Apportionment behind the Veil of Ignorance

First Draft: November 27, 2013

Current Draft: August 29, 2015

Junichiro Wada

wada@yokohama-cu.ac.jp

Yokohama City University

22-2 Seto, Kanazawa, Yokohama, Japan

Tel.: +81-45-787-2311 Fax: +81-45-787-2413

Abstract

To bring the apportionment closer to the quota or the apportionment quotient closer to the population quotient, instead of "distance," we introduce *f*-divergence for utilitarianism and Bregman divergence for suitable optimization. Even in our relaxed condition, we find that we must use α-divergence for optimization and show that the minimization of α-divergence induces the same divisor methods that correspond to the Kolm–Atkinson social welfare function, which is bounded by constant relative risk aversion.

Keywords: Apportionment; α–divergence; Kullback–Leibler divergence; Divisor method; Stolarsky mean

JEL classification codes: D63, D72

1. Introduction

The US Constitution decrees that representatives shall be apportioned among the several states according to their respective numbers. Since the Great Compromise, the philosophy of the House of Representatives has been equity between the people (i.e., "one-person one-vote, one-vote one-value") rather than equity between states. Hence, equity between the people must be the objective of the state-level apportionment of representatives.

If we could apportion representatives to states perfectly proportionally to their population, the principle of "one-person one-vote, one-vote one-value" would be mathematically possible. However, this is not usually the case. Historically, three methods have been considered to find a somewhat equal solution: (i) the comparison of two states based on "average" values, (ii) the divisor method, using infinitely many kinds of thresholds to round up or down, and (iii) the constrained optimization of certain objective functions.

Huntington (1928) pioneered inclusive research in the area. He compares two states by using data on the "average" values of each state *j*, such as per capita representatives (n_j/N_j) or per representative population (average district population in single-member district cases) (N_j/n_j).

The Hill method is used to reduce the relative difference between over-represented state *A* and under-represented state *B*, i.e., $((N_B/n_B)/(N_A/n_A))-1$ or $1-((n_B/N_B)/(n_A/N_A))$. On the contrary, if we try to reduce the absolute difference, we use the Dean method in the case of $(N_B/n_B)-(N_A/n_A)$ and the Webster method in the case of $(n_A/N_A)-(n_B/N_B)$. Huntington (1928) tests all these combinations and finds five traditional methods, namely the Adams, Dean, Hill, Webster, and Jefferson methods, and recommends the Hill method, which uses the relative differences between two states.

Balinski and Young (1982) advocate divisor methods, which are the only approaches to avoid the population paradox (they also all avoid the Alabama and new states paradoxes). All divisor methods choose a population-to-representatives ratio as a target and divide the populations of the states by this ratio or “divisor” x to obtain quotients. Each divisor method, however, has its own thresholds, with which the quotients are rounded up or down to the nearest whole number. The whole numbers obtained must then sum to the given number of seats: if the sum is too large (small), the divisor is adjusted upward (downward) until the correct sum results.

From the sets of serial non-negative integers, m-1 and m, these different methods use different thresholds. Adams uses the smaller ones, Dean uses the harmonic means, Hill uses the geometric means, Webster (Saint Laguë) uses the arithmetic means, and Jefferson (d’Hondt) uses the larger ones. Other than these traditional five divisor methods, we could choose any (infinite sets of) thresholds as in the case of modified Saint Laguë, Imperiali, Danish, and so on.

While divisor methods assure a certain degree of equity between states, they offer no assurance of equity between the people. Of the multitude of possible divisor methods, Balinski and Young (1982) choose the Webster method, as the only unbiased divisor method that eliminates any systematic advantage to either small or large states. For the final choice of methods, they also use equity between states as in the case of Huntington (1928).

Constrained optimization is a typical approach in economics or operations research. According to Wada (2010), popular indexes such as Rae, Loosemore–Hanby, Gallagher (least squares), and largest deviation are based on the distance (L_p-norm) between population quotient **N** (N_j / N) and apportionment quotient **n**

(n_j/n), or between quota **q** ($q_j = (N_j/N)n$) and apportionment **n** (n_j).

