの座標を示す行列 F の $(s, 1), (s, 2)$ 要素を f_{s1}, f_{s2} としたとき，次のように与えることができる．

$$\frac{(x - f_{s1})^2}{x_{s(\alpha)}^2} + \frac{(y - f_{s2})^2}{y_{s(\alpha)}^2} = 1$$

ただし，

$$\begin{aligned} x_{s(\alpha)} &= \left(\frac{\hat{p}_{s\cdot} + \hat{p}_{\cdot s}}{2} \right)^{-1/2} \mu_1 \sqrt{\frac{x_\alpha^2}{\hat{\Phi}^{(\lambda)} / \frac{\lambda(\lambda+1)}{2n\hat{\delta}(2^\lambda-1)}} \left(1 - \sum_{m=3}^M a_{sm}^2 \right)}, \\ y_{s(\alpha)} &= \left(\frac{\hat{p}_{s\cdot} + \hat{p}_{\cdot s}}{2} \right)^{-1/2} \mu_2 \sqrt{\frac{x_\alpha^2}{\hat{\Phi}^{(\lambda)} / \frac{\lambda(\lambda+1)}{2n\hat{\delta}(2^\lambda-1)}} \left(1 - \sum_{m=3}^M a_{sm}^2 \right)}, \end{aligned}$$

であり， $\hat{\Phi}^{(\lambda)}, \hat{\delta}$ は多項分布の下で $\{p_{ij}\}$ を最尤推定量 $\{\hat{p}_{ij}\}$ に置き換えた $\Phi^{(\lambda)}, \delta$ の plug-in 推定量， $\hat{p}_{s\cdot}, \hat{p}_{\cdot s}$ は $p_{s\cdot}, p_{\cdot s}$ の最尤推定量である． a_{sm} については，直交行列 A の (s, m) 要素である．また， x_α^2 は自由度 $S(S - 1)/2$ のカイ二乗分布における上側 $\alpha\%$ 点である．一般的な $r \times c$ 分割表において，この信頼領域は楕円形で与えられる．しかし，正方分割表における信頼領域の特徴として，歪対称行列から得られる特異値が $\mu_1 = \mu_2$ となることから，楕円形ではなく円形で与えられる．

2.3 数値実験

これまでに紹介した対称性の可視化手法により, power-divergence 型尺度 $\Phi^{(\lambda)}$ の可視化の振る舞いを示す. また, パラメータ λ 変化時の座標や寄与率の差異がどのような解釈を与えるかも示す. 解析用データについては, 表 1 を使用するとともに, パラメータに $\lambda = -1/2, 0, 1$ を与える. 各パラメータにおける power-divergence は, 特別な名前を持つ情報量であり, $\lambda = -1/2, 0, 1$ について, それぞれ Hellinger distance, Kullback-Leibler divergence, Pearson divergence と呼ばれる.

図 1-3 は, それぞれのパラメータで解析した結果である. 全てのパラメータについては Brim が原点に近い場所にプロットされており, $\lambda = 0, 1$ においては Nescafé も原点に近い. その一方で, High Point や Taster's Choice は原点から離れたところにプロットされている. これらのことから, Brim や Nescafé は強い対称性を持つため, 1, 2 回目の選択の移り変わりに違いが少ないことを示しており, High Point や Taster's Choice は何らかの影響を受け, 選択に違いが生じたと考えられる. 次に, それぞれのプロット間距離と信頼領域に着目すると, どのパラメータにおいても, Taster's Choice と Sanka, Brim と Nescafé は近い位置にあり, 信頼領域の重なる範囲も大きい. これは, 購入選択の遷移の傾向が似ていることを示しており, それぞれのペアとなるブランドは似た特徴や販売戦略などを持つのではないかと考察できる. また, 原点と信頼領域の被覆関係について, $\lambda = -1/2, 1$ の場合は全ブランドの信頼領域が原点を含まない一方, $\lambda = 0$ の場合では全てが含んでいることが確認できる. これは, 寄与率とパラメータごとの特徴が反映されており, $\lambda = 0$ は二次の軸までの寄与率の合計が約 90% に対し, $\lambda = -1/2, 1$ は約 70, 80% であることや, 各情報量での逸脱を測る感度に起因すると考えられる. これまでの結果と考察を踏まえると, パラメータ選択が課題として取り上げられる. 現実的なパラメータの選択方法として, 有名な情報量で寄与率が最大になるものを選ぶ, データの背景を踏まえた情報量の特徴を考慮して決定する, などが挙げられる. しかし, 現在でも数理的に妥当な方法は提案されておらず, 本研究でもパラメータの選択は重要な課題の 1 つである.

