情報量と不平等尺度

TOYODA, Takashi / 豊田, 敬

(出版者 / Publisher)
法政大学経営学会
(雑誌名 / Journal or Publication Title)
The Hosei journal of business / 経営志林
(巻 / Volume)
32
(号 / Number)
4
(開始ページ / Start Page)
197
(終了ページ / End Page)
201
(発行年 / Year)
1996-01-30
(URL)
https://doi.org/10.15002/00003419

〔研究ノート〕

情報量と不平等尺度

豊田　敬

0．エントロピーの概念を不平等尺度に応用するというアイデアは，Theil（1967）によつて導入された。Theilのこの不平等尺度は，各グループ内不平等度の加重平均とグループ間不平等度との和に分解できるという著しい特徴をもつ。その後，1970年代に不平等比較とSchur-凸性との関連が明らかになり，Bourguignon（1979），Cowell and Kuga（1981），Shorrocks（1980），Toyoda（1980）等によって，Schur-凸性と分解可能性とを兼ね備えた不平等尺度のクラスとしての一般化エントロピークラス（generalized entropy class）が提示された。[なお，一般化エントロピークラスという名称はCowell and Kugaによる。]

ところが，これらの研究はいずれも，分解可能性とSchur-凸性を基本条件として展開しているため，一般化エントロピークラスと名付けられてはいるものの，本来のエントロピーや情報量との関係が希薄になっているという憾みがある。

本稿の目的は，情報量と一般化エントロピークラスとの関係をより明確にすることにある。具体的に言えばつぎの2点である。第一点は，Theilの不平等尺度はKullback and Leibler（1951）の情報量に他ならないという事実を確認することである。第二点は，一般化エントロピークラスはCsiszār（1967）のf-ダイヴァージェンスのサブクラスであるという事実を指摘することである。[Csiszār, I.（1967），Information-type measures of difference in probability distributions and indirect observations, *Studia Sci. Math. Hungar.* 2，299−318．は入手できていない。そのため，引用文献には彼自身のサーベイ的論文Csiszār（1977）を挙げてある。]

1．記号の導入を兼ねながら，まずエントロピーについて触れておこう。

完全事象系の確率分布を

$$\boldsymbol{p} = (p_1,\ p_2,\ \cdots,\ p_N)$$

とする。

$$p_i \geqq 0,\ i = 1,\ \cdots,\ N, \qquad \sum_{i=1}^{N} p_i = 1$$

であることは言うまでもない。このとき，$\boldsymbol{p}$のエントロピーは，平均情報量

$$H(\boldsymbol{p}) = -\sum_{i=1}^{N} p_i \log p_i$$

で定義される。ただし，$p_i = 0$ に対しては

$$x \log x \to 0,\ (x \to +0)$$

という事実に従って，$0 \cdot \log \cdot 0 = 0$ と定める。以下，本稿では対数のlogは自然対数とする。

定義からただちに分かるとおり，$H(\boldsymbol{p})$ は，

$$Pr(\pi = p_i) = p_i,\ i = 1,\ \cdots,\ N,$$

なる確率変数πの凸関数 $-\log \pi$ の期待値にほかならない。

エントロピー$H(\boldsymbol{p})$ は，事象系の「不確かさ」を測る尺度であって，非負の数値をとる。特に，事象系に「不確かさ」がない $\boldsymbol{p} = (0,\ \cdots,\ 0,\ 1)$ のとき，最小値となる：

$$\min_{\boldsymbol{p}}\ H(\boldsymbol{p}) = H(0,\ \cdots,\ 0,\ 1) = 0。$$

他方，事象系の「不確かさ」が最も大きい一様分布 $\boldsymbol{p} = (1/N,\ 1/N,\ \cdots,\ 1/N)$ のとき，最大値 $\log N$ をとる：

$$\begin{aligned}\max_{\boldsymbol{p}}\ H(\boldsymbol{p}) &= H(1/N,\ 1/N,\ \cdots,\ 1/N) \\ &= -\sum_{i=1}^{N} (1/N)\log(1/N) \\ &= \log N。\end{aligned}$$