If we use these indexes for the objective functions, their optimal integer solution for apportionment is given by the Hamilton method (largest remainder; see Birkhoff 1976). The Hamilton method takes quota as the cue. This computes the quotas and then gives to each state the whole number contained in its quota. The seats left over are then distributed to the states that have the larger remainders.

Although the Hamilton method leads to the population paradox, the solution stays within the quota since it uses the quota as the cue.[1] Here, we focus our attention on the fact that each term summed without any weight for the objective function (L_p-norm) represents each state. This objective function concerns equity between states but not equity between representatives or between the people.

Let us use the numerical example of Saari (1994), a supporter of the Hamilton method. As Table 1 shows, the apportionment by the Hamilton method (the largest remainder method) minimizes the L_2-norm (Euclidean distance) between quota **q** and apportionment **n**. However, after apportionment, even if each state makes equal districts, the distance between the district quota and apportionment (1 for each district) is not minimized by using the Hamilton method (see Table 2). According to the state populations or their quotas, the Hamilton method provides a closer apportionment than do the Hill or Webster methods. However, after apportionment, single-member districts result. Indeed, the average district quotas of Hamilton show that the apportionment (1) is more unfair than those of Hill or Webster (1.57 for state A in Hamilton compared with 1.005 for state C in Hill or Webster). The cue of the Hamilton method is thus the quota; however, as noted above, this concerns only equity between states and not equity between representatives (districts) or populations.

[Tables 1 and 2 about here]

Wada (2012) considers representatives as income or wealth and uses the Kolm–Atkinson social welfare function as the objective function. As the following shows, the Kolm–Atkinson social welfare function is utilitarian; hence, we can consider its maximization as an expected utility maximization behind the veil of ignorance or on the constitutional stage:

$$KASWF^{\varepsilon} = \sum_j \frac{N_j}{N} \frac{1}{(1-\varepsilon)} \left(\left(\frac{\frac{n_j}{n}}{\frac{N_j}{N}} \right)^{(1-\varepsilon)} - 1 \right)$$

Here, $\varepsilon \to \infty$ makes it the Rawlsian social welfare function, $\varepsilon \to 1$ makes it the Nash social welfare function, and $\varepsilon = 0$ makes it the Benthamian social welfare function.[2]

Wada (2012) multiplies the function by $(-1/\varepsilon)$ and turns the maximization problem into a minimization problem of the generalized entropy index $(\alpha \equiv 1-\varepsilon)$:

$$E^{\alpha} = \frac{1}{\alpha(\alpha-1)} \sum_j \frac{1}{N} N_j \left(\left(\frac{\frac{n_j}{N_j}}{\frac{n}{N}} \right)^{\alpha} - 1 \right)$$

Here, $\alpha \to 0$ ($\varepsilon \to 1$), corresponding to the Nash, makes it mean log deviation, $\alpha \to 1$ ($\varepsilon \to 0$), corresponding the Benthamian, makes it the Theil index, and $\alpha = 2$ makes it half of the squared coefficient of variation.

Against the background of the veil of ignorance, by minimizing the generalized entropy indexes for the integer solutions, Wada (2012) succeeds in finding divisor methods with the thresholds of the Stolarsky mean, which includes four of the five traditional divisor methods as well as the divisor method with the thresholds of the logarithmic mean founded by the Nash social welfare function and the divisor method with the thresholds of the identric mean founded by the Benthamian social welfare function.

Since Wada (2012) starts with the Kolm–Atkinson social welfare function, we could say that this concerns equity between the people. However, although the utility function with constant relative risk aversion is the most often used utility function in economics, the Kolm–Atkinson social welfare function depends on the specialization of the utility function form. Moreover, it does not derive the Dean method[3] or the divisor methods used in the real world such as modified Saint Laguë, Imperiali, Danish, and so on.

For equity between the people behind the veil of ignorance, we should consider using a more general objective function form. In this paper, we use "quasi-distance" or divergence instead of "distance" associated with the L_p-norm in order to bring the apportionment vector **n** closer to the quota vector **q**, or to bring the apportionment quotient vector **Q** closer to the population quotient vector **P**.

