情報処理学会論文誌 Vol.54 No.8 1988–1995 (Aug. 2013)

分布間ダイバージェンスと議席配分方式の関係について

一森 哲男[1,a)]

受付日 2012年11月15日, 採録日 2013年5月18日

概要：本論文の目的は分布間ダイバージェンスと議席配分方式の関係を明らかにすることである．最近，Rényi ダイバージェンスと緩和除数方式と呼ばれる議席配分方式の間の関係が明らかにされた．そこで本論文では，この結果を一般化し，f ダイバージェンスと呼ばれるダイバージェンスのクラスと議席配分方式間の関係を明らかにする．得られた主な結果は，(i) f ダイバージェンスが緩和比例方式に対応すること，(ii) f ダイバージェンスのクラスに属する全変動距離（ℓ_1 距離）が最大剰余法に対応すること，および，(iii) Chernoff の α ダイバージェンスが 1 対 1 に緩和除数方式に対応することである．また，それぞれの分野で得られている知見から他方の分野の結果を比較検討する．

キーワード：ダイバージェンス，議員定数配分，最適化

Relationship between Divergences and Apportionment Methods

Tetsuo Ichimori[1,a)]

Received: November 15, 2012, Accepted: May 18, 2013

Abstract: In this paper, we develop the relationship between some divergences and apportionment methods of representatives. Recently Ichimori has established the correspondence between the Rényi divergence and the relaxed divisor method of apportionment. The aim of this paper is to generalize his results and to develop the correspondence between the class of f-divergences and apportionment methods. The results obtained are (i) the class of f-divergences are in correspondence with the relaxedly proportional methods, (ii) the total variation distance (i.e., ℓ_1 norm) is in correspondence with the largest remainder method and (iii) the class of Chernoff's α-divergences are in one-to-one correspondence with the relaxed divisor methods.

Keywords: divergence, apportionment, optimization

1. はじめに

情報理論では，2 つの分布間の距離を測るためにダイバージェンスという情報量が定義されている．もちろん，ベクトル解析のダイバージェンス（発散）とは別のものである．最も有名なダイバージェンスは Kullback-Leibler によるものであろう [1]（これは相対エントロピーや情報ダイバージェンスなどさまざまな名前で呼ばれている [2]）．これ以外にも，Rényi ダイバージェンス [3] や Chernoff による α ダイバージェンス [4] などいくつものダイバージェンスが知られている．

[1] 大阪工業大学
Osaka Institute of Technology, Hirakata, Osaka 573–0196, Japan
[a)] ichimori@is.oit.ac.jp

政治学では得票や人口に応じて議席を配分することが重要な問題となっている．比例代表制では，政党の獲得した票数に応じて政党間で議席を配分する．アメリカのような連邦制では，州の人口に応じて州間で議席を配分する．政党を州と読み替え，政党の得票を州の人口と読み替えれば，両者は同じ問題と考えられる．この問題は議員定数配分問題と呼ばれ 200 年以上にわたり激しい論争が続けられている [5]．わが国でも 1 票の格差の問題としてよく知られている．議席配分方式とは国勢調査により定まった人口分布に応じて，定められた数の議席を州間で配分するアルゴリズムのことである．つまり，このアルゴリズムにより議席分布が定まる．人口分布と議席分布の両者を確率分布と考えれば，これらの分布間のダイバージェンスが定義できる．この値が最小であれば，議席分布は人口分布に最も近いと

解釈できる．

最近，先行研究で，Rényi ダイバージェンスの最小化が緩和除数方式という配分方式に対応することが明らかにされた [6]．

本論文の目的は，この先行研究の結果を拡張し，さまざまなダイバージェンスがどのような配分方式に対応しているかを明らかにすることである．さらに，それぞれの分野で得られている知見から他方の分野を観察し考察する．このことは，分布間のダイバージェンスや議席配分方式に関して，今後の研究の発展に役立つものと期待できる．

2. f ダイバージェンス

分布間ダイバージェンスとして，Csiszár の f ダイバージェンス [7]（Ali-Silvey 距離ともいう [8]）は最も一般的なものとして知られている．2 つの離散確率分布 $\mathcal{U}=(u_1,u_2,\ldots,u_n)$ と $\mathcal{V}=(v_1,v_2,\ldots,v_n)$ が与えられたとする．ここで，$u_k>0$ $(k=1,2,\ldots,n)$，かつ $\sum_{k=1}^{n}u_k=1$ であり，$v_k>0$ $(k=1,2,\ldots,n)$，かつ $\sum_{k=1}^{n}v_k=1$ である．このとき，両分布間の f ダイバージェンスとは

$$D_f(\mathcal{U}||\mathcal{V})=\sum_{k=1}^{n}v_kf\left(\frac{u_k}{v_k}\right)$$

と定義されている．正の実数の集合を $\mathbb{R}_+$ とし，実数の集合を $\mathbb{R}$ と書く．関数 $f:\mathbb{R}_+\to\mathbb{R}$ は狭義凸で，さらに，$f(1)=0$ と正規化されている．

f ダイバージェンスは数学で定義される距離関数とは異なり，対称性や三角不等式が成り立たない．そのため，厳密には距離とはいえないが，しばしば，$D_f(\mathcal{U}||\mathcal{V})$ は $\mathcal{U}$ から $\mathcal{V}$ まで[*1]の距離という．有名な Jensen の不等式より

$$\sum_{k=1}^{n}v_kf\left(\frac{u_k}{v_k}\right)\geq f\left(\sum_{k=1}^{n}v_k\frac{u_k}{v_k}\right)=f(1)=0$$

が導かれ，その値が非負 $D_f(\mathcal{U}||\mathcal{V})\geq 0$ であることが分かる．また，関数の狭義凸性より，$D_f(\mathcal{U}||\mathcal{V})=0$ となるのは，すべての $k\in\{1,2,\ldots,n\}$ に対し $u_k=v_k$ となるとき，および，そのときのみである．

