Quotient dissimilarities, Euclidean embeddability, and Huygens' weak principle

François Bavaud

Section d'Informatique et de Méthodes Mathématiques,
Faculté des Lettres, Université de Lausanne
CH-1015 Lausanne-Dorigny
(e-mail: Francois.Bavaud@imm.unil.ch)

Abstract. We introduce a broad class of categorical dissimilarities, the quotient dissimilarities, for which aggregation invariance is automatically satisfied. This class contains the chi-square, ratio, Kullback-Leibler and Hellinger dissimilarities, as well as presumably new "power" and "threshold" dissimilarity families. For a large subclass of the latter, the product dissimilarities, we show that the Euclidean embeddability property on one hand and the weak Huygens' principle on the other hand are mutually exclusive, the only exception being provided by the chi-square dissimilarity D^{χ}. Various suggestions are presented, aimed at generalizing Factorial Correspondence Analysis beyond the chi-square metric, by non-linear distortion of departures from independence. In particular, the central inertia appearing in one formulation turns out to exactly amount to the mutual information of Information Theory. [1]

1 Introduction and notations

Let n_{jk} be an $(|J| \times |K|)$ contingency table, with relative frequency $f_{jk} := n_{jk}/n$, row profiles $w_{jk} := n_{jk}/n_{j\bullet}$, column profiles $w^*_{kj} := n_{jk}/n_{\bullet k}$ and marginal (strictly positive) profiles $\rho^*_j := n_{j\bullet}/n = f_{j\bullet}$ and $\rho_k := n_{\bullet k}/n = f_{\bullet k}$, where $n_{j\bullet} := \sum_{k\in K} n_{jk}$ are the row marginals, $n_{\bullet k} := \sum_{j\in J} n_{jk}$ are the column marginals, and $n := n_{\bullet\bullet}$ is the grand total. By construction, $f_{jk} = \rho^*_j w_{jk} = \rho_k w^*_{kj}$; also, the row and column profiles transform as $w_{jk} = \rho_k w^*_{kj}/\rho^*_j$ and $w^*_{kj} = \rho^*_j w_{jk}/\rho_k$.

Independence quotients constitute a most convenient cell-by-cell representation of departures from independence. Surprisingly enough, they are not routinely used in standard Statistics, with the exception of quantitative Geography and Economics, where they are referred (when J or K enumerates regions) to as *location quotients*.

[1] Stimulating discussions with Martin Rajman and Jean-Cédric Chappelier are gratefully acknowledged. This work has benefited from a join grant from the University of Lausanne and the Swiss Federal Institute of Technology of Lausanne.

Definition 1 Independence quotients q_{jk} are the ratios of the observed counts to the expected counts under independence:

$$q_{jk} := \frac{n_{jk}\, n}{n_{j\bullet}\, n_{\bullet k}} = \frac{f_{jk}}{\rho_j^*\, \rho_k} = \frac{w_{jk}}{\rho_k} = \frac{w_{kj}^*}{\rho_j^*} \tag{1}$$

Property 1 *$q_{jk} = 1$ for all cells iff perfect independence holds. Also, in any case,*

$$\sum_{j\in J} \rho_j^*\, q_{jk} = 1 \quad \forall k \in K \qquad\qquad \sum_{k\in K} \rho_k\, q_{jk} = 1 \quad \forall j \in J \tag{2}$$

Conversely, independence quotients q_{jk} obeying (2) determine a contingency table unique up to a multiplicative constant by $n_{jk} := n\, \rho_j^*\, \rho_k\, q_{jk}$.