なお，$H(\boldsymbol{p})$ は$\boldsymbol{p}$の各要素に関して対称である。つまり，$(1,\ 2,\ \cdots,\ N)$ の任意の順列 $(i_1,\ i_2,\ \cdots,\ i_N)$ に対して，

$$H(p_1, p_2, \cdots, p_N) = H(p_{i1}, p_{i2}, \cdots, p_{iN})$$

が成り立つという明白な事実もここで記しておく。

2．さて，Theilの不平等尺度について概説しよう。経済主体あるいはより一般な主体がN個からなる社会Sを想定し，各主体に帰属するなんらかの量

$$x_i,\ i = 1, \cdots, N,$$

の格差，不平等度あるいは集中度と称される，いささかあいまいな概念を測定することを考える。ただし，ここで対象とする量自体はあいまいではなく，共通の測定単位でもって非負の実数で表現され，加減乗除の演算が一定の意味を有し，そして総和が零でないものとする：

$$x_i \geqq 0,\ i = 1, \cdots, N, \quad \sum_{i=1}^{N} x_i \neq 0\,。$$

以下，言い回しを簡単にするために，各主体を「個人」と称し，対象とする量x_i，$i = 1$，…，N，を「所得」と称することにする。もちろん，ここでいう「個人」は，世帯，企業，事業所等いずれであっても構わない。「所得」の方も，「個人」の内容に対応して資産，貯蓄残高，資本金，出荷額などさまざまでありうる。また，格差や集中度という言葉をすべて「不平等度」あるいは「平等度」と称する。

Theilの提案になるエントロピー$H(\boldsymbol{p})$を用いた所得不平等尺度は，つぎのような考えにもとづいて導出される。

いま，N人の所得をx_i，$i = 1$，…，N，とする。このとき，全所得に占める各個人の所得シェアは

$$p_k = \frac{x_k}{\Sigma x_i},\ k = 1, \cdots, N,$$

であり，所得シェアベクトル
$\boldsymbol{p} = (p_1, p_2, \cdots, p_N)$は分布の条件

$$p_i \geqq 0,\ i = 1, \cdots, N, \qquad \Sigma p_i = 1$$

を満たす。

所得シェアベクトル$\boldsymbol{p}$において，

$$p_i = 1/N,\ i = 1, \cdots, N,$$

のときは，各個人の所得がすべて等しい「完全平等」の場合に相当する。一方，ある一つの添字iに対して$p_i = 1$，この添字i以外の他のすべての添字$j (j \neq i)$に対して$p_j = 0$のときは，その第i番目の1個人が全所得を独占し，他のどの個人の所得もすべて零という「完全不平等」の場合に相当する。

したがって，前者の「完全平等」の場合に，エントロピー$H(\boldsymbol{p})$は最大値$\log N$となり，後者の「完全不平等」の場合に，エントロピー$H(\boldsymbol{p})$は最小値0となる。つまり，所得シェアベクトル$\boldsymbol{p}$にエントロピー$H(\boldsymbol{p})$を適用すると，$H(\boldsymbol{p})$は「平等度」を測る尺度とみることができるわけである。

ところで，平等尺度$H(\boldsymbol{p})$を，非負値の不平等尺度に転換するには，その最大値$\max_{p} H(\boldsymbol{p}) = \log N$から$H(\boldsymbol{p})$を引算すればよい。こうして，次式であらわされるTheilの不平等尺度が得られる。

$$\begin{aligned} T(\boldsymbol{p}) &= \log N - H(\boldsymbol{p}) \\ &= \log N + \sum_{i=1}^{N} p_i \log p_i \\ &= \sum_{i=1}^{N} p_i (\log N + \log p_i) \\ &= \sum_{i=1}^{N} p_i \log N p_i \\ &= \sum_{i=1}^{N} p_i \log \frac{p_i}{1/N} \end{aligned} \tag{1}$$

3．社会Sがいくつかのグループから構成されるときのTheilの尺度の表現を示すことで，それがKullback-Leiblerの情報量と同一であることを確認しよう。