たとえば，$f(t)=t\log t$ $(t>0)$ とおけば，f ダイバージェンスは

$$\sum_{k=1}^{n}u_k\log\frac{u_k}{v_k}$$

となり，Kullback-Leibler ダイバージェンスが得られる．**表 1** に関数 $f(t)$ とそれにより定まるダイバージェンスをいくつか与える．

表 1 の D_{KL} は Kullback-Leibler ダイバージェンスを示す．D_{χ^2} は Pearson の χ^2 ダイバージェンス [9] を表している．D_{H} は Hellinger 距離の 2 乗（Hellinger ダイバージェンス）[10] と呼ばれ，$\mathcal{U}$ と $\mathcal{V}$ に関して対称である．すなわち，$D_{\mathrm{H}}(\mathcal{U}||\mathcal{V})$ は $D_{\mathrm{H}}(\mathcal{V}||\mathcal{U})$ に等しい．D_{V} は全変動距離（ℓ_1 距離）といわれ，これも $\mathcal{U}$ と $\mathcal{V}$ に関して対称である．

*1　逆に，これを $\mathcal{V}$ から $\mathcal{U}$ までの距離と定義する人もいる．

表 1　代表的なダイバージェンス

Table 1　Typical examples of divergences.

$f(t)$ $(t>0)$	ダイバージェンス
$t\log t$	$D_{\mathrm{KL}}(\mathcal{U}\|\|\mathcal{V})=\sum_{k=1}^{n}u_k\log\frac{u_k}{v_k}$
$-\log t$	$D_{\mathrm{KL}}(\mathcal{V}\|\|\mathcal{U})=\sum_{k=1}^{n}v_k\log\frac{v_k}{u_k}$
$(t-1)^2$	$D_{\chi^2}(\mathcal{U}\|\|\mathcal{V})=\sum_{k=1}^{n}\frac{(u_k-v_k)^2}{v_k}$
$\frac{(t-1)^2}{t}$	$D_{\chi^2}(\mathcal{V}\|\|\mathcal{U})=\sum_{k=1}^{n}\frac{(v_k-u_k)^2}{u_k}$
$\frac{1}{2}\left(\sqrt{t}-1\right)^2$	$D_{\mathrm{H}}(\mathcal{U}\|\|\mathcal{V})=\sum_{k=1}^{n}\frac{1}{2}\left(\sqrt{u}_k-\sqrt{v}_k\right)^2$
$\|t-1\|$	$D_{\mathrm{V}}(\mathcal{U}\|\|\mathcal{V})=\sum_{k=1}^{n}\|u_k-v_k\|$
$\frac{1}{\alpha(\alpha-1)}(t^\alpha-1)$	$D_\alpha(\mathcal{U}\|\|\mathcal{V})=\frac{1}{\alpha(\alpha-1)}\left(\sum_{k=1}^{n}\frac{u_k^\alpha}{v_k^{\alpha-1}}-1\right)$
$\frac{1}{\alpha-1}(t^\alpha-1)$	$D_{\mathrm{T}}(\mathcal{U}\|\|\mathcal{V})=\frac{1}{\alpha-1}\left(\sum_{k=1}^{n}\frac{u_k^\alpha}{v_k^{\alpha-1}}-1\right)$

最後から 2 つ目は Chernoff の α ダイバージェンスであり，これは 1 つのパラメータ $-\infty<\alpha<+\infty$ を含んでいる．しばしば，

$$\frac{4}{1-\alpha^2}\left(1-\sum_{k=1}^{n}u_k^{\frac{1+\alpha}{2}}v_k^{\frac{1-\alpha}{2}}\right)$$

の形で表現されるが，ここでは比較のため，特に，その下の Tsallis ダイバージェンスと比較しやすいような形で表現した．具体的には，上記の式の指数部の $(1+\alpha)/2$ を新たに α とおき直している．$\alpha=1$ の場合は極限操作により

$$f(t)=\lim_{\alpha\to 1}\frac{1}{\alpha(\alpha-1)}(t^\alpha-1)=t\log t$$

となり，Kullback-Leibler ダイバージェンス $D_{\mathrm{KL}}(\mathcal{U}||\mathcal{V})$ が導かれる．$\alpha=0$ の場合は極限操作により $f(t)=-\log t$ となり，Kullback-Leibler ダイバージェンス $D_{\mathrm{KL}}(\mathcal{V}||\mathcal{U})$ が導かれる．

最後の D_{T} は Tsallis ダイバージェンス [11] を表しており，これも 1 つのパラメータ $\alpha>0$ を含んでいる．$\alpha=1$ の場合，極限操作により $f(t)=t\log t$ となり，Kullback-Leibler ダイバージェンスが導かれる．この Tsallis ダイバージェンス，および，これのもとになる Tsallis エントロピーは Tsallis の発見以前から知られていたようである．表 1 の中のダイバージェンスの式を比較することにより，Tsallis ダイバージェンスが Chernoff の α ダイバージェンスの特別な場合であることが分かる．このことからも，Tsallis の発見は新規でないと思われる．

一般に，凸関数はいくらでも存在する．また，凸関数と凸関数の和（より一般的には，凸結合）もまた，凸関数となるので，それに応じて f ダイバージェンスはいくらでも

存在する．たとえば，関数 $f(t) = t \log t$ と $f(t) = -\log t$ から，$f(t) = t \log t + (-\log t) = (t-1)\log t$（$t > 0$）もダイバージェンスを定義する [1]．これは Jeffreys ダイバージェンス [12] と呼ばれるもので，これを記号 $D_{\mathrm{J}}(\mathcal{U}||\mathcal{V})$ で表せば，当然，2 つの Kullback-Leibler ダイバージェンスの和になる．つまり，

$$D_{\mathrm{J}}(\mathcal{U}||\mathcal{V}) = D_{\mathrm{KL}}(\mathcal{U}||\mathcal{V}) + D_{\mathrm{KL}}(\mathcal{V}||\mathcal{U})$$

となる．これは $\mathcal{U}$ と $\mathcal{V}$ に関して対称で，$D_{\mathrm{J}}(\mathcal{U}||\mathcal{V}) = D_{\mathrm{J}}(\mathcal{V}||\mathcal{U})$ である．