A dissimilarity D between J objects is a symmetric $(J \times J)$ non-negative matrix $D_{jj'}$ with a null diagonal. The most popular dissimilarity for contingency tables is the chi-square dissimilarity between rows j and j' expressing in terms of independence quotients as

$$D_{jj'}^{\chi} := \sum_{k\in K} \frac{n}{n_{\bullet k}} \left(\frac{n_{jk}}{n_{j\bullet}} - \frac{n_{j'k}}{n_{j'\bullet}}\right)^2 = \sum_{k\in K} \rho_k\, (q_{jk} - q_{j'k})^2 \tag{3}$$

D^{χ} is well known to be aggregation invariant (definition 3). Furthermore, the chi-square measure of row/column dependence obtains as half the weighted average of the dissimilarity between each pairs of rows, or *equivalently*, as the weighted average of the dissimilarity between each row and the average profile (remark 1):

$$\frac{\chi^2}{n} = \frac{1}{2} \sum_{jj'} \rho_j^*\, \rho_{j'}^*\, D_{jj'}^{\chi} = \sum_j \rho_j^*\, D_{j\rho}^{\chi} \qquad \text{where } D_{j\rho}^{\chi} := \sum_k \rho_k (q_{jk} - 1)^2$$

The equivalence emphasized above constitutes the *weak Huygens' principle* (definition 6). In the present case, the *strong Huygens' principle* $\sum_j \rho_j^*\, D_{ja}^{\chi} = \sum_j \rho_j^*\, D_{j\rho}^{\chi} + D_{a\rho}^{\chi}$ also holds (definition 6). Finally, the dissimilarities D^{χ} can be represented as Euclidean square distances in a space with at most $\min(|J|, |K|) - 1$ dimensions; this Euclidean embeddability property (definition 4) allows visualization in Factorial Correspondence Analysis.

What are the necessary implications, if any, between the above properties for a *general* dissimilarity D? In a previous publication (Bavaud 2000), we addressed in part the same question for *weights dissimilarities*, i.e. of the form $D_{jj'} = \sum_k G(\rho_k)\, F(w_{jk}, w_{j'k})$, for which necessary and sufficient conditions for aggregation invariance were established. In the present self-contained paper, we focus on *quotients dissimilarities* (definition 2), for which aggregation invariance is automatically satisfied (theorem 1)[2]. As far as dissimilarities

[2] to the best of our knowledge, this elementary but fundamental property has gone so far unnoticed.

are concerned, quotients dissimilarities appear as more natural objects than weights dissimilarities, and easier to manipulate as well.

Somewhat to our surprise, we realized that the Euclidean embeddability condition and the weak Huygens' principle seem to work in *competition* rather than in association. More specifically, we consider here a broad but strict subclass of quotients dissimilarities, namely the product dissimilarities (definition 7), covering all the (aggregation-invariant versions of) dissimilarities we remember having encountered in the literature. For this subclass, we prove (theorem 3, our main result) that Euclidean embeddability and the weak Huygens' principle *are mutually exclusive*, the only exception being provided by the chi-square dissimilarity D^{χ}.

2 Main results

Definition 2 Quotients dissimilarities $D_{jj'}$ between rows $j, j' \in J$ and quotients dissimilarities $D^*_{kk'}$ between columns $k, k' \in K$ are respectively defined as

$$D_{jj'} := \sum_k \rho_k \; F(q_{jk}, q_{j'k}) \qquad\qquad D^*_{kk'} := \sum_j \rho^*_j \; F(q_{jk}, q_{jk'}) \qquad (4)$$

where $F(q,q') \geq 0$ is a non-negative function obeying $F(q,q') = F(q',q)$ and $F(q,q) = 0$, $\forall q, q' \geq 0$.

Remark 1 Definition (4) enables to compute dissimilarities such as D_{ja} where j is one of the original rows and a is a supplementary row, whose quotient profile $\{a_k\}_{k\in K}$ satisfies $a_k \geq 0$ and $\sum_k \rho_k \; a_k = 1$ in virtue of (2). Similarly, $D^*_{k a^*}$ is the dissimilarity between column k and supplementary column a^* of quotient profile $\{a^*_j\}_{j\in J}$ satisfying $a^*_j \geq 0$ and $\sum_j \rho^*_j \; a^*_j = 1$. As typical examples, consider the average row ρ whose associated quotient profile is $a_k = q_{\rho k} = \rho_k/\rho_k = 1$, as well as the average column ρ^* with associated quotient profile $a^*_j = q_{j\rho^*} = 1$.

