

# Location Powered Quotient

- 組成データ解析に基づく地域特化係数の提案 -

吉田崇紘\*・村上大輔\*\*・瀬谷創\*\*\*

## Location Powered Quotient: A Compositional Data Analysis-based Industrial Concentration Measure

Takahiro YOSHIDA\*, Daisuke MURAKAMI\*\*, Hajime SEYA\*\*\*

**Abstract:** A typical measure of industrial concentration is the Location Quotient (LQ), which is simply calculated as the regional and national ratios of employment in each industrial sector. However, its calculation focuses on a single sector and thus ignores relationships with other sectors. Therefore, we propose an alternative version of LQ based on compositional data analysis, which is commonly used and developed in geosciences. The proposed index, Location Powered Quotient (LPQ), has the following properties. (1) LPQ is derived from the powering operator in Aitchison's vector space structure, (2) LPQ considers not only specialization but also the balance of composition. We apply this LPQ to an analysis of Japanese industry data to examine how the LPQ is interpreted.

**Keywords:** 地域特化係数 (location quotient) , 産業集積 (industrial concentration) , 組成データ解析 (compositional data analysis)

### 1. はじめに

地域における産業の相対的な集積度を簡便に計測可能な地域特化係数 (Location Quotient: LQ, Florence, 1939) は, 地域科学分野において幅広く利用されてきた. LQ は, 経済活動の空間分布の分析や地域経済の強みである基盤産業の把握を行う代表的な係数のひとつである (e.g., 黒田ほか, 2008).

地域  $i$  における産業  $p$  の地域特化係数  $LQ_{i,p}$  は式 (1) で定義される:

$$LQ_{i,p} = \frac{x_{i,p}}{g_p}. \quad (1)$$

ここで,  $x_{i,p}$  は地域  $i$  における産業  $p$  の従業者数シェア率,  $g_p$  は全国における産業  $p$  の従業者数シェア率である.  $LQ_{i,p} = 1$  であれば, 地域  $i$  における産業  $p$  のシェア率は, 全国レベルと同じであることを示す.  $LQ_{i,p} > 1$  であれば, 全国平均を超えて地域  $i$  に産業  $p$  が集中していることを示し基盤産業とみなされる. 一方,  $LQ_{i,p} < 1$  であれば, 非基盤産業とみなされる.

LQ は様々な改良版が提案されている. 式 (1) の計算に必要なデータは従業者数のみのため, 比較的データが集めやすく計算も容易で, 解釈もわかりやすいといった利点がある. 一方で, 基盤産業と非基盤産業の区別に用いる閾値の設定に批判があり, この点の改良が長年行われている (Guimaraes et al., 2009). O'Donoghue and Gleave (2004) は, 基準化した LQ を提案し, 正規性の検定によって区別する方法を提案している. Tian (2013) は LQ を対数化した上で基準化を行い, bootstrap 法を用いて統計的検定を行う評価方法を提案している. この他に, LQ は経済波及効果の分析に用いられる産業連関表の交易係数の推定手法としても広く用いられており, 推定精度の向上を目的として LQ をベースとした開発が近年注目されている (e.g., Pereira-Lopez et al., 2020; Buendia et al., 2022).

本研究ではさらなる LQ の改良に向けて, 新たな課題を指摘する. 式 (1) の分子, 分母のシェア率は,

---

\* 正会員 東京大学大学院工学系研究科 (The University of Tokyo)  
〒113-8656 東京都文京区本郷 7-3-1 E-mail : yoshida.takahiro@up.t.u-tokyo.ac.jp

\*\* 正会員 統計数理研究所データ科学研究系 (The Institute of Statistical Mathematics)

\*\*\* 正会員 神戸大学大学院工学研究科 (Kobe University)

組成データ (compositional data) と呼ばれる, 各変量の総和が一定 (たとえば, 百分率であれば 100, 割合であれば 1) という定数和制約の掛かったベクトルの一変量を抜き出したものである. 産業  $p$  のシェア率の計算に含まれる産業  $p$  以外のシェア率の影響を捨象しており, 全産業のシェア率を考慮の上で産業  $p$  に着目する指標になってはいない.