3. 議席配分方式

議席配分方式は最大剰余法と除数方式を考慮対象とするのが現実的である．最大剰余法は世界中でよく使われており，わが国でも地域に議席を配分する際には必ず使用されている．たとえば，小選挙区制で選出される衆議院議員の 300 議席は 47 都道府県間で最大剰余法を用いて配分されている．また，残りの 180 議席は 11 ブロック（北海道ブロック，東北ブロックなど）間で最大剰余法を用いて配分されている．一方の除数方式は 1 つの配分方式ではなく，無数の方式を含む配分方式のクラスである．たとえば，11 ブロックに配分された議席は，それぞれ，各ブロック内での比例代表制の選挙結果に基づき，ドント法（別名 Jefferson 方式）という除数方式を用いて政党間で配分されている．

3.1 最大剰余法

議席を州間で配分する問題を考える．州の総数を s，議席の総数を h とする．州 i の人口を p_i とし，そこに配分する議席数を a_i で表す．州全体の集合を $S = \{1, \ldots, s\}$ と定義する．

議席数 h を州の人口に完全に比例させて配分したとき，各州 i の理想の分け前は

$$q_i = h \times \frac{p_i}{\pi}$$

である．ここで，π は総人口で $\sum_{i=1}^{s} p_i$ に等しい．この q_i を州 i の「取り分」という．万一，すべての取り分が整数であれば，問題が解決し，各州の議席配分数は取り分に等しくなる．しかしながら，普通は，取り分は整数ではないため，その端数を切り上げるか切り捨てるかして，取り分を整数に丸める必要がある．ただし，一般的な四捨五入ではうまくいかないことが多い．そこで，小数点以下の端数の大きなものから順にその端数を切り上げ，すべての議席が配分されるようにしたものを「最大剰余法」という．

定義 1（最大剰余法） 議席分布（$a_1, \ldots, a_s$）を定める．

(1) すべての州 $i \in S$ に対して取り分 $q_i = hp_i/\pi$ を求める．
(2) すべての州 $i \in S$ に対して配分議席数を $a_i = \lfloor q_i \rfloor$ とする[*2]．
(3) 残余議席数 $\eta = h - \sum_{i=1}^{s} a_i$ を求める．
(4) すべての州 $i \in S$ に対して端数 $q_i - a_i$ を求め，最大の端数を持つ η 個の州 i に対しては 1 議席を追加する：$a_i = a_i + 1$．

たとえば，ある国が 3 つの州から構成されているとする．州 1 の人口は 235 万人，州 2 の人口は 333 万人，州 3 の人口は 432 万人とする．議席総数を 10 とする．つまり，$s = 3$，$h = 10$，$p_1 = 2{,}350{,}000$，$p_2 = 3{,}330{,}000$，$p_3 = 4{,}320{,}000$ である．計算により，取り分は $q_1 = 2.35$，$q_2 = 3.33$，$q_3 = 4.32$ となる．上記アルゴリズムのステップ (2) での配分は $a_1 = 2$，$a_2 = 3$，$a_3 = 4$ となる．残余の議席数は $\eta = 10 - 2 - 3 - 4 = 1$ となり，端数が最大の州 1 が 1 議席の追加を受け，最終の配分は $a_1 = 3$，$a_2 = 3$，$a_3 = 4$ となる．

上記の数値例に対し，議席総数を 1 割だけ増加させてみる．すなわち，$h = 11$ とする．人口は変化しないとすると，各州の取り分も 1 割だけ増加し，具体的には，$q_1 = 2.585$，$q_2 = 3.663$，$q_3 = 4.752$ となる．ステップ (2) での配分は順に 2 議席，3 議席，4 議席となり，残余の議席数は 2 となる．端数の大きな 2 州は州 2 と 3 なので，それぞれの州に 1 議席が追加され，最終の配分は順に 2 議席，4 議席，5 議席と定まる．以前に比べ州 1 の配分議席数が 3 から 2 に減少している．原因は議席総数の増加だけであるため，この現象はきわめて奇妙である．

実際，アメリカで 1880 年度の国勢調査結果に基づき，最大剰余法を用いて議席を配分しているとき，議席総数が 299 ではアラバマ州に 8 議席が与えられるが，議席総数を 300 に増やすとアラバマ州に配分される議席数は 7 に減ってしまった．そのため，この現象を「アラバマ・パラドックス」と呼んでいる．

最大剰余法はこのような奇妙な現象を引き起こすが，「取り分制約」というきわめて重要な性質を満足する．取り分制約とは，議席配分アルゴリズムにより定められた議席分布（$a_1, \ldots, a_s$）が不等式

$$\lfloor q_i \rfloor \le a_i \le \lceil q_i \rceil \quad \forall i \in S$$

を満たすことである[*3]．そのため，最大剰余法はわが国をはじめ多くの国々で現在も使用され続けている．

3.2 除数方式

アラバマ・パラドックスを避ける方法として知られているのが除数方式と呼ばれる配分方式である．ただし，この方式は取り分制約を満たさない [5]．除数方式の基本は人口 x 人につき 1 議席を配分する．この x の値は「除数」と呼ばれ，すべての議席が配分されるように調整して定められる．このため，除数 x は整数に限定されているのではなく，正の実数とする．最大剰余法は取り分の値を整数に丸

*2 記号 $\lfloor x \rfloor$ は床関数で，x 以下の最大の整数である．

*3 記号 $\lceil x \rceil$ は天井関数で，x 以上の最小の整数である．

めたが，除数方式は人口を x で割った値を整数に丸める．そのときに使われるのが「丸め関数」と呼ばれるものである．非負の整数の集合を $\mathbb{Z}_0$，正の整数の集合を $\mathbb{Z}_+$，非負の実数の集合を $\mathbb{R}_0$ する．丸め関数 $d: \mathbb{Z}_0 \to \mathbb{R}_0$ は狭義増加である．各 $a \in \mathbb{Z}_0$ に対し $a \leq d(a) \leq a+1$ を満たす．また，$d(b) = b$ かつ $d(c) = c+1$ を満たす整数のペア (b, c) は存在しない．ただし，$b \in \mathbb{Z}_+$，$c \in \mathbb{Z}_0$ である．