The remainder of the paper is structured as follows. In section 2, we introduce the general idea of "quasi-distance" or divergence. Sections 3 and 4 introduce f-divergence for utilitarianism and Bregman divergence for suitable optimization. In section 5, we find, even in our relaxed condition, that we must use α-divergence for optimization, while section 6 shows that the minimization of α-divergence induces the same divisor methods induced by the Kolm–Atkinson social welfare function that is bounded by constant relative risk aversion. Section 7 discusses and provides concluding remarks.

2. Divergence

Distance is a function that assigns a real number $d(\mathbf{u}\|\mathbf{v})$ to every ordered pair of points (**u, v**). The distance axioms are as follows:

a. Non-negativity

For all **u**, **v**, $d(\mathbf{u}\|\mathbf{v}) \geq 0$.

b. Zero property

$d(\mathbf{u}\|\mathbf{v}) = 0$ if, and only if, $\mathbf{u} = \mathbf{v}$.

c. Symmetry

For all **u**, **v**, $d(\mathbf{u}\|\mathbf{v}) = d(\mathbf{v}\|\mathbf{u})$.

d. Triangle inequality

For all **u**, **v**, and **w**, $d(\mathbf{u}\|\mathbf{v}) \leq d(\mathbf{u}\|\mathbf{w}) + d(\mathbf{w}\|\mathbf{v})$.

Like distance, divergence is a function that assigns a real number $D(\mathbf{u}\|\mathbf{v})$ to every ordered pair of points (**u, v**). The divergence axioms are as follows:

a. Non-negativity

For all **u**, **v**, $D(\mathbf{u}\|\mathbf{v}) \geq 0$.

b. Zero property

$D(\mathbf{u}\|\mathbf{v}) = 0$ if and only if $\mathbf{u} = \mathbf{v}$.

c. Positive definite

Points **u** and **v** are located near to each other and the coordinates are **u** and $\mathbf{u} + d\mathbf{u}$. We can express $D(\mathbf{u}\|\mathbf{u} + d\mathbf{u})$ as a Taylor expansion,

$$D(\mathbf{u} \parallel \mathbf{u} + d\mathbf{u}) = \frac{1}{2}\sum g_{ij}(\mathbf{u}) du_i du_j \ .$$

Here $G(\mathbf{u})=(g_{ij}(\mathbf{u}))$ is strictly positive definite.

Divergence is often used in statistics and information geometry to judge how close distributions are. It is thus a suitable function with which to judge the closeness of the apportionment distribution to the population distribution. Here, we need neither triangle inequality nor symmetry. In the apportionment problem, the origin must be a population distribution and we simply choose the closest apportionment distribution. There are many kinds of divergences in mathematics, and we must choose the form that is suitable for our apportionment problem.

3. *f*-divergence

The *f*-divergence from vector **u** to vector **v** is defined as follows:

$$D^f(\mathbf{u} \parallel \mathbf{v}) = \sum u_j f\left(\frac{v_j}{u_j}\right)$$

$$f : \Re^+ \to \Re, \text{ convex}, \quad f(1) = 0$$

$$0f\left(\frac{0}{0}\right) = 0, f(0) = \lim_{t \to 0} f(t), 0f\left(\frac{a}{0}\right) = \lim_{t \to 0} tf\left(\frac{a}{t}\right)$$

Since we intend to measure the divergence from the quota vector **q** to the apportionment vector **n**, or the divergence from the population quotient vector **P** to the apportionment quotient vector **Q**, the requirement of the domain $\Re^+$ is suitable. As in the case of the United States, we can district each state equally. Thus, we can use $\frac{n_j}{q_j}$ $\left(\frac{n_j}{\frac{N_j}{N}n} = \frac{n_j}{N_j} \cdot \frac{N}{n}\right)$ or $\frac{Q_j}{P_j}$ $\left(\frac{\frac{n_j}{\sum n_j}}{\frac{N_j}{\sum N_j}} = \frac{n_j}{N_j} \cdot \frac{N}{n}\right)$ for the value of $\frac{v_j}{u_j}$ and weight each f with its "population," q_j or P_j, for the perspective of the people. Furthermore, if we define $U(t) = -f(t)$, $U(t)$ becomes a concave function; hence, the minimization of the divergence problem would be understood as the maximization of utilitarian social welfare with a general individual utility function, $U(t)$ or $U\left(\frac{n_j}{N_j}\cdot\frac{N}{n}\right)$ in both cases. We should thus use f-divergence for utilitarianism or optimization behind the veil of ignorance.