そこで本研究では, 地質学や鉱山学で用いられてきた組成データに対する解析手法 (Compositional Data Analysis: CoDA, Aitchison, 1982) を援用し, 全産業のシェア率を考慮した新たな地域特化係数を提案する. CoDA は, 組成データの特性である定数和制約について, 対数比変換に基づいた方法で対処を試みる統計解析手法の枠組みである (e.g., Pawlowsky-Glahn et al., 2015). CoDA を援用することにより, 定数和制約に起因する問題に対応し, 特定の産業だけでなく産業構成全体の情報も取り入れた新たな地域特化係数を開発できる可能性がある.

以下, まず第 2 章では, CoDA の基本演算子を導入し, 新たな地域特化係数 Location Powered Quotient (LPQ) を提案する. 次に第 3 章では, 本研究で用いる産業組成データについて説明し, LPQ の分析例を従来の LQ と比較しながら考察する. 最後に第 4 章で本研究の今後の展望と課題を述べる.

## 2. Location Powered Quotient (LPQ)

以下では, まず組成データ解析の基本演算子を導入し, LQ の改良版を提示する. 変量数が  $D$ , 各変量が比率で与えられた組成データベクトル  $\mathbf{x} = (x_1, x_2, \dots, x_D)$  の標本空間 (単体空間) は式 (2) で定義することができる:

$$\mathcal{S}^D = \left\{ \mathbf{x} \mid x_p > 0, p = 1, 2, \dots, D, \sum_{p=1}^D x_p = 100 \right\}. \quad (2)$$

CoDA において,  $\mathcal{S}^D$  のベクトル空間は, 摂動演算 (perturbation operation, 式 3):

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}(x_1 \cdot y_1, \dots, x_D \cdot y_D) \in \mathcal{S}^D \quad (3)$$

と, 冪演算 (powering operation, 式 4):

$$\alpha \odot \mathbf{x} = \mathcal{C}(x_1^\alpha, \dots, x_D^\alpha) \in \mathcal{S}^D \quad (4)$$

によって定義される (Aitchison, 1986). ここで,  $\mathbf{x}, \mathbf{y} \in \mathcal{S}^D$ ,  $\alpha \in \mathbb{R}$  であり,  $\mathcal{C}(\cdot)$  は式 (5) で表されるように, 非負ベクトル  $\mathbf{z} = (z_1, z_2, \dots, z_D) \in \mathbb{R}_+^D$  の閉鎖演算 (closure operation) である:

$$\mathcal{C}(\mathbf{z}) = \left( \frac{100 \cdot z_1}{\sum_{p=1}^D z_p}, \dots, \frac{100 \cdot z_D}{\sum_{p=1}^D z_p} \right) \in \mathcal{S}^D. \quad (5)$$

摂動演算, 冪演算はそれぞれ単体空間上の足し算, 掛け算に相当する.

以上の演算子に基づき, 新たな地域特化係数  $\alpha_i \in \mathbb{R}^D$  を導入する. 地域  $i$  の産業組成を  $\mathbf{x}_i \in \mathcal{S}^D$ , 全国の産業組成を  $\mathbf{g} \in \mathcal{S}^D$  として, 次の冪演算を考える:

$$\begin{aligned} \mathbf{x}_i &= \alpha_i \odot \mathbf{g} \\ &= \mathcal{C}(g_1^{\alpha_{i,1}}, \dots, g_D^{\alpha_{i,D}}) \\ &= \left( \frac{g_1^{\alpha_{i,1}}}{g_1^{\alpha_{i,1}} + \dots + g_D^{\alpha_{i,D}}}, \dots, \frac{g_D^{\alpha_{i,D}}}{g_1^{\alpha_{i,1}} + \dots + g_D^{\alpha_{i,D}}} \right). \end{aligned} \quad (6)$$

なお, 紙幅の都合,  $\alpha_i, \mathbf{x}_i$  の添え字  $i$  を以下省略する.