Definition 3 A dissimilarity D is **aggregation invariant** if its values $D_{jj'}$ remain unchanged when two identical profiles $w^*_{k_1 j} = w^*_{k_2 j}$ (or equivalently $q_{jk_1} = q_{jk_2}$) associated to distinct columns k_1 and k_2 are further aggregated into a single column denoted $[k_1 \cup k_2]$ yielding the same profile $w^*_{[k_1\cup k_2]j}$.

Theorem 1. *Quotients dissimilarities D and D^* are aggregation invariant.*

Proof: D of the form (4) is aggregation invariant iff

$$\rho_{k_1} \; F(q_{jk_1}, q_{j'k_1}) + \rho_{k_2} \; F(q_{jk_2}, q_{j'k_2}) = \rho_{[k_1\cup k_2]} \; F(q_{j[k_1\cup k_2]}, q_{j'[k_1\cup k_2]})$$

for all $j \neq j'$, whenever $q_{jk_1} = q_{jk_2} =: b_j$ for all j. The above identity follows form $\rho_{[k_1\cup k_2]} = \rho_{k_1} + \rho_{k_2}$ and

$$q_{j[k_1\cup k_2]} = \frac{n_{j[k_1\cup k_2]} \; n}{n_{j\bullet} \; n_{\bullet[k_1\cup k_2]}} = \frac{(n_{jk_1} + n_{jk_2}) \; n}{n_{j\bullet} \; (n_{\bullet k_1} + n_{\bullet k_2})} = b_j = q_{jk_1} = q_{jk_2}$$

Aggregation invariance for D^* is proved similarly. □

Definition 4 A dissimilarity D is **Euclidean embeddable** if there exist coordinates x_{jl} for $j \in J$ and $l = 1, \ldots, r < \infty$ such that $D_{jj'} = \sum_{l=1}^{r}(x_{jl} - x_{j'l})^2$. [3]

Definition 5 Let $\{h_j\}_{j\in J}$ be a signed vector (not necessarily non-negative) normed to $\sum_{j\in J} h_j = 1$, and let a be a row of J or a supplementary row whose quotient profile obeys $\sum_k \rho_k\, a_k = 1$.

a)the **"type 1 products matrix"** B^h associated to dissimilarity D and center h is $B^h := -\frac{1}{2}(I - H)D(I - H')$, that is:

$$B^h_{jj'} := -\frac{1}{2}(D_{jj'} - \sum_{l\in J} h_l\, D_{lj'} - \sum_{l\in J} h_l\, D_{jl} + \sum_{l,l'\in J} h_l\, h_{l'}\, D_{ll'})$$

where $H := \mathbf{1}\, h'$, that is $(H)_{ll'} = h_{l'}$.

b) the **"type 2 products matrix"** ${}^a\!B$ associated to dissimilarity D and center a is ${}^a\!B := -\frac{1}{2}(D_{jj'} - D_{ja} - D_{aj'})$.

Note the type 2 products matrix ${}^{j_0}\!B$ associated to the center $j_0 \in J$ to be at same time a type 1 products matrix B^{h^0} with center $h^0_j = \delta_{jj_0}$. In general, the dissimilarities obtain from the "products matrices" as $D_{jj'} = B^h_{jj} + B^h_{j'j'} - 2B^h_{jj'}$, respectively $D_{jj'} = {}^a B_{jj} + {}^a B_{j'j'} - 2\, {}^a B_{jj'}$, and the well-known Schoenberg characterization holds:

Theorem 2. a) *D is Euclidean embeddable iff B^h is positive semi-definite (p.s.d.). In this case, $B^{\tilde h}$ is also p.s.d. for any signed $\tilde h$ obeying $\sum_j \tilde h_j = 1$.*

b)*D is Euclidean embeddable iff ${}^a\!B$ is p.s.d. In this case, ${}^{\tilde a}\!B$ is also p.s.d. for any supplementary row $\tilde a$.*

Proof: see e.g. Schoenberg (1935) or Gower (1982).