4. Bregman divergence

Bregman divergence, which is often used for optimization problems, from vector **u** to vector **v** is defined as follows:

$$D^{Bregman}(\mathbf{u} \,\|\, \mathbf{v}) = \phi(\mathbf{v}) - \phi(\mathbf{u}) - \langle \mathbf{v} - \mathbf{u}, \nabla\phi(\mathbf{u}) \rangle$$

Here, ϕ is a strictly convex and continuously differentiable function. If we use the squared magnitude of vector $\|\mathbf{u}\|^2$ for ϕ, $D^{Bregman}$ becomes the squared Euclidean distance. Figure 1 shows Bregman divergence in one dimension, indicating that the requirement of strict convexity is good for optimization. Indeed, it is necessary for apportionment problems without arbitrariness.

[Figure 1 about here]

5. α-divergence and Kullback–Leibler (KL) divergence

Amari's (2009) finding that "α-divergence is unique, belonging to both f-divergence and Bregman divergence classes" shows that the class of α-divergence is the intersection of the classes of f-divergence and Bregman divergence in a manifold of positive measures. KL divergence and its duality are the only divergences belonging to the intersection of f-divergence and Bregman divergence in the space of the probability distribution.

For vector **u** to vector **v**, α-divergence in a manifold of positive measures can be defined as follows (here $u=\sum u_j$ and $v=\sum v_j$.):

$$D^{\alpha}(\mathbf{u}\,\|\,\mathbf{v})=\sum u_j\frac{1}{\alpha(\alpha-1)}\left(\left(\frac{v_j}{u_j}\right)^{\alpha}-1\right)+\frac{v-u}{1-\alpha}$$

$$=\sum\frac{1}{\alpha(\alpha-1)}\left((v_j)^{\alpha}(u_j)^{1-\alpha}-u_j\right)+\frac{v-u}{1-\alpha}$$

$$=\sum v_j\frac{1}{\alpha(\alpha-1)}\left(\left(\frac{u_j}{v_j}\right)^{1-\alpha}-1\right)+\frac{u-v}{\alpha}=D^{1-\alpha}(\mathbf{v}\,\|\,\mathbf{u})$$

$\alpha\to 0$

$$D^{0}(\mathbf{u}\,\|\,\mathbf{v})=\sum u_j\left(-\log\left(\frac{v_j}{u_j}\right)\right)+u-v$$

$$=\sum v_j\frac{u_j}{v_j}\log\left(\frac{u_j}{v_J}\right)+u-v=D^{1}(\mathbf{v}\,\|\,\mathbf{u})$$

$\alpha\to 1$

$$D^{1}(\mathbf{u}\,\|\,\mathbf{v})=\sum u_j\frac{v_j}{u_j}\log\left(\frac{v_j}{u_j}\right)+v-u$$

$$=\sum v_j\left(-\log\left(\frac{u_j}{v_j}\right)\right)+v-u=D^{0}(\mathbf{v}\,\|\,\mathbf{u})$$

KL divergence is α-divergence in the case of $\alpha\to 0$, and thus it can be defined as follows in the space of the probability distribution where u=v=1:

$$D^{0}(\mathbf{u}\,\|\,\mathbf{v})=\sum u_j\left(-\log\left(\frac{v_j}{u_j}\right)\right)$$

Its duality is the case of $\alpha\to 1$:

$$D^{1}(\mathbf{u}\,\|\,\mathbf{v})=\sum u_j\frac{v_j}{u_j}\log\left(\frac{v_j}{u_j}\right)$$

6. Apportionment for minimizing α-divergence and KL divergence

Under our requirement of using Bregman divergence and *f*-divergence, if we intend to minimize the quasi-distance or divergence from the quota vector **q** to the apportionment vector **n**, since $\sum n_j = n = \sum q_j$, we must use α-divergence:

$$D^{\alpha}(\mathbf{q} \,\|\, \mathbf{n}) = \sum \frac{1}{\alpha(\alpha-1)}\left((n_j)^{\alpha}(q_j)^{1-\alpha} - q_j\right)$$