式 (6) より次の連立方程式が得られる:

$$\begin{cases} x_1 = \frac{g_1^{\alpha_1}}{g_1^{\alpha_1} + \dots + g_D^{\alpha_D}} \\ \vdots \\ x_D = \frac{g_D^{\alpha_D}}{g_1^{\alpha_1} + \dots + g_D^{\alpha_D}} \end{cases} \quad (7)$$

$$\Leftrightarrow \begin{cases} \log x_1 = \alpha_1 \log g_1 - \log(g_1^{\alpha_1} + \dots + g_D^{\alpha_D}) \\ \vdots \\ \log x_D = \alpha_D \log g_D - \log(g_1^{\alpha_1} + \dots + g_D^{\alpha_D}) \end{cases}.$$

式 (7) の両辺で和をとり整理する:

$$\begin{aligned} &\log x_1 + \dots + \log x_D \\ &= (\alpha_1 \log g_1 + \dots + \alpha_D \log g_D) \\ &\quad - D \log(g_1^{\alpha_1} + \dots + g_D^{\alpha_D}) \\ \Leftrightarrow &\log(x_1 \dots x_D) = \log \frac{g_1^{\alpha_1} \dots g_D^{\alpha_D}}{(g_1^{\alpha_1} + \dots + g_D^{\alpha_D})^D} \\ \Leftrightarrow &x_1 \dots x_D = \frac{g_1^{\alpha_1} \dots g_D^{\alpha_D}}{(g_1^{\alpha_1} + \dots + g_D^{\alpha_D})^D} \\ \Leftrightarrow &g_1^{\alpha_1} + \dots + g_D^{\alpha_D} = \frac{(g_1^{\alpha_1} \dots g_D^{\alpha_D})^{\frac{1}{D}}}{(x_1 \dots x_D)^{\frac{1}{D}}}. \end{aligned} \quad (8)$$

式 (8) を式 (7) の連立方程式に代入し整理する:

$$\begin{aligned} & \begin{cases} \frac{x_1}{(x_1 \dots x_D)^{\frac{1}{D}}} = \frac{g_1^{\alpha_1}}{(g_1^{\alpha_1} \dots g_D^{\alpha_D})^{\frac{1}{D}}} \\ \vdots \\ \frac{x_D}{(x_1 \dots x_D)^{\frac{1}{D}}} = \frac{g_D^{\alpha_D}}{(g_1^{\alpha_1} \dots g_D^{\alpha_D})^{\frac{1}{D}}} \end{cases} \\ \Leftrightarrow & \begin{cases} \log \left( \frac{x_1}{(x_1 \dots x_D)^{\frac{1}{D}}} \right) = \log \left( \frac{g_1^{\alpha_1}}{(g_1^{\alpha_1} \dots g_D^{\alpha_D})^{\frac{1}{D}}} \right) \\ \vdots \\ \log \left( \frac{x_D}{(x_1 \dots x_D)^{\frac{1}{D}}} \right) = \log \left( \frac{g_D^{\alpha_D}}{(g_1^{\alpha_1} \dots g_D^{\alpha_D})^{\frac{1}{D}}} \right) \end{cases} \\ \Leftrightarrow & \begin{cases} \log x_1 - \frac{1}{D} \sum_{p=1}^D \log x_p = \alpha_1 \log g_1 - \frac{1}{D} \sum_{p=1}^D \alpha_p \log g_p \\ \vdots \\ \log x_D - \frac{1}{D} \sum_{p=1}^D \log x_p = \alpha_D \log g_D - \frac{1}{D} \sum_{p=1}^D \alpha_p \log g_p \end{cases} \quad (9) \end{aligned}$$

式 (9) の連立方程式を行列で表す：

$$\begin{aligned} & \begin{bmatrix} \log x_1 - \frac{1}{D} \sum_{p=1}^D \log x_p \\ \vdots \\ \log x_D - \frac{1}{D} \sum_{p=1}^D \log x_p \end{bmatrix} \\ & = \begin{bmatrix} \frac{D-1}{D} \log g_1 & \dots & -\frac{1}{D} \log g_D \\ \vdots & \ddots & \vdots \\ -\frac{1}{D} \log g_1 & \dots & \frac{D-1}{D} \log g_D \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_D \end{bmatrix} \\ & = \mathbf{A}\boldsymbol{\alpha}. \quad (10) \end{aligned}$$