いま，社会Sが，互いに素なn個のグループS_1，S_2，…，S_nから構成され，各グループには，それぞれN_1人，N_2人，…，N_n人の個人が所属しているとする：

$$S = S_1 \cup S_2 \cup \cdots \cup S_n,\ S_i \cap S_j = \emptyset,\ i \neq j$$
$$N = N_1 + N_2 + \cdots + N_n\,。$$

社会SにおいてグループS_iの占める所得シェアをp_i，$i = 1$，…，nであらわす。S_iにはN_i人が所属しているのであるから，各個人1人当りの所得シェアはp_i/N_iである。

したがって，個人ベースの所得シェアベクトルは

$$p_B = \Big(\underbrace{\frac{p_1}{N_1}, \cdots, \frac{p_1}{N_1}}_{N_1}, \underbrace{\frac{p_2}{N_2}, \cdots, \frac{p_2}{N_2}}_{N_2}, \cdots\cdots, \underbrace{\frac{p_n}{N_n}, \cdots, \frac{p_n}{N_n}}_{N_n}\Big)$$

とあらわされる。もちろん，$\boldsymbol{p}_B$ も分布の条件を満たしており，これにTheilの尺度を適用すれば

$$\begin{aligned} T(\boldsymbol{p}_B) &= \sum_{i=1}^{n} N_i \cdot \frac{p_i}{N_i} \log N \cdot \frac{p_i}{N_i} \\ &= \sum_{i=1}^{n} p_i \log \frac{p_i}{N_i/N} \end{aligned} \qquad (2)$$

となる。この $T(\boldsymbol{p}_B)$ はグループ間の所得不平等度を計測したものである。特に，$N_1 = N_2 = \cdots = N_n = 1$ の場合，$n = N$ であって，グループ内の不平等度は消滅し，（1）式に帰着する。

ここで改めて，グループベースの所得シェアベクトル（所得シェア分布）を $\boldsymbol{p} = (p_1, p_2, \cdots, p_n)$ であらわす。また，

$$q_i = N_i/N,\ i = 1, \cdots, n,$$
$$\boldsymbol{q} = (q_1, q_2, \cdots, q_n)$$

とする。$\boldsymbol{q}$ は，グループベースの人的シェアベクトル（人的シェア分布）である。ここでも分布の条件

$$p_i \geqq 0,\ i = 1, \cdots, n, \qquad \Sigma p_i = 1$$
$$q_i \geqq 0,\ i = 1, \cdots, n, \qquad \Sigma q_i = 1$$

は満たされている。さらに，

$$q_i = 0 \Rightarrow p_i = 0$$

である。

このように記号を定めると，Theil の尺度の（2）式は

$$\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} = \sum_{i=1}^{n} p_i \left(-\log \frac{q_i}{p_i}\right) \qquad (3)$$

と書きあらわすことができる。ただし，

$$0 \cdot \log(0/q_i) = 0, \quad 0 \cdot \log(0/0) = 0$$

と定める。

（3）式は Kullback-Leibler の情報量あるいはI-ダイヴァージェンス（I-divergence）と呼ばれるものに他ならない。以下これを，KL 情報量と略称して，$I_0(\boldsymbol{p}|\boldsymbol{q})$ と表記する：

$$I_0(\boldsymbol{p}|\boldsymbol{q}) = \sum_{i=1}^{n} p_i\,(-\log(q_i/p_i)) \qquad (4)$$

添字の 0 の意味については後で触れる。

KL 情報量 $I_0(\boldsymbol{p}|\boldsymbol{q})$ を不平等尺度に引き寄せて言えば，所得シェア p_i と人的シェア q_i との乖離を凸関数 $-\log(q_i/p_i)$ で測って，それを所得シェア p_i で加重平均したものである。この意味で $I_0(\boldsymbol{p}|\boldsymbol{q})$ は，所得シェア分布 $\boldsymbol{p}$ と人的シェア分布 $\boldsymbol{q}$ との乖離あるいは逸脱をあらわす量と解される。