When $\alpha \neq 0$ and $\alpha \neq 1$, the optimal apportionment must satisfy the following condition:

$\forall s,t$, $s \neq t$ and $n_s > 0$, $n_t \geq 0$

$$\frac{1}{\alpha(\alpha-1)}\left((n_s-1)^{\alpha} N_s^{1-\alpha} + (n_t+1)^{\alpha} N_t^{1-\alpha}\right) \geq \frac{1}{\alpha(\alpha-1)}\left(n_s^{\alpha} N_s^{1-\alpha} + n_t^{\alpha} N_t^{1-\alpha}\right)$$

The above optimal condition implies:

$$\min_{n_s>0}\left(\frac{N_s}{\left(\frac{n_s^{\alpha} - (n_s-1)^{\alpha}}{\alpha}\right)^{-\frac{1}{1-\alpha}}}\right) \geq \max_{n_t \geq 0}\left(\frac{N_t}{\left(\frac{(n_t+1)^{\alpha} - n_t^{\alpha}}{\alpha}\right)^{-\frac{1}{1-\alpha}}}\right)$$

Since this apportionment satisfies the divisor methods (Balinski and Young 1982), we can restate it as follows. Find a divisor x so that n_js, which are the “special rounded” numbers of the quotients of states, N_j/x, add up to the required total, n. Here, special rounded refers to being rounded up when the quotient is equal to or greater than the Stolarsky mean, $\left(\frac{n_j^{a} - (n_j-1)^{a}}{a}\right)^{\frac{1}{a-1}}$, of both side integers ((n_j-1) and n_j).

In brief, we obtain the following proposition:

Proposition 1

To minimize Bregman divergence and *f*-divergence, or α-divergence, from quotas to apportionment, we must use the divisor apportionment method with the threshold of the Stolarsky mean of both side integers ($(n_j - 1)$ and n_j).

When $\alpha \to -\infty$, we obtain the minimum number ($n_j - 1$). This represents the Adams method (1+d'Hondt method). $\alpha = -1$ gives the geometric mean and refers to the Hill method (US House of Representatives method). When $\alpha \to 0$, the threshold becomes[4] the logarithmic mean, $\frac{1}{\log n_j - \log(n_j - 1)}$, and when $\alpha \to 1$, the threshold becomes[5] the identric mean, $\frac{n_j^{\,n_j}}{e(n_j - 1)^{(n_j - 1)}}$. $\alpha = 2$ gives the arithmetic mean and the Webster method (Saint Laguë method). When $\alpha \to \infty$, we arrive at the maximum number, n_i. This represents the Jefferson method (d'Hondt method). We show the results in Table 3.

[Table 3 about here]

Under our requirement of using Bregman divergence and *f*-divergence, if we want to minimize the quasi-distance or divergence from the population quotient vector **P** ($P_j = \frac{N_j}{\sum N_j}$) to the apportionment quotient vector **Q** ($Q_j = \frac{n_j}{\sum n_j}$), we must use KL divergence or its duality:

$$D^0(\mathbf{P}\,\|\,\mathbf{Q})=\sum\frac{N_j}{N}\log\left(\frac{\frac{N_j}{N}}{\frac{n_j}{n}}\right)\qquad D^1(\mathbf{P}\,\|\,\mathbf{Q})=\sum\frac{N_j}{N}\left(\frac{\frac{N_j}{N}}{\frac{n_j}{n}}\right)\log\left(\frac{\frac{N_j}{N}}{\frac{n_j}{n}}\right)$$

This is the same as the case of α-divergence with $\alpha\to 0$ and $\alpha\to 1$, and we obtain [6] a divisor method with the threshold of the logarithmic mean, $\frac{n_j-(n_j-1)}{\log n_j-\log(n_j-1)}$ and a divisor method with the threshold of the identric mean,

$$\frac{{n_j}^{n_j}}{e(n_j-1)^{(n_j-1)}}.$$

Proposition 2

To minimize Bregman divergence and *f*-divergence, or KL divergence and its duality, from the population quotient to the apportionment quotient, we must use the divisor apportionment method with the threshold of the logarithmic mean of both side integers ((n_j-1) and n_j) and the threshold of the identric mean of them.