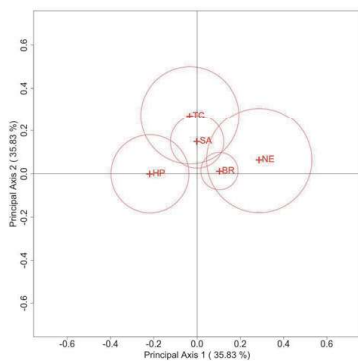


図 1 $\lambda = -1/2$ の場合

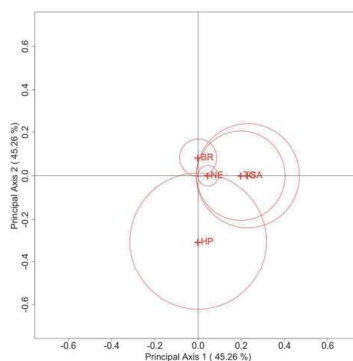


図 2 $\lambda = 0$ の場合

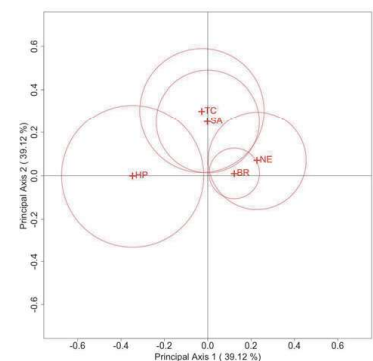


図 3 $\lambda = 1$ の場合

3 可視化と非対称モデルとの関係

3.1 非対称モデル

対称モデルといくつかの非対称モデルには、分解可能な関係にあるものが存在する。その関係を持つ非対称モデルの例に、CS モデルと GS モデルがある。McCullagh [10] により提案された CS モデルは、 $\delta > 0$ のもと、

$$p_{ij} = \begin{cases} \delta \psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases} \quad \text{ただし, } \psi_{ij} = \psi_{ji}$$

で与えられ、Read [11] の GS モデルは、

$$\delta_U = \delta_L,$$

ただし、 $\delta_U = \sum \sum_{i < j} p_{ij}$, $\delta_L = \sum \sum_{i > j} p_{ij}$ で与えられる。一目見てわかるように、CS モデルは定数倍の逸脱構造があることを仮定し、GS モデルは上、下三角成分で釣り合いが取れることを仮定している。

対称性の尺度 $\Phi^{(\lambda)}$ と同様に、CS モデルと GS モデルの逸脱度を評価する尺度 Φ_{CS} , Φ_{GS} がそれぞれ提案されており、

$$\begin{aligned} \Phi_{CS} &= \frac{1}{\log 2} \sum_{i \neq j} \frac{p_{ij}^* + p_{ji}^*}{2} \left\{ \log \left(\frac{\delta p_{ij}^*}{\delta_U (p_{ij}^* + p_{ji}^*)} \right) + \log \left(\frac{\delta p_{ji}^*}{\delta_L (p_{ij}^* + p_{ji}^*)} \right) \right\}, \\ \Phi_{GS} &= \frac{1}{\log 2} \sum_{i \neq j} \frac{p_{ij}^* + p_{ji}^*}{2} \left\{ 1 - \frac{1}{\log 2} H_{ij}^{(0)} \left(\left\{ \frac{\delta_U}{\delta}, \frac{\delta_L}{\delta} \right\} \right) \right\}, \end{aligned}$$

で定義される (Tahata *et al.* [12]). これらの尺度の性質についても、 $[0, 1]$ 閉区間で逸脱の強度を評価し、完全な CS モデルや GS モデルの構造を持つ場合、それぞれ $\Phi_{CS} = 0$, $\Phi_{GS} = 0$ となり、その一方、CS モデルや GS モデルとの隔たりが最大となる構造を持つ場合、それぞれ $\Phi_{CS} = 1$, $\Phi_{GS} = 1$ となる。これらに対し、 $\Phi^{(\lambda)}$ 、特に $\lambda = 0$ の場合を Φ_S と置くと、

$$\Phi_S = \Phi_{CS} + \Phi_{GS}$$

の分解関係が成り立つ。この関係に基づき、各カテゴリの Φ_{CS} , Φ_{GS} の影響を可視化から調べたいが、1つの問題が生じる。それは、 Φ_{GS} は $\Phi^{(\lambda)}$ と同様に、 Φ_{GS} が導ける歪対称行列を構成可能だが、 Φ_{CS} は常に構成できないことである。そのため、直接的な各カテゴリの GS モデルと CS モデルとの逸脱の評価を同時に行えないので、 Φ_S , Φ_{GS} を別々に可視化することで、間接的な評価を下す。本稿は、 Φ_{CS} , Φ_{GS} による CS モデルと GS モデルへの分解を取り上げたが、そのほかの分解可能な非対称モデルでも同様に考えることができる。

3.2 数値実験

3.2.1 シミュレーション

シミュレーション実験として、完全な CS モデル, GS モデルの構造を持つ場合の 4×4 確率表を用いて, Φ_S , Φ_{GS} による可視化を行った. 図 4-5 は CS モデルを, 図 6-7 では GS モデルを仮定した場合の結果を示している.