ここで、行列  $\mathbf{A}$  は特異行列となるため、ムーア・ペンローズ形一般逆行列により疑似逆行列  $\mathbf{A}^+$  を得る。式 (11) の両辺に左から  $\mathbf{A}^+$  を掛けることで、最小二乗解  $\boldsymbol{\alpha}$  を得ることができる。冪演算から導出できることから、新たな地域特化係数  $\boldsymbol{\alpha}$  を Location Powered Quotient (LPQ) と呼ぶ。

### 3. 分析と考察

#### 3.1. 産業組成データ

本研究で用いるデータは平成 28 年経済センサス活動調査の産業別従業者数である。データは政府統計窓口 e-Stat からダウンロードした。今回の分析では、現行の日本標準産業分類（大分類）の各産業別従業者数を、地域が 47 都道府県、産業が 3 分類になるよう集計した。産業 3 分類は、「第 1 次産業」に A 農業・林業、B 漁業、C 鉱業・採石業・砂利採取業、「第 2 次産業」に D 建設業、E 製造業、F 電気・ガス・熱供給・水道業、そして「第 3 次産業」に G 情報通信業、H 運輸業・郵便業、I 卸売業・小売業、J 金融業・保険業、K 不動産業・物品賃貸業、L 学術研究・専門・技術サービス業、M 宿泊業・飲食サービス業、N 生活関連サービス業・娯楽業、O 教育・学習支援業、P 医療・福祉、Q 複合サービス事業、R サービス業(他に分類されないもの)を割り当てた。なお、S 公務(他に分類されるものを除く)、T 分類不能の産業は、e-Stat から取得可能なデータに含まれていないため、集計から除外した。地域特化係数を計算するにあたり、定数和が 100 となるよう閉鎖演算 (式 5) により組成データに変換した。図 1 に集計した産業組成データの三角図を示す。シェア率は、「第 1 次産業」が 10%以下、「第 2 次産業」が 40%以下、「第 3 次産業」が 60%以上にいずれの都道府県も位置していることが確認できる。