Wada (2012) multiplies the Kolm–Atkinson social welfare function by $-\frac{1}{\varepsilon}\left(=\frac{1}{(\alpha-1)}\right)$ to turn the maximizing social welfare problem into the minimizing generalized entropy problem. This leads not only to the Adams-, Hill-, and Nash-based methods but also to the Webster-, Jefferson-, and Benthamian-based methods. Here, we must be careful that the function $f(t)$ used for *f*-divergence is convex but not necessarily monotonic; this means $U(t)$ is concave but not necessarily increasing monotonically. If we keep the principle of equity between the people or optimization behind the veil of ignorance, we should use $\alpha<1$ ($\varepsilon>0$). This is especially so if we are concerned about the quasi-distance from the population quotient to the apportionment quotient, as this is the expected utility in the space of the probability distribution; hence, we should use KL divergence ($\alpha \to 0$ ($\varepsilon \to 1$)) and the divisor method with the threshold of the logarithmic mean. Since α-divergence has a character of duality as shown in section 5, we could transform the case of divergence α>1 into the optimization problem of equity between representatives. However, in a democracy, we should keep the principle of equity between the people, rather than between the politicians (districts) or states.

7. Discussion and concluding remarks

By using the Kolm–Atkinson social welfare function, which is supported by a utility function with constant relative risk aversion, Wada (2012) derives the divisor apportionment method with the threshold of the Stolarsky mean from generalized entropy. As the form of *f*-divergence and Bregman divergence shows, our condition should be more extensive than the utility function with constant relative risk aversion. However, the result is exactly the same. Thus, if we choose optimization behind the veil of ignorance based on the principle of “one-person one-vote, one-vote one-value” (or equity between the people), we may not need to consider using other utility functions. In other words, the methods of Dean, modified Saint Laguë, Imperiali, Danish, and so on are not purely based on the equity between the people principle.

As Wada (2012) shows, the divisor method using the logarithmic mean is supported by the Nash social welfare function, which has some good characteristics (Kaneko and Nakamura 1979). Here, we add one more supporting reason: if we need to minimize to a suitable divergence from the population quotient to the apportionment quotient, especially if we want to keep the foundation of utilitarianism, we must use the divisor method using the logarithmic mean.

The Kolm–Atkinson social welfare function corresponding to α-divergence would be useful for evaluating representation in the real world with a micro-level foundation. α-divergence at the district level (left-hand side of the following equations) can be collated as the sum of apportionment equity (equity at the level between states) and the weighted $\left(\left(\frac{n_j}{n}\right)^{\alpha}\left(\frac{N_j}{N}\right)^{1-\alpha}\right)$ sum of districting equity (equity at the level within states.) This character would also be useful for further research.

$$\sum_{j=1}^{k}\sum_{i=1}^{k_j}\frac{N_{ji}}{N}\frac{1}{\alpha(\alpha-1)}\left(\left(\frac{\frac{n_{ji}}{n}}{\frac{N_{ji}}{N}}\right)^{\alpha}-1\right)$$

$$=\sum_{j=1}^{k}\frac{N_j}{N}\frac{1}{\alpha(\alpha-1)}\left(\left(\frac{\frac{n_j}{n}}{\frac{N_j}{N}}\right)^{\alpha}-1\right)+\sum_{j=1}^{k}\left(\frac{n_j}{n}\right)^{\alpha}\left(\frac{N_j}{N}\right)^{1-\alpha}\sum_{i=1}^{k_j}\frac{N_{ji}}{N_j}\frac{1}{\alpha(\alpha-1)}\left(\left(\frac{\frac{n_{ji}}{n_j}}{\frac{N_{ji}}{N_j}}\right)^{\alpha}-1\right)$$

$$\alpha\neq 0,1$$

$\alpha\to 0$ (KL divergence)

$$\sum_{j=1}^{k}\sum_{i=1}^{k_j}\frac{N_{ji}}{N}\left(-\log\left(\frac{\frac{n_{ji}}{n}}{\frac{N_{ji}}{N}}\right)\right)$$

$$=\sum_{j=1}^{k}\frac{N_j}{N}\left(-\log\left(\frac{\frac{n_j}{n}}{\frac{N_j}{N}}\right)\right)+\sum_{j=1}^{k}\frac{N_j}{N}\sum_{i=1}^{kj}\frac{N_{ji}}{N_j}\left(-\log\left(\frac{\frac{n_{ji}}{n_j}}{\frac{N_{ji}}{N_j}}\right)\right)$$

