1 因子情報路モデルからの 隔たりを測る尺度

東京理科大学理工学部 富澤貞男・寒河江雅彦・国沢清典

A Measure of Departure from One-Factor Information Channel Model

Sadao TOMIZAWA, Masahiko SAGAE and Kiyonori KUNISAWA

|||||||||||||||||||| 要 旨 |||

国沢 (1959, 1975) は, 社会現象における大衆の選択行動を説明するのに, 1 次元確率分布のモデルとして 1 因子情報路モデルを導入した. 本稿は, 1 因子情報路モデルからの隔たりの程度を測るために Kullback-Leibler 情報量に基づいた簡単な尺度を導入する. この尺度は, 1 因子情報路モデルからの隔たりの程度が, 最大のときに比べてどれくらいであるかという相対比を示しており, いくつかのデータにおいてその隔たりの程度を比較するのに役立つ. この尺度の近似的信頼区間を求め, 例を示す.

||||||||||||||||||||

1. はじめに

社会現象における大衆の選択行動を説明するのに, 国沢 (1959, 1975) は, 1 因子情報路モデルを導入した. そのモデルを以下説明しよう. たとえば, 1 つの商品について r 個の銘柄 $A_1, A_2, \dots, A_r$ が販売されているものとする. 銘柄 $A_1, A_2, \dots, A_r$ の既知の販売価格を $0 < u_1 \leq u_2 \leq \dots \leq u_r$ とする. 購買者は, この販売価格以外の情報をもっていないと仮定する. 購買者が銘柄 $A_k (k=1, 2, \dots, r)$ を選択する確率を p_k とする. ここに, $\sum p_k = 1$ である. 1 因子情報路モデルとは p_k が次のように定まるといふ仮説である:

$$H_0: p_k = p_k^*, k=1, 2, \dots, r, \quad (1)$$

ただし, $p^* = (p_1^*, \dots, p_r^*)$ は, 与えられた $u = (u_1, \dots, u_r)$ にたいして

(I) 平均販売価格 $\bar{u} = u_1 p_1 + \dots + u_r p_r$ を最小,

(II) エントロピー $H(p) = -p_1 \log p_1 - \dots - p_r \log p_r$ を最大,

を満たす $p = (p_1, \dots, p_r)$ の値である. この $\{p_k^*\}$ は $\{u_k\}$ より具体的に次のように求まる. すなわ

Key words: Kullback-Leibler information, Extremal entropy, Extremal mean.

ち,

$$p_k^* = w_0^{-u_k}, k=1, 2, \dots, r, \quad (2)$$

ただし, w_0 は方程式

$$w^{-u_1} + w^{-u_2} + \dots + w^{-u_r} = 1$$

の正根である (正根は, つねに存在して 1 つに限る). [詳細は, 国沢 (1959, 1975) を参照せよ. なお, いろいろな正の整数 $\{u_k\}$ に対して計算された w_0 の値が, たとえば, 国沢 (1975, 付表) に掲載されている.] 条件 (I) から, 販売価格 $\{u_k\}$ の比, $u_1 : u_2 : \dots : u_r$, が一定ならば, 1 因子情報路モデルは不変である. したがって, 一般性を失うことなく販売価格は簡単な正の整数としてよい.

本稿の目的は, 1 因子情報路モデルからの隔たりの程度を表す尺度として次節の Ψ を導入することである.

2. 1 因子情報路モデルからの隔たりの程度を測る尺度

1 因子情報路モデル H_0 からのある p の隔たりの程度を測るために次のような尺度を導入する:

$$\begin{aligned} \Psi &= \Psi(u_1, \dots, u_r; p_1, \dots, p_r) \\ &= \frac{1}{u_r \log w_0} [(\log w_0) \bar{u} - H(p)] \\ &= \frac{1}{u_r \log w_0} [(\log w_0) \sum_{k=1}^r u_k p_k - (-\sum_{k=1}^r p_k \log p_k)] \\ &= \frac{1}{u_r \log w_0} D(p, p^*), \end{aligned}$$

ただし

$$D(p, p^*) = \sum_{k=1}^r p_k \log (p_k / p_k^*).$$

この尺度は, 確率分布 $\{p_k\}$ と 1 因子情報路モデルを満たす構造をもつ確率分布 $\{p_k^*\}$ との距離を (上限の値で) 修正した Kullback-Leibler 情報量で測っている. Ψ の取り得る範囲は, $0 \leq \Psi \leq 1$ であって,