図 4-5 を見ると, 両方のプロットが完全に一致していることがわかる. これは, 完全な CS モデルの構造を持つとき, CS モデルからの逸脱は存在しないために $\Phi_{CS} = 0$ であることから, $\Phi_S = \Phi_{GS}$ が成り立つとともに, Φ_S と Φ_{GS} を導出できる歪対称行列が等しくなることが理由である. その一方で, 図 6-7 はプロットにかなりの乖離があることが見てわかる. 完全な GS モデルの構造を持つとき, GS モデルからの逸脱は存在しないため, $\Phi_{GS} = 0$ は明らかである. そのため, Φ_{GS} を導く歪対称行列は零行列となることから, 図 7 が示すように全カテゴリが原点に集中している. また, 図 7 の寄与率が NaN の理由は, 零行列の特異値が全て 0 となるためである. 図 6 は, $\Phi_S = \Phi_{CS}$ を踏まえ, Φ_{CS} による歪対称行列が構成可能な特殊な例であり, GS モデルの構造が成り立つことにより, CS モデルにおける δ , δ_U , δ_L が一意に定まるため, S モデルの逸脱と同様に表せることを示している. これらの結果は, 非対称モデルとの関係を直接的に評価できないが, 間接的に十分な評価が可能であることを示唆している.

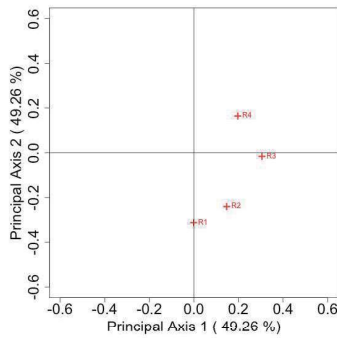


図 4 Φ_S の可視化 (完全な CS モデル)

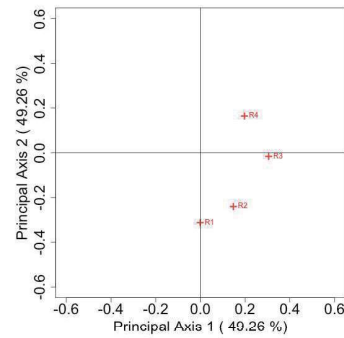


図 5 Φ_{GS} の可視化 (完全な CS モデル)

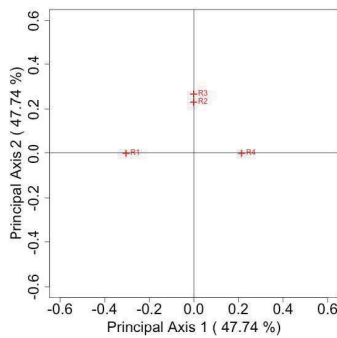


図 6 Φ_S の可視化 (完全な GS モデル)

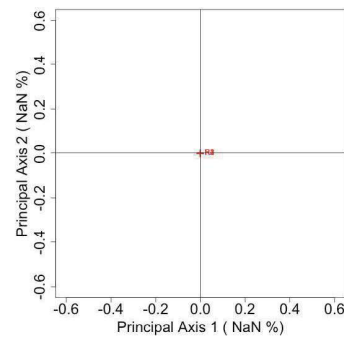


図 7 Φ_{GS} の可視化 (完全な GS モデル)

3.2.2 実データ

実データへの適用例として、表 1 を再び考える。2.3 節の $\lambda = 0$ における図 2 の結果から、対称性からの逸脱が大きいカテゴリがいくつかあることが確認できるが、それらの逸脱に CS モデルと GS モデルがどのように関係するかを考察する。

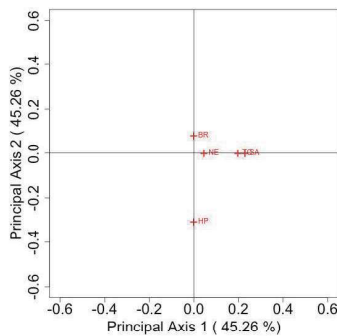


図 8 Φ_S の可視化 ($\hat{\Phi}_S = 0.0596$)