除数方式では，人口を除数で割った値（実数）を整数に丸めて州の配分議席数を決めるが，より一般的に正の実数 z を丸め関数 $d(a)$ を用いて，整数に丸める規則を以下のように定義する．

定義 2（丸め規則） 実数 $z > 0$ を丸めた整数を $[z]$ で表す．

(1) $z < d(0)$ ならば $[z] = 0$ とする．

(2) $d(a) < z < d(a+1)$ となる $a \in \mathbb{Z}_0$ が存在すれば，$[z] = a+1$ とする．

(3) $z = d(a)$ となる $a \in \mathbb{Z}_0$ が存在すれば，$[z] = a$ または $[z] = a+1$ とする．

たとえば，丸め関数を $d(a) = a+0.5$ $(a \in \mathbb{Z}_0)$ とする除数方式（Webster 方式と呼ばれる）を考える．上記 3.1 節の数値例（ただし，$h = 10$ とする）に対し，除数を $x = 955{,}000$ に選ぶと，$p_1/x = 2.46$，$p_2/x = 3.49$，$p_3/x = 4.52$ が得られる．順に，$d(1) < 2.46 < d(2)$，$d(2) < 3.49 < d(3)$，$d(4) < 4.52 < d(5)$ となり，上記の丸め規則 (2) より，配分は順に 2 議席，3 議席，5 議席と定まる．

いくつかの除数方式とそれにともなう丸め関数 $d(a)$ を**表 2** に与える．Adams 方式，Dean 方式，Hill 方式，Webster 方式，Jefferson 方式の丸め関数は文献 [5] から引用した．残りの丸め関数は文献 [13] から引用した．さらに，丸め関数 $d(a)$ は連続する 2 整数 a と $a+1$ の間の値をとるが，多くは両者の何らかの平均になっている場合が多いので，その両者間の関係も同表に与える．表 2 に現れる Theil 方式の identric 平均と緩和除数方式の Stolarsky 平均は文献 [14] で定義されている．Theil 方式は文献 [15] の脚注 1 で，T&S 方式は文献 [16] でそれぞれ述べられているが，前者は提案だけで，後者はその付録 E で丸め関数が対数平均であることを示している．

表の中の丸め関数の中には，$d(0)$ が定義されていないものもあるが，そのような場合，すべて，極限操作により定義する．よって，T&S 方式では $d(0) = 0$，Theil 方式では $d(0) = 1/e \approx 0.37$ と定まる．最後の緩和除数方式に対して，$\theta \leq 0$ では極限操作により $d(0) = 0$ となり，$\theta > 0$ かつ $\theta \neq 1$ の場合は，そのまま $a = 0$ を代入して，$d(0) = (1/\theta)^{1/(\theta-1)}$ である．

この緩和除数方式で $\theta = 0, 1$ の場合の丸め関数も極限操作で定義する．このとき，

表 2 代表的な除数方式
Table 2 Typical examples of divisor methods.

除数方式名	$d(a)$	a と $a+1$ の関係
Adams	a	小さい方
Dean	$\dfrac{2a(a+1)}{2a+1}$	調和平均
Hill	$\sqrt{a(a+1)}$	幾何平均
T&S	$\dfrac{1}{\log \frac{a+1}{a}}$	対数平均
Theil	$\dfrac{1}{e}\dfrac{(a+1)^{a+1}}{a^a}$	identric 平均
Webster	$a+\frac{1}{2}$	算術平均
Jefferson	$a+1$	大きい方
緩和除数	$\left(\dfrac{(a+1)^\theta - a^\theta}{\theta}\right)^{\frac{1}{\theta-1}}$	Stolarsky 平均

$$\lim_{\theta \to 0} \left(\frac{(a+1)^\theta - a^\theta}{\theta}\right)^{\frac{1}{\theta-1}} = \frac{1}{\log \frac{a+1}{a}}$$

および

$$\lim_{\theta \to 1} \left(\frac{(a+1)^\theta - a^\theta}{\theta}\right)^{\frac{1}{\theta-1}} = \frac{1}{e}\frac{(a+1)^{a+1}}{a^a}$$

となることから，$\theta = 0$ の場合の緩和除数方式は T&S 方式に等しく，$\theta = 1$ の場合の緩和除数方式は Theil 方式に等しくなる．また，簡単な計算より，$\theta = -1$ は Hill 方式に，$\theta = 2$ は Webster 方式に等しくなることが分かる．

3.3 最適化による配分方式の表現

議席の配分方式を最適化問題として記述する試みは古くから行われている．すべての方式が適切な形をした最適化問題として表現されるとは限らないし，1 つの方式をさまざまな形の最適化問題として表現できる可能性もある．ここでは，一般的によく知られた形で表現されている配分方式を取り上げる．

最初に，最適化問題の制約条件を述べる．決定変数は州への配分議席数 a_i $(i \in S)$ である．丸め規則の定義 (1) より，$d(0) = 0$ となる除数方式を用いれば，議席を受け取らない州は存在しないので，そのような配分方式を最適化問題で表現するには，制約：$a_i \in \mathbb{Z}_+$ が必要である．他方，$d(0) > 0$ となる除数方式を含め，その他の配分方式に対しては，制約：$a_i \in \mathbb{Z}_0$ が必要である．さらに，制約条件：$\sum_{i=1}^{s} a_i = h$ が必要である．すなわち，州全体に配分される議席の和は議席総数に等しい．

配分議席数 $a_i = 0$ を許す・許さないの違いはあるものの，それ以外の点では，制約条件はすべての配分方式に共通なので，最小化問題として記述するならば，各配分方式を特徴づける違いは目的関数だけとなる．**表 3** にいくつかの配分方式と，それらを記述する目的関数をそれぞれ示す．Adams 方式，Jefferson 方式の目的関数は文献 [17] に与えられている．最大剰余法，Hill 方式，Webster 方式の

表 3　最適化表現による配分方式
Table 3　Apportionment methods via optimization.