$\alpha\to 1$

$$\sum_{j=1}^{k}\sum_{i=1}^{k_j}\frac{N_{ji}}{N}\left(\frac{\frac{n_{ji}}{n}}{\frac{N_{ji}}{N}}\log\left(\frac{\frac{n_{ji}}{n}}{\frac{N_{ji}}{N}}\right)\right)$$

$$=\sum_{j=1}^{k}\frac{N_j}{N}\left(\frac{\frac{n_j}{n}}{\frac{N_j}{N}}\log\left(\frac{\frac{n_j}{n}}{\frac{N_j}{N}}\right)\right)+\sum_{j=1}^{k}\frac{n_j}{n}\sum_{i=1}^{kj}\frac{N_{ji}}{N_j}\left(\frac{\frac{n_{ji}}{n_j}}{\frac{N_{ji}}{N_j}}\log\left(\frac{\frac{n_{ji}}{n_j}}{\frac{N_{ji}}{N_j}}\right)\right)$$

Acknowledgements

A previous version was presented at a Japan Law and Economics Association seminar, to the Public Choice Society, and to the Japanese Economic Association under the title of "Apportionment Method from the Viewpoint of Divergence." I am grateful to Tatsuo Oyama, Kazuyuki Fujii, Tatsuo Suzuki, Steven Brams, Koichi Suga, Donald Saari, Bernard Grofman, Rein Taagepera, and Shun'ichi Amari. This work was supported by a Grant-in-Aid for Scientific Research (C 25380161) from the Japan Society for the Promotion of Science.

References

Amari S (2009) α-divergence is unique, belonging to both *f*-divergence and Bregman divergence classes. *IEEE Transactions on Information Theory* 55(11): 4925–4931.

Balinski ML and Young HP (1982) *Fair Representation*. New Haven: Yale University Press.

Birkhoff G (1976) House monotone apportionment schemes. *Proceedings of the National Academy of Sciences, USA* 73: 684–686.

Huntington EV (1928) The apportionment of representatives in Congress. *Transactions of the American Mathematical Society* 30: 85–110.

Kaneko M and Nakamura K (1979) The Nash social welfare function. *Econometrica* 47(2): 423–435.

Saari D (1994) *Geometry of Voting*. Berlin: Springer-Verlag.

Wada J (2010) Evaluating the unfairness of representation with the Nash Social welfare function. *Journal of Theoretical Politics* 22(4): 445–467.

Wada J (2012) A divisor apportionment method based on the Kolm–Atkinson social welfare function and generalized entropy. *Mathematical Social Sciences* 63(3): 243–247.

Table 1. Distance between state quota and state apportionment

			apportionment methods		
			Hamilton largest remainde	Hill US House	Webster Saint Laguë
state	state population	state quota	apportionment	apportionment	apportionment
A	1570	1.570	1	2	2
B	26630	26.630	27	27	27
C	171800	171.800	172	171	171
	Euclidean distance between state quota and state apportionment		**0.708**	**0.981**	**0.981**