$\boldsymbol{p} \neq \boldsymbol{q}$ であれば，

$$I_0(\boldsymbol{p}|\boldsymbol{q}) \neq I_0(\boldsymbol{q}|\boldsymbol{p})$$

であるから，KL 情報量は，$\boldsymbol{p}$，$\boldsymbol{q}$ に関して対称でない。$\boldsymbol{p} = \boldsymbol{q}$ ならば，

$$I_0(\boldsymbol{p}|\boldsymbol{q}) = I_0(\boldsymbol{q}|\boldsymbol{p}) = 0$$

である。$\boldsymbol{p} = \boldsymbol{q}$ は，すべてのグループで所得シェアと人的シェアが等しい場合，すなわち

$$p_i = q_i,\ i = 1, \cdots, n,$$

の場合であって，完全平等の場合に該当する。

KL 情報量 $I_0(\boldsymbol{p}|\boldsymbol{q})$ はまた，人的シェア分布 $\boldsymbol{q} = (q_1, q_2, \cdots, q_n)$ に関する期待値として，

$$I_0(\boldsymbol{p}|\boldsymbol{q}) = \sum_{i=1}^{n} q_i \cdot (p_i/q_i) \log(p_i/q_i) \qquad (4)'$$

と書きあらわすこともできる。この（4）'式は，人的シェア q_i と所得シェア p_i との乖離を凸関数 $(p_i/q_i)\log(p_i/q_i)$ で測って，それを人的シェア q_i で加重平均したものである。

4．（4）式の KL 情報量を一般化したタイプの情報量として，$\mathbb{R}^+ = (0, \infty)$ で定義された凸関数 f を用いた

$$I_f(\boldsymbol{p}|\boldsymbol{q}) = \sum_{i=1}^{n} p_i f(q_i/p_i) \qquad (5)$$

なるものが考えられる。これは，Csiszār と Ali and Silvey による f-ダイヴァージェンスと呼ばれているものである。特に，$f(t) = -\log t$ とおけば，（5）式は KL 情報量の（4）式に帰着する。

なお，f の定義域にない $f(0)$ 等についてはつぎのように定める：

$$f(0) = \lim_{t\to -0} f(t), \qquad 0\cdot f(0/0) = 0,$$

$$0\cdot f(q/0) = q \lim_{t\to +0} t\cdot f(1/t)。$$

ここで，f を

$$f_b(t) = \frac{1}{b}(t^{-b}-1),\ b > -1,\ b \neq 0$$

とおく。b はパラメータである。$b=0$ に対しては，

$$f_0(t) = -\log t$$

とする。これは，l'Hopital のルールに従うと

$$\lim_{b\to 0} f_b(t) = \lim_{b\to 0} \frac{-t^{-b}\log t}{1} = -\log t$$

となることにもとづく。

この関数のクラス $f_b(t)(b > -1)$ は，t の単調減少な凸関数で，$f_b(1)=0$ であり，$b=0$ のときが KL 情報量に対応する。したがって，$f_b(t)$ を用いた時の（5）式は，f-ダイヴァージェンスのサブクラスであって，KL 情報量を含む。

$f_b(t)$ を用いたときの（5）式を $I_b(\boldsymbol{p}|\boldsymbol{q})$ であらわすことにする。$b=0$ のときは $I_0(\boldsymbol{p}|\boldsymbol{q})$ となるが，これは KL 情報量そのものであり，（4）式左辺の添字 0 とも表記が符合する。

$b > -1$，$b \neq 0$ に対して，$I_b(\boldsymbol{p}|\boldsymbol{q})$ はつぎのように書きあらわされる：

$$I_b(\boldsymbol{p}|\boldsymbol{q}) = (1/b)\sum_{i=1}^{n} p_i[(q_i/p_i)^{-b}-1] \qquad (6)$$
$$= (1/b)\sum_{i=1}^{n}(p_i^{1-b}q_i^{-b}-1)$$
$$= (1/b)\sum_{i=1}^{n} q_i[(p_i/q_i)^{1+b}-1] \qquad (6)'$$

$b=0$ のときは，（4）式，（4）' 式と同じで，

$$I_0(\boldsymbol{p}|\boldsymbol{q}) = \sum_{i=1}^{n} p_i(-\log(q_i/p_i)) \qquad (4)$$
$$= \sum_{i=1}^{n} q_i\cdot(p_i/q_i)\log(p_i/q_i) \qquad (4)'$$