配分方式	目的関数
最大剰余	$\sum_{i=1}^{s} \left\| a_i - p_i \times \frac{h}{\pi} \right\|$
Adams	$\sum_{i=1}^{s} \frac{a_i^2 - a_i}{p_i}$
Dean	$\sum_{i=1}^{s} \frac{a_i^2 - \varphi(a_i)}{2p_i}$
Hill	$\sum_{i=1}^{s} a_i \left(\frac{p_i}{a_i} - \frac{\pi}{h}\right)^2$
T&S	$\sum_{i=1}^{s} p_i \log \frac{p_i}{a_i}$
Theil	$\sum_{i=1}^{s} a_i \log \frac{a_i}{p_i}$
Webster	$\sum_{i=1}^{s} p_i \left(\frac{a_i}{p_i} - \frac{h}{\pi}\right)^2$
Jefferson	$\sum_{i=1}^{s} \frac{a_i^2 + a_i}{p_i}$
緩和除数	$\sum_{i=1}^{s} \frac{a_i^\theta}{\theta(\theta-1)p_i^{\theta-1}}$

目的関数は文献 [5] に与えられている．Webster 方式の目的関数を最初に考案したのは Saint-Laguë [18] であり，Hill 方式の目的関数の正当性を示したのは Huntington [19] である．T&S 方式，Theil 方式および緩和除数方式の目的関数は文献 [20] で与えられている．Dean 方式の目的関数は文献 [21] で与えられているが，これに含まれる関数 $\varphi(a)$ は奇数の調和数と呼ばれ，各 $a \in \mathbb{Z}_+$ に対し，

$$\varphi(a) = 1 + \frac{1}{3} + \frac{1}{5} + \ldots + \frac{1}{2a-1}$$

と定義される．

4.　ダイバージェンスと配分方式との関係

2 章で述べたダイバージェンスと 3 章で述べた議席配分方式との関係を調べる．最初に，f ダイバージェンス・クラスに属する代表的なダイバージェンスと議席配分方式との関係を調べる．次に，f ダイバージェンスそのものがどのような配分方式のクラスに対応するかを明らかにする．最後に，ここで得られた対応関係について考察を与える．

4.1　代表的なダイバージェンスと配分方式との関係

議席配分問題の枠組みの中で，それぞれの f ダイバージェンスを焼き直してみる．すなわち，$n = s$ とし，すべての州 $i \in S$ に対して $u_i = a_i/h$，$v_i = p_i/\pi$，または，$u_i = p_i/\pi$，$v_i = a_i/h$ とする．前者のダイバージェンスを $D_f(\mathcal{A}||\mathcal{P})$ と書き，後者を $D_f(\mathcal{P}||\mathcal{A})$ と書く．

最初に，Kullback-Leibler ダイバージェンスを考えてみる．$D_{\mathrm{KL}}(\mathcal{A}||\mathcal{P})$ は

$$\sum_{k=1}^{n} u_k \log \frac{u_k}{v_k} = \sum_{i=1}^{s} \frac{a_i}{h} \log \frac{\frac{a_i}{h}}{\frac{p_i}{\pi}} = \sum_{i=1}^{s} a_i \log \frac{a_i}{p_i} + h \log \frac{\pi}{h}$$

となり，これの最小化は表 3 より，Theil 方式に対応していることが分かる．ただし，ここで少し注意が必要である．Theil 方式では $d(0) \neq 0$ なので，人口の少ない州では $a_i = 0$ となる可能性がある．本論文のダイバージェンスの定義では $u_i > 0$ を仮定していたので，これに対処するため，慣例に従い，$0 \log 0 = \lim_{t \to 0} t \log t = 0$ と定めることにする [2]．

さらに，$D_{\mathrm{KL}}(\mathcal{P}||\mathcal{A})$ は

$$\sum_{k=1}^{n} u_k \log \frac{u_k}{v_k} = \sum_{i=1}^{s} \frac{p_i}{\pi} \log \frac{\frac{p_i}{\pi}}{\frac{a_i}{h}} = \sum_{i=1}^{s} p_i \log \frac{p_i}{a_i} + \pi \log \frac{h}{\pi}$$

となり，これの最小化は T&S 方式に対応する．T&S 方式では $d(0) = 0$ なので $a_i \geq 1$ となり，上式の分母に a_i がきても問題はない．

次に，χ^2 ダイバージェンスを考えてみる．$D_{\chi^2}(\mathcal{A}||\mathcal{P})$ は

$$\sum_{k=1}^{n} \frac{(u_k - v_k)^2}{v_k} = \sum_{i=1}^{s} \frac{(\frac{a_i}{h} - \frac{p_i}{\pi})^2}{\frac{p_i}{\pi}} = \frac{\pi}{h^2} \sum_{i=1}^{s} p_i \left(\frac{a_i}{p_i} - \frac{h}{\pi}\right)^2$$

となり，Webster 方式に対応することが分かる．さらに，$D_{\chi^2}(\mathcal{P}||\mathcal{A})$ は

$$\sum_{k=1}^{n} \frac{(v_k - u_k)^2}{u_k} = \sum_{i=1}^{s} \frac{(\frac{p_i}{\pi} - \frac{a_i}{h})^2}{\frac{a_i}{h}} = \frac{h}{\pi^2} \sum_{i=1}^{s} a_i \left(\frac{p_i}{a_i} - \frac{\pi}{h}\right)^2$$

となり，Hill 方式に対応することが分かる．

次に，Hellinger ダイバージェンス $D_{\mathrm{H}}(\mathcal{A}||\mathcal{P}) = D_{\mathrm{H}}(\mathcal{P}||\mathcal{A})$ を考えてみる．これは

$$\sum_{i=1}^{s} \frac{1}{2} \left(\sqrt{\frac{a_i}{h}} - \sqrt{\frac{p_i}{\pi}}\right)^2 = 1 - \sqrt{\frac{1}{h\pi}} \sum_{i=1}^{s} \sqrt{a_i p_i}$$