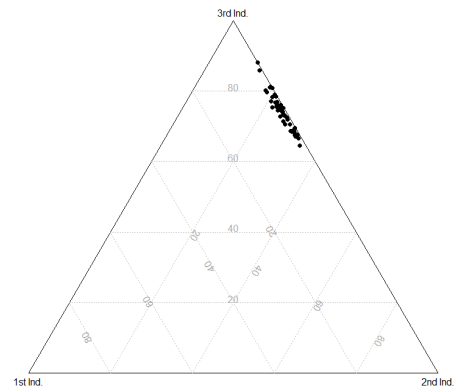


図 1. 産業組成データの分布

#### 3.2. 分析結果と考察

LQ と比較しながら提案した LPQ の解釈について

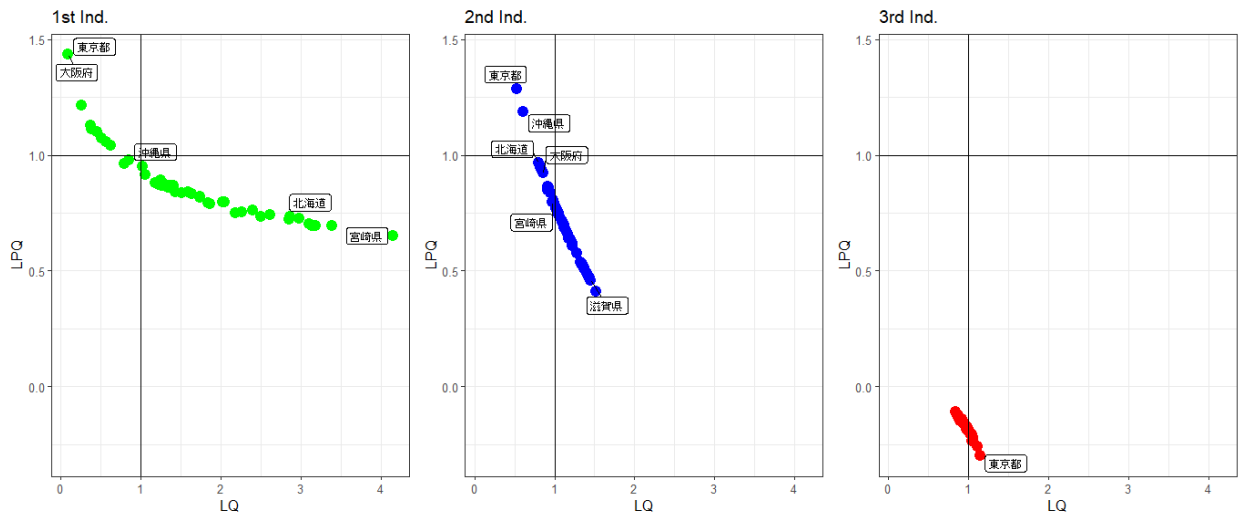


図 2 : (LQ) - (LPQ) の散布図

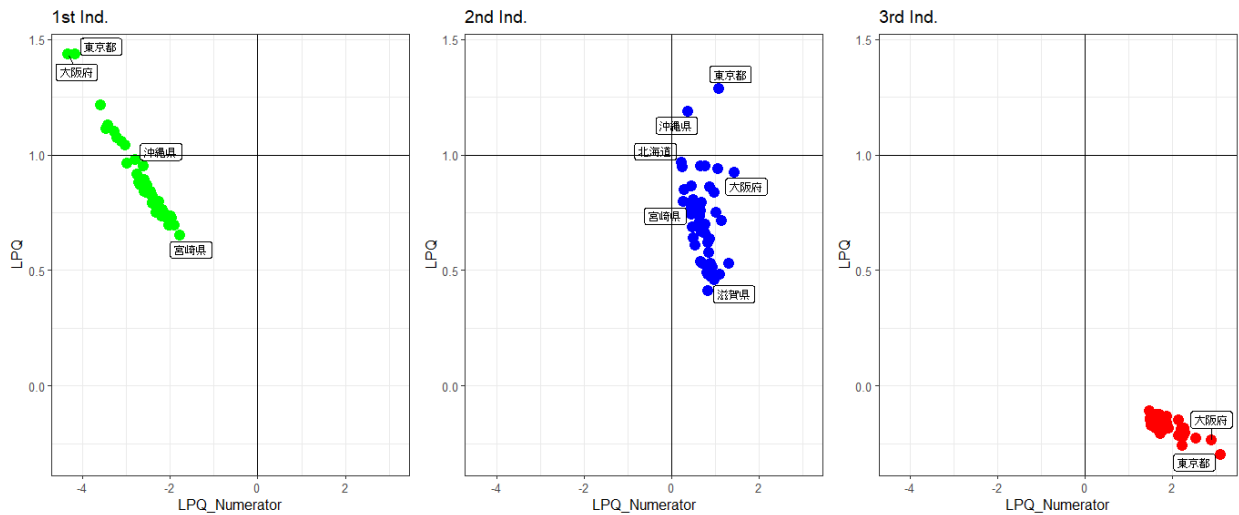


図 2 : (式 (10) の左辺) - (LPQ) の散布図

考察する．図 2 は，LQ と LPQ の散布図に両軸で 1 となる水平線と鉛直線を加えたものである．LQ は前述の通りであるが，LPQ も同様に，式 (9) あるいは (10) において  $\alpha = 1$  であるとき両辺が一致することから，両係数とも 1 を解釈上の閾値とした．LPQ は LQ ほど直観的な解釈ができないが，LQ でいうところの分子に相当するのは地域における産業構成内の偏差であり，分母に相当するのも変量数で調整された全国における産業構成内の偏差である．ここで，「第 1 次産業」のパネルにおける宮崎県は，LQ では最も大きい値であり 1 より大きく，LPQ では最も小さい値であり 1 より小さいところに位置している．

宮崎県の産業組成データは，(i) 「第 1 産業」のシェア率が他の都道府県のそれと比べて最も大きい，(ii) 他の都道府県と同様に宮崎県内における「第 1 次産業」のシェア率は最も小さい．LQ は (i) のみを考慮し，LPQ は (i) と (ii) を同時に考慮する指標である．(ii) を反映した結果，宮崎県の LPQ は値が 1 より小さくなった可能性があり，直観にも沿う．一方で，(i) を反映するのであれば，「第 1 産業」のシェア率が他の都道府県のそれと比べて最も小さい大阪府や東京都の LPQ は，宮崎県より小さな値をとるのが自然であり，解釈には課題が残る．