である。

つぎに，$I_b(\boldsymbol{q}|\boldsymbol{p})$ を書き下すと，$b > -1$，$b \neq 0$ に対して，

$$I_b(\boldsymbol{q}|\boldsymbol{p}) = (1/b)\sum_{i=1}^{n} q_i[(p_i/q_i)^{-b}-1] \qquad (7)$$

$b=0$ のときは，

$$I_0(\boldsymbol{q}|\boldsymbol{p}) = \sum_{i=1}^{n} q_i(-\log(p_i/q_i)) \qquad (8)$$

である。

人的シェア分布 $\boldsymbol{q}$ に関する期待値の型に着目してパラメータを調整整理しよう。（6）' 式，（4）' 式において $a = 1+b$ とおけば，$a > 0$，$a \neq 1$ に対して

$$I_a(\boldsymbol{p}|\boldsymbol{q}) = \frac{1}{a-1}\sum_{i=1}^{n} q_i[(p_i/q_i)^a-1],$$

$a=1$ に対して

$$I_1(\boldsymbol{p}|\boldsymbol{q}) = \sum_{i=1}^{n} q_i\cdot(p_i/q_i)\log(p_i/q_i),$$

となる。また，（7）式，（8）式において $a = -b$ とおけば，$a < 1$ に対して

$$I_a(\boldsymbol{q}|\boldsymbol{p}) = \frac{1}{-a}\sum_{i=1}^{n} q_i[(p_i/q_i)^a-1],$$

$a=0$ に対して

$$I_0(\boldsymbol{q}|\boldsymbol{p}) = \sum_{i=1}^{n} q_i(-\log(p_i/q_i)),$$

となる。最後に，関数の凸性に留意して以上の 4 つの式をまとめると，次の一般化エントロピークラスが得られる：

$$I_a = \frac{1}{a(a-1)}\sum_{i=1}^{n} q_i[(p_i/q_i)^a-1];\ a \neq 1, 0,$$
$$I_1 = \sum_{i=1}^{n} q_i\cdot(p_i/q_i)\log(p_i/q_i); \qquad a = 1,$$
$$I_0 = \sum_{i=1}^{n} q_i(-\log(p_i/q_i)); \qquad a = 0。$$

特に，$n = N$ の場合，

$$q_i = 1/N,\ i = 1, \cdots, N,$$
$$y_i = (p_i/q_i) = (x_i/\bar{x}),\ i = 1, \cdots, N,$$

ただし，$\bar{x} = (1/N)\sum_{i=1}^{N} x_i$，

であるから，

$$I_a = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{a(a-1)}(y_i^a-1); \qquad a \neq 1, 0,$$
$$I_1 = \frac{1}{N}\sum_{i=1}^{N} y_i\log y_i; \qquad a = 1,$$
$$I_0 = \frac{1}{N}\sum_{i=1}^{N}(-\log y_i); \qquad a = 0,$$

とあらわされ，相対所得 y_i，$i=1$，…，N，($\Sigma y_i = N$) の凸関数の平均となる。

［引用文献］

Ali, S. M. and S. D. Silvey (1966), A general class of coefficients of divergence of one distribution from another, *J. R. Statist. Soc. B*, 28, 131－142.

Bourguignon, F. (1979), Decomposable income inequality measures, *Econometrica*, 47, 901－920.

Csiszâr, I. (1977), Information measures : a critical survey, *Trans. 7th Prague Conf. Information Theory Statistical Decision Functions Random Processes*, D. Reidel, Boston.

Cowell, F. A. and K. Kuga (1981), Additivity and the entropy concept : an axiomatic approach to inequality measurement, *J. Econ. Theory*, 25, 131－143.

Kullback, S. and R. A. Leibler (1951), On information and sufficiency, *Ann. Math. Statist.* 22, 79－86.

Shorrocks, A. F. (1980), The class of additively decomposable inequality measures, *Econometrica*, 48, 613－625.

Theil, H. (1967), *Economics and Information Theory, North-Holland Publishing Co., Amsterdam.*

Toyoda, T. (1980), Decomposability of inequality measures, *Econ. Studies Quarterly*, 31, 207－216.

［付記］本稿は1994年度法政大学特別研究助成を受けた研究をまとめたものである。