となり，$\theta = 1/2$ の緩和除数方式に対応することが分かる．現在のところ，この配分方式の名前は付けられていないようである．上式の右辺の項 $\sum_i \sqrt{a_i p_i}$ は Bhattacharyya 係数と呼ばれ，$-\log(\sum_i \sqrt{a_i p_i})$ は Bhattacharyya 距離と呼ばれている [22]．だから，$\theta = 1/2$ の緩和除数方式は Bhattacharyya 距離を最小にしているともいえる．

次に，Chernoff の α ダイバージェンスを考えてみる．$D_\alpha(\mathcal{A}||\mathcal{P})$ は

$$\sum_{i=1}^{s} \frac{1}{\alpha(\alpha-1)} \left(\sum_{i=1}^{s} \left(\frac{a_i}{h}\right)^{\alpha} \Big/ \left(\frac{p_i}{\pi}\right)^{\alpha-1} - 1 \right)$$

と書ける．ここで，$\alpha = \theta$ とおくと，

$$\sum_{i=1}^{s} \frac{1}{\theta(\theta-1)} \left(\frac{\pi^{\theta-1}}{h^\theta} \sum_{i=1}^{s} \frac{a_i^\theta}{p_i^{\theta-1}} - 1 \right)$$

が得られる．これの最小化は緩和除数方式に対応していることが分かる．いい換えれば，$\alpha = \theta$ とおいた α ダイバージェンスは，パラメータ θ を持つ緩和除数方式に 1 対 1 に

表 4 ダイバージェンスと配分方式の対応
Table 4 The correspondence between divergences and methods of apportionment.

ダイバージェンス	配分方式	パラメータ θ の値
$D_{\mathrm{V}}(\mathcal{A}\|\|\mathcal{P}) = D_{\mathrm{V}}(\mathcal{P}\|\|\mathcal{A})$	最大剰余	該当せず
$D_{\mathrm{KL}}(\mathcal{A}\|\|\mathcal{P})$	Theil	1
$D_{\mathrm{KL}}(\mathcal{P}\|\|\mathcal{A})$	T&S	0
$D_{\chi^2}(\mathcal{A}\|\|\mathcal{P})$	Webster	2
$D_{\chi^2}(\mathcal{P}\|\|\mathcal{A})$	Hill	-1
$D_{\mathrm{H}}(\mathcal{A}\|\|\mathcal{P}) = D_{\mathrm{H}}(\mathcal{P}\|\|\mathcal{A})$	名無し	$\frac{1}{2}$
$D_{\alpha}(\mathcal{A}\|\|\mathcal{P})$	緩和除数	$(-\infty, +\infty)$
$D_{\mathrm{ET}}(\mathcal{A}\|\|\mathcal{P})$	緩和除数	$(-\infty, +\infty)$

対応している．

Tsallis ダイバージェンスは機械学習や物理学をはじめさまざまな研究に応用されている．これは正のパラメータ $\alpha > 0$ に対して定義されているが，これをすべての値のパラメータに対して定義できるように拡張する．すなわち，$-\infty < \alpha < +\infty$ に対して，

$$D_{\mathrm{ET}}(\mathcal{U}||\mathcal{V}) = \frac{1}{\alpha(\alpha-1)}\left(\sum_{k=1}^{n} u_k^{\alpha} v_k^{1-\alpha}\right)$$

と定義する．しかしながら，これは Chernoff の α ダイバージェンスに等しい．だから，これも緩和除数方式に対応する．

次に，分布に関して対称となる全変動距離（ℓ_1 距離）$D_{\mathrm{V}}(\mathcal{A}||\mathcal{P})$ を考えてみる．これは

$$\sum_{k=1}^{n}|u_k - v_k| = \sum_{i=1}^{s}\left|\frac{a_i}{h} - \frac{p_i}{\pi}\right| = \frac{1}{h}\sum_{i=1}^{s}\left|a_i - p_i \times \frac{h}{\pi}\right|$$

と書けるので，最大剰余法に対応することが分かる．

以上の結果を**表 4** にまとめる．この表では，最大剰余法以外はすべて緩和除数方式であり，それぞれ対応するパラメータ θ の値を明示している．

最後に，これも分布に関して対称となる Jeffreys ダイバージェンス $D_{\mathrm{J}}(\mathcal{A}||\mathcal{P})$ を考えてみる．これは 2 つの Kullback-Leibler ダイバージェンスの和で定義されているので，両者の中間的な配分方式になる．これは両方式の目的関数の定数項を除いた和，すなわち，

$$\sum_{i=1}^{s}\left(a_i \log\frac{a_i}{p_i} + p_i \log\frac{p_i}{a_i}\right)$$

を最小にする配分方式である．ただし，これはあまり満足のいく配分方式ではなく，除数方式にもならないし，取り分制約も満たさない．

配分方式が除数方式になるためには，任意の正の定数 $\lambda > 0$ に対して，人口分布 $(p_1, \ldots, p_s)$ が $(\lambda p_1, \ldots, \lambda p_s)$ に変化しても議席分布 $(a_1, \ldots, a_s)$ は不変とならなければならない [5]．しかしながら，上式の人口 p_i を λp_i に置き換えてみると，上式は

$$\sum_{i=1}^{s}\left(a_i \log\frac{a_i}{p_i} + \lambda p_i \log\frac{p_i}{a_i}\right) - h\log\lambda + \pi\lambda\log\lambda$$

に変化する．十分小さな λ に対しては，上式の最小化は Theil 方式に対応し，十分大きな λ に対しては，T&S 方式に対応することが分かる．すなわち，これは除数方式ではなく，アラバマ・パラドックスを受ける可能性がある．

また，Theil 方式と T&S 方式の与える議席分布が同一となる人口分布の数値例を作ることは可能である．この数値例を用いると，明らかに，Jeffreys ダイバージェンスに対応する配分方式の与える議席分布もこれらの議席分布に一致するはずである．一方，Theil 方式と T&S 方式は除数方式であるため，取り分制約は満たさない．よって，Jeffreys ダイバージェンスに対応する配分方式も取り分制約を満たさない．