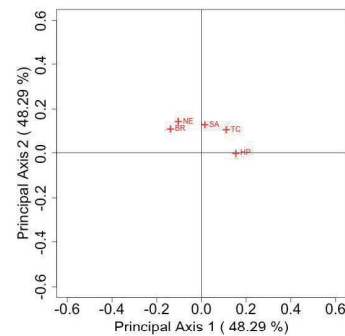


図 9 Φ_{GS} の可視化 ($\hat{\Phi}_{GS} = 0.0236$)

図 8-9 は、表 1 を $\hat{\Phi}_S$, $\hat{\Phi}_{GS}$ により可視化した結果である。図 9 を見ると、全てのカテゴリに GS モデルからの逸脱が同程度存在することがわかる。しかし、 $\hat{\Phi}_S = 0.0596$, $\hat{\Phi}_{GS} = 0.0236$ を踏まえると、対称性からの逸脱構造が GS モデルからの逸脱だけでなく、CS モデルからもある程度存在すると判断できる。そこで、図 8-9 の各カテゴリと原点の距離を観察すると、High Point が大きく異なることが確認できる。このことは、High Point は対称性からの逸脱が存在するなかで、CS モデルからの逸脱を他ブランドよりも強く持つ特徴がある、と考えることができる。

4 まとめと今後の課題

本稿は、対称性からの逸脱度を定量的に測る divergence 型尺度 $\Phi^{(\lambda)}$ の適用から、二次元座標上への可視化手法の提案を行うことで、信頼領域も絡めて各カテゴリの対称性を評価するとともに、各カテゴリ間の関係性も考察可能であることを紹介した。また、対称性からの逸脱が見られる場合には、いくつかの非対称モデルへの分解関係を踏まえ、どのような非対称構造が存在するかの間接的評価が十分に可能であることを確認した。今後の課題としては、本研究手法における power-divergence に与えるべき最適なパラメータ選択や、非対称モデルへの分解を踏まえた直接的な可視化による評価への理論的保証を与えたいと考えている。

謝辞

RIMS 共同研究「確率モデルと統計的推測」にて、貴重な公演の機会を頂きまして、誠にありがとうございました。また、本研究における有益なコメントを頂いた、京都女子大学の土田潤先生に感謝を申し上げます。最後に、本研究は JSPA 科研費 20K03756 の助成を受けたものです。

参考文献

- [1] Alan Agresti. A simple diagonals-parameter symmetry and quasi-symmetry model. *Statistics & probability letters*, Vol. 1, No. 6, pp. 313–316, 1983.
- [2] Eric J Beh. Elliptical confidence regions for simple correspondence analysis. *Journal of Statistical Planning and Inference*, Vol. 140, No. 9, pp. 2582–2588, 2010.
- [3] Eric J Beh and Rosaria Lombardo. Visualising departures from symmetry and Bowker's X^2 statistic. *Symmetry*, Vol. 14, No. 6, p. 1103, 2022.
- [4] Albert H Bowker. A test for symmetry in contingency tables. *Journal of the american statistical association*, Vol. 43, No. 244, pp. 572–574, 1948.
- [5] Harald Cramér. *Mathematical Methods of Statistics*. Princeton university press, 1946.
- [6] Leo A Goodman. The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing entries: Ra fisher memorial lecture. *Journal of the American Statistical Association*, Vol. 63, No. 324, pp. 1091–1131, 1968.
- [7] Leo A Goodman. Multiplicative models for square contingency tables with ordered categories. *Biometrika*, Vol. 66, No. 3, pp. 413–418, 1979.
- [8] Michael Greenacre. Correspondence analysis of square asymmetric matrices. *Journal of the Royal Statistical Society Series C: Applied Statistics*, Vol. 49, No. 3, pp. 297–310, 2000.
- [9] Rajiv Grover and V Srinivasan. A simultaneous approach to market segmentation and market structuring. *Journal of Marketing Research*, Vol. 24, No. 2, pp. 139–153, 1987.
- [10] Peter McCullagh. A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika*, Vol. 65, No. 2, pp. 413–418, 1978.
- [11] Timothy RC Read and Noel AC Cressie. *Goodness-of-fit Statistics for Discrete Multivariate Data*. Springer Science & Business Media, 1988.
- [12] Kouji Tahata, Ryohei Auch, Shuji Ando, and Sadao Tomizawa. Separation of the refined estimator of the measure for symmetry in square contingency tables. *Communications in Statistics-Simulation and Computation*, pp. 1–17, 2023.
- [13] Sadao Tomizawa, Takashi Seo, and Hideharu Yamamoto. Power-divergence-type measure of departure from symmetry for square contingency tables that have nominal categories. *Journal of Applied Statistics*, Vol. 25, No. 3, pp. 387–398, 1998.

〒 278-8510 千葉県野田市山崎 2641

Tel : 04-7124-1501

E-mail : 6323701@ed.tus.ac.jp