- (i) $\Psi=0$ であるための必要十分条件は, 1 因子情報路モデルが成り立つこと, すなわち, $\{p_k\}$ が (I) と (II) を満たすときであり,
- (ii) $\Psi=1$ であるための必要十分条件は, 与えられた既知の販売価格が $u_1 \leq u_2 \leq \dots \leq u_{s-1} < u_s = u_{s+1} = \dots = u_r$ であって $p_s, p_{s+1}, \dots, p_r$ のいずれかが 1 のときである (付録参照せよ). すなわち, $\{p_k\}$ が

(III) 平均販売価格 $\bar{u}$ を最大,

(IV) エントロピー $H(p)$ を最小,

を満たすときである (証明略).

尺度 Ψ の値が増加するにつれて、1 因子情報路モデルからの隔たりの程度は、(Kullback-Leibler 情報量の意味で) 増加する。ここに、 $\Psi=1$ のとき、すなわち、 $\{p_k\}$ が (III) と (IV) を満たすとき、1 因子情報路モデルからの隔たりの程度は最大とみなされる [これは (I) と (II) と全く逆の場合であるから自然である]。なお、販売価格 $\{u_k\}$ の比、 $u_1:u_2:\cdots:u_r$ 、が一定ならば、 Ψ も一定となることを注意しておく。

いま、 n 人の購買者のうち、銘柄 $A_k(k=1, 2, \dots, r)$ を選択した人数を n_k とする。ただし、 $\sum n_k = n$ であり、各購買者は r 銘柄のうち 1 つしか選択できないとする。観測値 $\{n_k\}$ は一つの多項分布に従うものと仮定する。Delta method を使って尺度 Ψ の近似の信頼区間を考えよう。[Delta method の説明は、たとえば、Bishop et al. (1975, Sec. 14.6) を見よ。] 推定された尺度 $\hat{\Psi}$ は、 $\{p_k\}$ を $\{\hat{p}_k\}$ で置き換えた Ψ により与えられる。ここに $\hat{p}_k = n_k/n$ である。Delta method を使って、 $\sqrt{n}(\hat{\Psi} - \Psi)$ は漸近的に ($n \rightarrow \infty$ として) 平均 0 そして分散 σ^2 の正規分布に従う。ここに、

$$\sigma^2 = \frac{1}{(u_r \log w_0)^2} \sum_{k=1}^r p_k (\log p_k + u_k \log w_0)^2 - \Psi^2$$

である。(なお、 $\Psi=0$ と $\Psi=1$ のときは、つねに $\sigma^2=0$ となり、この漸近分布は適用できないことに注意する。) いま、 $\{p_k\}$ を $\{\hat{p}_k\}$ によって置き換えた σ^2 を $\hat{\sigma}^2$ で表す。このとき $\hat{\sigma}/\sqrt{n}$ は $\hat{\Psi}$ に対する一つの推定された近似の標準誤差であり、そして $\hat{\Psi} \pm z_{\alpha/2} \hat{\sigma}/\sqrt{n}$ は Ψ に対する近似の $100(1-\alpha)\%$ の信頼区間である。ただし $z_{\alpha/2}$ は、標準正規分布の両側の裾の確率が α の確率点である。

3. 例

表 1 は、中部電力 OR グループ (国沢, 1975, pp. 57-59) によるもので、1960 年度における日本の 4 つの電力会社 (中部電力、北陸電力、北海道電力、東京電力) における家庭用電灯ランク別 [5A (アンペア), 10 A, 20 A, 30 A] の契約数を表している。基本料金の比はアンペア数の比と同じであり、ここに、 $u_1=1, u_2=2, u_3=4, u_4=6$ とする。このとき、 $w_0=1.7885$ である [詳細は国沢 (1975) を見よ]。

各電力会社のデータに適用された $\hat{\Psi}$ そして Ψ の 95% 信頼区間 (表 2) より、すべての電力会社において、1 因子情報路モデルからの隔たりの程度は、それが最大のときに比べて、1/100 以下である。(とりわけ、東京電力においては、1/1000 以下である。) 加えて、1 因子情報路モデルからの隔たりの程度を、電力会社間で比較すると、(i) 中部電力が 1 番大きく、(ii) 北陸電力が 2 番目に大きく、(iii) 北海道電力が 3 番目に大きく、そして (iv) 東京電力が 1 番小さい。したがって、これら 4 つの電力会社のうちで東京電力における契約世帯が契約に際して

(I) 基本料金支出をできるだけ少なくし、

(II) 一方、料金にとらわれずに自由な選択をとろうとする (したがって、エントロピーをできるだけ大きくする)、

という志向がもっとも強く働いていると考えられる。

表 1. 1960 年度の日本の各電力会社における家庭用電灯ランク別の契約数.
中部電力 OR グループ (国沢, 1975, pp. 57-59) より引用.