4.2 f ダイバージェンスと配分方式の関係

ここでは，より一般的に f ダイバージェンスがどのような配分方式に対応しているかを議論する．いま，u_k（$k = 1, \ldots, n$）を正の変数と考え，v_k（$k = 1, \ldots, n$）は正の定数として，非線形計画問題

$$\min_{\boldsymbol{u}} D_f(\mathcal{U}||\mathcal{V}) \ \mathrm{s.t.} \sum_{k=1}^{n} u_k = 1, u_k > 0\,(k = 1, \ldots, n)$$

を考える．記号 s.t. は条件を表す．最小化は変数 $\boldsymbol{u} = (u_1, \ldots, u_n)$ に関して行う．f ダイバージェンスの性質より $D_f(\mathcal{U}||\mathcal{V})$ の最小値は 0 であり，唯一の最適解は $u_k = v_k$（$k = 1, \ldots, n$）となる．このことから，この最適化問題を（連続緩和された）議席配分問題に焼き直すと

$$\min_{\boldsymbol{a}} D_f(\mathcal{A}||\mathcal{P}) \ \mathrm{s.t.} \sum_{i=1}^{s} a_i = h, a_i > 0\,(i \in S)$$

が得られる．最小化は議席分布 $\boldsymbol{a} = (a_1, \ldots, a_s)$ に関して行う．ただし，州 i に配分される議席数 a_i は整数ではなく正の実数，つまり，連続量としている．いい換えれば，配分議席数の整数条件を線形条件に緩和（連続緩和）している．これの唯一の最適解はすべての州 i に対して $a_i = (h/\pi) \times p_i$ となり，配分議席数 a_i が人口 p_i に完全に比例している．このような性質を持つ配分方式を「緩和比例方式」と呼んでいる [23]．いい換えれば，f ダイバージェンスから導かれる配分方式はすべて緩和比例方式となる．さらに，緩和比例方式のクラスには最大剰余法と緩和除数方式が含まれる [23]（**表 5** を参照）．

最大剰余法は全変動距離 $D_{\mathrm{V}}(\mathcal{A}||\mathcal{P})$ から導かれ，これは取り分制約を満たす．緩和除数方式は Hill 方式，T&S 方式，Theil 方式，Webster などを含む配分方式のクラスであり，これはパラメータ $\alpha = \theta$（$-\infty < \alpha < +\infty$）を持つ α ダイバージェンスに 1 対 1 に対応する [13]．緩和除数方

表 5　対応のまとめ
Table 5　Results of correspondence.

ダイバージェンス	配分方式	特色
Csiszár の f	緩和比例	連続緩和で議席と人口が比例
全変動距離 ℓ_1	最大剰余	取り分制約を満たす
Chernoff の α	緩和除数	アラバマ・パラドックスを回避

式には，Adams 方式，Jefferson 方式，Dean 方式は含まれない[*4]．これらの配分方式が緩和除数方式でないことの証明は文献 [21] で与えられている．厳密にいえば，これらの配分方式が緩和比例方式でないことを証明している．一般に，Adams 方式と Dean 方式は人口の少ない州に非常に有利であり，Jefferson 方式は人口の多い州に非常に有利であることはよく知られているため，好ましい配分方式とはいえない．つまり，緩和除数方式は好ましくない配分方式を排除している．最近，除数方式かつ緩和比例方式となる配分方式は緩和除数方式だけであることが示されており [24]，緩和除数方式の妥当性が増大した．

4.3　対応関係に関する考察

情報理論では，ダイバージェンス間の優劣についての一般的な議論はなされてこなかったようである．しかしながら，ダイバージェンスの中で最も基本的で重要なものは Kullback-Leibler ダイバージェンス $D_{\mathrm{KL}}(\mathcal{A}||\mathcal{P})$ および $D_{\mathrm{KL}}(\mathcal{P}||\mathcal{A})$ であろう．これらに対応する配分方式は，それぞれ，Theil 方式と T&S 方式である．すなわち，情報理論からすれば，Theil 方式と T&S 方式が最も基本的で重要となる．

しかしながら，議席配分の世界では，Theil 方式と T&S 方式は配分方式としてはあまり注目を浴びてこなかった．実のところ，この分野の基本的な著書 [5] の中でも，一言も述べられていない．一方，この世界では，Webster 方式と Hill 方式が最も優れた配分方式と考えられている．これらに対応するダイバージェンスはそれぞれ χ^2 ダイバージェンス $D_{\chi^2}(\mathcal{A}||\mathcal{P})$ と $D_{\chi^2}(\mathcal{P}||\mathcal{A})$ である．さらに，Webster 方式と Hill 方式の一方を選ぶ問題では，必ずしも完全に認められているわけではないが，Webster 方式が Hill 方式より優れているといわれている [5], [13]．議席配分の世界から見ると，χ^2 ダイバージェンス $D_{\chi^2}(\mathcal{A}||\mathcal{P})$ が最も優れている．議席を配分する問題では，さまざまな配分方式を用いて議席を配分し，その中でどの議席分布が人口分布に最も近いかを論じている．だから，議席分布と人口分布では，真の分布は人口分布の方である．この結論を一般的にいえば，分布 $\mathcal{V}$ を真の分布としたときの χ^2 ダイバージェンス $D_{\chi^2}(\mathcal{U}||\mathcal{V})$ が最も優れているといえる．

一般に，ダイバージェンスは対称でないが，多くの応用では対称の方が好ましい．そこで，Jeffreys ダイバージェンスのように，強制的に対称となるダイバージェンスが作られているが，議席配分の観点からすると，このようなことは好ましいことではない．なぜならば，これらに対応する配分方式は，Jeffreys ダイバージェンスに対応する配分方式のように，ほとんどは，取り分制約を満たさず，アラバマ・パラドックスを受けてしまうからである．