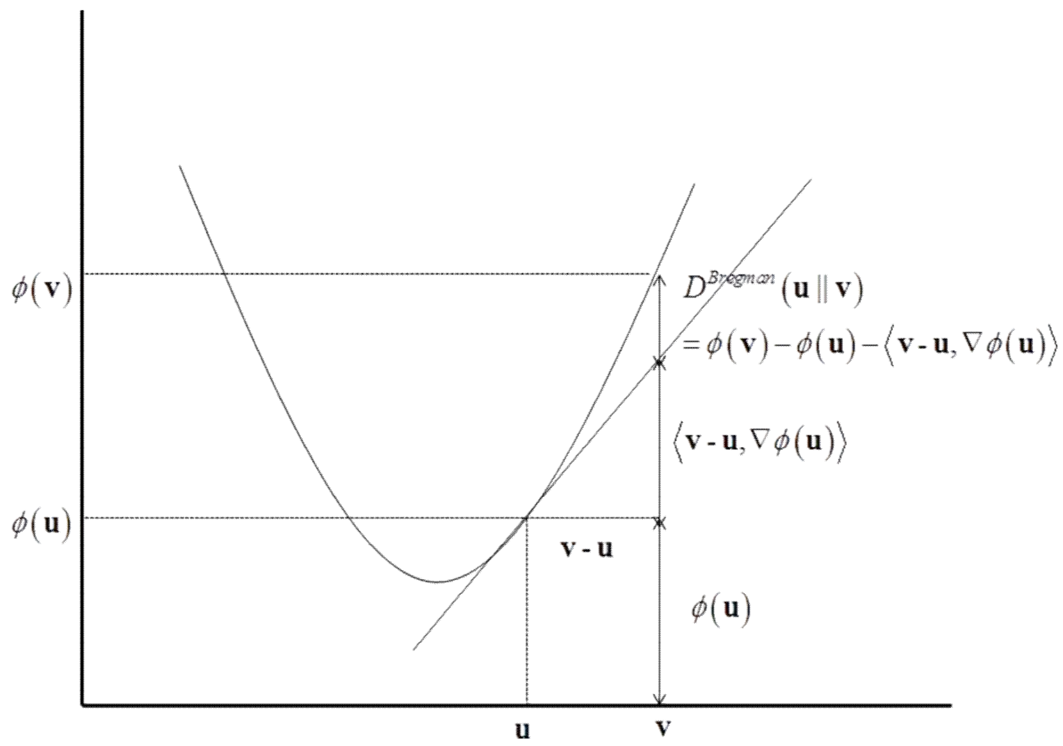
Table 2. Distance between district quota and district apportionment (1)

		apportionment methods		
		Hamilton largest remainde	Hill US House	Webster Saint Laguë
state district (apportionment)	A	1	2	2
	B	27	27	27
	C	172	171	171
average district population	A	1570	785	785
	B	986	986	986
	C	999	1005	1005
average district quota	A	1.570	0.785	0.785
	B	0.986	0.986	0.986
	C	0.999	1.005	1.005
Euclidean distance between district quota nd district apportionment (1)		**0.575**	**0.318**	**0.318**

Table 3. Quotients needed for seats

Traditional name of the methods	Adams 1+d'Hondt	Hill US House			Webster Sainte-Lague	Jefferson d'Hondt
Kolm–Atkinson SWF	*Rawlsian SWF*		**Nash SWF**	*Benthamian SWF*		
ε	**∞**	2	**1**	*0*		
G.entropy (Wada, 2012)		$1/2cv^2$	**MLD**	Theil index	$1/2CV^2$	
α-divergence	-∞	-1	**0 (KL-divergence)**	*1*	2	∞
Stolarsky mean	minimum	geometric mean	**logarithmic mean**	identric mean	arithmetic mean	maximum
Quotient needed for a seats						
1	0	0	**0**	*0.3679*	0.5	1
2	1	1.4142	**1.4427**	*1.4715*	1.5	2
3	2	2.4495	**2.4663**	*2.4832*	2.5	3
4	3	3.4641	**3.4761**	*3.4880*	3.5	4
5	4	4.4721	**4.4814**	*4.4907*	4.5	5
6	5	5.4772	**5.4848**	*5.4924*	5.5	6
7	6	6.4807	**6.4872**	*6.4936*	6.5	7
8	7	7.4833	**7.4889**	*7.4944*	7.5	8
9	8	8.4853	**8.4902**	*8.4951*	8.5	9
10	9	9.4868	**9.4912**	*9.4956*	9.5	10

Figure 1. Image of Bregman divergence



[1] Actually, no method avoids the population paradox and always stays within the quota (Balinski and Young 1982).

[2] The Adams method corresponds to the Rawlsian social welfare function and the Hill method corresponds to the case of ε=2.

[3] The harmonic mean, which is the threshold for the Dean method, is not the Stolarsky mean.

[4] We can consider $0^0 = 1$.

[5] We can consider $0 \log 0 = 0$, and $0^0 = 1$.

[6] We can consider $0^0 = 1$.