中部電力					
アンペア	5	10	20	30	計
戸 数	998,650	743,265	214,659	58,695	2,015,269
比 率	49.55	36.88	10.65	2.91	100
北陸電力					
アンペア	5	10	20	30	計
戸 数	306,313	153,919	39,223	10,222	509,677
比 率	60.10	30.20	7.70	2.01	100
北海道電力					
アンペア	5	10	20	30	計
戸 数	349,080	187,649	74,330	13,045	624,104
比 率	55.93	30.07	11.91	2.09	100
東京電力					
アンペア	5	10	20	30	計
戸 数	2,277,606	1,471,377	411,391	114,758	4,275,132
比 率	53.28	34.42	9.62	2.68	100

表 2. 表 1 のデータに適用された Ψ の推定値, $\hat{\Psi}$ に対する推定された近似の標準誤差, そして Ψ の近似の 95% 信頼区間

適用されたデータ	隔たり尺度の推定値 $\hat{\Psi}$	標準誤差 $\hat{\sigma}/\sqrt{n}$	信頼区間 $\hat{\Psi} \pm z_{0.025} \hat{\sigma}/\sqrt{n}$
中部電力	0.0026	0.00003	(0.0025, 0.0027)
北陸電力	0.0018	0.00004	(0.0017, 0.0019)
北海道電力	0.0012	0.00003	(0.0011, 0.0013)
東京電力	0.0007	0.00001	(0.0007, 0.0007)

4. おわりに

本稿で導入した尺度 Ψ は, 第 2 節で述べたように, Kullback-Leibler 情報量 $D(p, p^*)$ をその上限で修正し, 区間 $[0, 1]$ に入るように基準化したものである. したがって, Ψ は, 1 因子情報路モデルからの隔たりの相対的大きさを知りたいときに役立つ. また, 第 3 節のように, いくつかのデータにおいてその隔たりの程度を比較するのに役立つ.

1 因子情報路モデルの適合度を検定するための尤度比カイ 2 乗統計量 G^2 は, $\hat{\Psi}$ を使って, $G^2 = 2n(u_r \log w_0) \hat{\Psi}$ と表せる. したがって, G^2 の取り得る範囲は, $0 \leq G^2 \leq 2n(u_r \log w_0)$ である. それゆえ, G^2 は (表 1 のデータのように) 観測値 n の大きさの異なるいくつかのデータにおいて, 1 因子情報路モデルからの隔たりの程度を比較するのにそのまま用いることはできないことを注

意しておく。(ちなみに, 表1の各電力会社別データに対する G^2 の値を大きい方から小さい方へ並べると中部, 東京, 北陸そして北海道の順になり, 第3節(表2)で述べた $\hat{\Psi}$ の大きさの順序とは一致しない.)

付 録

第2節における $0 \leq \Psi \leq 1$ と (i), (ii) の証明を述べる. 与えられた既知の販売価格を $u_1 \leq u_2 \leq \dots \leq u_{s-1} < u_s = u_{s+1} = \dots = u_r$ として, $w_0^{u_r} > 1$ とそして $D(p, p^*)$ と $H(p)$ が共に非負であることを使えば

$$\begin{aligned} 0 \leq \Psi &= \frac{1}{u_r \log w_0} D(p, p^*) = \frac{1}{u_r \log w_0} [(\log w_0) \bar{u} - H(p)] \\ &\leq \frac{1}{u_r \log w_0} [(\log w_0) \bar{u}] \leq \frac{1}{u_r} (u_r \sum p_k) = 1 \end{aligned}$$

が成り立つ. また, 等号条件を考えれば, (i) と (ii) が成り立つことは明白である.

参 考 文 献

- Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975): *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: MIT Press.
 国沢清典 (1959): OR のための情報の理論入門. 日科技連出版.
 国沢清典 (1975): エントロピー・モデル. 日科技連出版.

(1990年12月25日受付 1991年1月17日最終修正)

著者連絡先: 〒278 千葉県野田市山崎東亀山2641 東京理科大学理工学部情報科学科