[*4] Adams 方式，Jefferson 方式，Dean 方式は緩和除数方式ではないが，除数方式ではあるので，アラバマ・パラドックスを避ける．

5.　結論

本論文では，Csiszár の f ダイバージェンスのクラスに属する代表的なダイバージェンスのいくつかと議席配分方式との対応を明らかにした．このような研究は本論文が初めてと思われる．さらに，その結果を一般化して，f ダイバージェンスに対応する配分方式はすべて緩和比例方式であり，最大剰余法と緩和除数方式を含んでいることを明らかにした．

議席配分方式は無限に存在するが，実際に世界中で使用されている主なものは，それほど多くはない．主な配分方式の例として，最大剰余法はわが国でも長年使用され，かつ，さまざまな国々でも使われている．また，緩和除数方式の中の Hill 方式がアメリカで使用され，Webster 方式がヨーロッパをはじめ，他の多くの国々で使われている．実際に使用している国はないと思われるが，Theil 方式と T&S 方式も妥当な方式と考えられている．いい換えれば，世界中でよく使われている配分方式，あるいは，十分妥当と思われる配分方式は，主に f ダイバージェンスから導かれる．

情報理論や確率論では，f ダイバージェンスは分布間の距離を測る妥当な尺度と考えられるが，今回の結果は，政治学の議席配分の研究でも，f ダイバージェンスが妥当なものということを意味している．

参考文献

[1] Kullback, S. and Leibler, R.A.: On Information and Sufficiency, *Annals of Mathematical Statistics*, Vol.22, pp.79–86 (1951).

[2] Cover, T.M. and Thomas, J.A.: *Elements of Information Theory, 2nd Edition*, Wiley, Hoboken, New Jersey (2006).

[3] Rényi, A.: On Measures of Entropy and Information, *Proc. 4th Berkeley Symposium on Mathematics, Statistics and Probability*, pp.547–561 (1960).

[4] Chernoff, H.: A Measure of Asymptotic Efficieny for Tests of a Hypothesis Based on a Sum of Observations, *Annals of Mathematical Statistics*, Vol.23, No.4, pp.493–507 (1952).

[5] Balinski, M.L. and Young, H.P.: *Fair Representation, Meeting the Ideal of One Man, One Vote, 2nd Edition*, Brookings Institution Press, Washington, D.C. (2001).

[6] 一森哲男：レニーのエントロピーを最大にする議席配分方式について，日本応用数理学会論文誌，Vol.22, No.3, pp.81–96 (2012).

[7] Csiszár, I.: Information-type Measures of Difference of

Probability Distributions and Indirect Observation, *Studia Sci. Math. Hungarica*, Vol.2, pp.299–318 (1967).

[8] Ali, S.M. and Silvey, S.D.: A General Class of Coefficients of Divergence of One Distribution from Another, *Journal of the Royal Statistical Society Series B*, Vol.28, pp.131–142 (1966).

[9] Pearson, K.: On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such That it can be Reasonably Supposed to Have Arisen from Random Sampling, *Philosophical Magazine*, Vol.50, No.302, pp.157–172 (1900).

[10] Hellinger, E.: Neue Begründung der Theorie der Quadratischen Formen von Unendlichenvielen Veränderlichen, *Journal für die Reine und Angewandte Mathematik*, Vol.136, pp.210–271 (1909).

[11] Tsallis, C.: Generalized Entropy-based Criterion for Consistent Testing, *Physical Review E*, Vol.58, No.2, pp.1442–1445 (1998).

[12] Jeffreys, H.: An Invariant Form for the Prior Probability in Estimation Problems, *Proc. Roy. Soc. Lond.*, Ser.A, Vol.186, pp.453–461 (1946).

[13] Ichimori, T.: Relaxed Divisor Methods and their Seat Biases, *Journal of the Operations Research Society of Japan*, Vol.55, No.1, pp.63–72 (2012).

[14] Stolarsky, K.B.: Generalizations of the Logarithmic Mean, *Mathemaitcs Magazine*, Vol.48, pp.87–92 (1975).

[15] Theil, H.: The Desired Political Entropy, *American Political Science Review*, Vol.63, pp.521–525 (1969).

[16] Theil, H. and Schrage, L.: The Apportionment Problem and the European Parliament, *European Economic Review*, Vol.9, pp.247–263 (1977).

[17] Ibaraki, T. and Katoh, N.: *Resource Allocation Problems, Algorithmic Approaches*, MIT Press, Cambridge, Massachusetts (1988).

[18] Sainte-Laguë, A.: La Repésentation Proportionelle et la Méthode des Moindres Carrés, *Compte Rendus Hebdomadaires des Séances de l'Académie des Science*, Vol.151, pp.377–378 (1910).

[19] Huntington, E.V.: The Apportionment of Representatives in Congress, *Trans. American Mathematical Society*, Vol.30, pp.85–110 (1928).

[20] 一森哲男：情報エントロピーと不平等指数について，情報処理学会論文誌，Vol.52, No.11, pp.2984–2988 (2011).

[21] 一森哲男：連続平等性と対称性の観点からみた議員定数配分方法と大域的最適化問題，日本応用数理学会論文誌，Vol.21, No.1, pp.103–124 (2011).

[22] Bhattacharyya, A.: On a Measure of Divergence between Two Statistical Populations Defined by Their Probability Distributions, *Bulletin of the Calcutta Mathematical Society*, Vol.35, pp.99–109 (1943).

[23] Ichimori, T.: New Apportionment Methods and their Quota Property, *JSIAM Letters*, Vol.2, pp.33–36 (2010).

[24] 一森哲男：議員定数配分問題の離散最適化による解法について，日本応用数理学会論文誌，Vol.23, No.1, pp.15–35 (2013).

一森 哲男　（正会員）

昭和28年生．昭和57年大阪大学大学院工学研究科応用物理学専攻博士後期課程修了．同年広島大学工学部助手．昭和60年より大阪工業大学工学部専任講師．昭和63年より大阪工業大学工学部助教授．平成8年より大阪工業大学情報科学部教授．システムの最適化関する研究に従事．工学博士．昭和62年日本オペレーションズ・リサーチ学会事例研究奨励賞受賞．日本オペレーションズ・リサーチ学会，日本応用数理学会各会員．