図 3 は，上記の課題の検証として符号が LPQ の解

に依らない式 (10) の左辺と LPQ の値の散布図である。式 (10) の左辺は地域の産業構成内の偏差とみることができることから閾値を 0 とし、LPQ は同様に閾値を 1 とした。各産業のパネルより、式 (10) の左辺の符号の分布は、「第 1 次産業」は負、「第 2 次産業」は 0 付近、「第 1 次産業」は正であり直観に整合的である。一方で LPQ の順序はいずれも直観に反しており解釈には議論の余地がある。

#### 4. おわりに

本研究では組成データ解析手法に基づく演算子を応用し、特定の産業だけでなく産業構成全体を考慮に入れた地域特化係数を提案した。提案した係数を日本の産業分類データに応用し、その地域特化の程度の解釈を検討した。本研究と同様のアプローチは、Local Moran'  $I$  統計量 (Anselin, 1995) や  $H$  統計量 (Ord and Getis, 2012) といった空間的自己相関、空間的異質性を測る一変量データを入力としている空間統計量を定数と制約付き多変量データ用に拡張する際にも役立つ可能性がある。

今後の展開には、まずは LPQ の解釈についてさらに検討することが挙げられる。その後、空間統計量への拡張に取り組みたい。また、組成データ解析手法の大きな課題である 0 値の対応についても、Andreano (2019) などの最近の検討を参考にしながら、分類数を増やした場合にも柔軟に適用可能で(分類を細かくするほど 0 値が生じやすい)、より詳細に地域特化の程度を測る方法について検討したい。さらに、従来の LQ の改良として長年検討されてきた統計的検定による評価方法を提案手法に採用することを検討したい。

#### 謝辞

本研究は JSPS 科研費 17K18554, 18H03628, 21K13153 の助成を受けたものである。

#### 参考文献

黒田達朗・田淵隆俊・中村良平 (2008) 『都市と地域の経済学 (新版)』, 有斐閣。

Aitchison, J. (1982) The statistical analysis of

compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, **44** (2), 139–160.

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data*. Chapman & Hall, London, UK.

Andreano, M.S., Benedetti, R., Piersimoni, F. (2019) A distance correlation index of spatial dependence for compositional data. *Papers in Regional Science*, **98** (6), 2371–2389.

Anselin, L. (1995) Local Indicators of Spatial Association—LISA. *Geographical Analysis*, **27** (2), 93–115.

Buendía Azorín, J. D., Martínez Alpañez, R., Sanchez de la Vega, M. D. M. (2022) A new proposal to model regional input–output structures using location quotients. An application to Korean and Spanish regions. *Papers in Regional Science*, in press.

Florence, P.S. (1939) *Report on the Location of Industry*. Political and Economic Planning, London, UK.

Guimarães, P., Figueiredo, O., Woodward, D. (2009) Dartboard tests for the location quotient. *Regional Science and Urban Economics*, **39** (3), 360–364.

O'Donoghue, D., Gleave, B. (2004) A note on methods for measuring industrial agglomeration. *Regional Studies*, **38** (4), 419–427.

Ord, J.K., Getis, A. (2012) Local spatial heteroscedasticity (LOSH). *The Annals of Regional Science*, **48** (2), 529–539.

Pawlowsky-Glahn, V., Egozcue, J.J., and Tolosana-Delgado, R. (2015) *Modelling and Analysis of Compositional Data*. John Wiley & Sons, Chichester, UK.

Pereira-Lopez, X., Carrascal-Incera, A., Fernandez-Fernandez, M. (2020) A bidimensional reformulation of location quotients for generating input-output tables. *Spatial Economic Analysis*, **15** (4), 476–493.

Tian, Z. (2013) Measuring agglomeration using the standardized location quotient with a bootstrap method. *Journal of Regional Analysis & Policy*, **43** (2), 186–197.