Definition 6 Huygens' principles relative to a quotient dissimilarity D are

$$\sum_j \rho^*_j\, D_{ja} = \sum_j \rho^*_j\, D_{j\rho} + D_{\rho a} \qquad \text{strong Huygens' principle} \qquad (5)$$

$$\sum_{jj'} \rho^*_j\, \rho^*_{j'}\, D_{jj'} = 2\sum_j \rho^*_j\, D_{j\rho} \qquad \text{weak Huygens' principle} \qquad (6)$$

where $a_k \geq 0$ is any quotient profile obeying $\sum_k \rho_k\, a_k = 1$ and ρ is the average quotient profile, identically equal to 1 (remark 1).

Writing $a_k = q_{j'k}$ and applying the weighted average operator $\sum_{j'} \rho^*_{j'} \ldots$ shows the strong formulation (5) to entail the weak one (6).

[3] in presence of metric properties, $D_{jj'}$ thus behaves as the *square* of an Euclidean distance.

Definition 7 Product dissimilarities, **f-** and **g-dissimilarities** are quotient dissimilarities (definition 2) defined as

$$\begin{array}{ll} D_{jj'} = \sum_{k\in K} \rho_k \, (g(q_{jk}) - g(q_{j'k}))(h(q_{jk}) - h(q_{j'k})) & \text{(product dissimilarity)} \\ D_{jj'} = \sum_{k\in K} \rho_k \, (f(q_{jk}) - f(q_{j'k}))^2 & (f\text{-dissimilarity}) \\ D_{jj'} = \sum_{k\in K} \rho_k \, (g(q_{jk}) - g(q_{j'k}))(q_{jk} - q_{j'k}) & (g\text{-dissimilarity}) \end{array} \quad (7)$$

where f, g and h are non-decreasing functions obeying $f(1) = g(1) = h(1) = 1$. When differentiable, we impose the normalization $f'(1) = g'(1) = h'(1) = 1$ to hold in addition.

Theorem 3. *The product dissimilarity (7)*

- *is Euclidean embeddable (definition 4) iff $g = h =: f$, that is iff the product dissimilarity is a f-dissimilarity.*
- *obeys Huygens' weak principle (6) iff $g(q) = q$ or $h(q) = q$, that is iff the product dissimilarity is a g-dissimilarity.*
- *is Euclidean embeddable and satisfies Huygens' weak principle iff $g(q) = h(q) = q$, that is iff the product dissimilarity is the chi-square dissimilarity D^{χ}. Also, D^{χ} is the only product dissimilarity obeying the strong Huygens' principle.*

Proof : the first part of the third point is a direct consequence of the first and second points. To prove the first point, observe (definitions 4, 7 and remark 1) that, if D is a product dissimilarity, then

$$^{\rho}B_{jj'} = \frac{1}{2} \sum_k \rho_k \, [(g(q_{jk}) - 1) \, (h(q_{j'k}) - 1) + (g(q_{j'k}) - 1) \, (h(q_{jk}) - 1)]$$

However, $^{\rho}B_{jj'}$ is p.s.d. for any quotient profile iff it is of the form $B_{jj';\rho} = \sum_k \rho_k \, G(q_{jk}) \, G(q_{j'k})$, that is iff $g = h =: f$.

To prove the second point, observe (definition 6) that, if D is a product dissimilarity, then

$$\sum_{jj'} \rho^*_j \, \rho^*_{j'} \, D_{jj'} - 2 \sum_j \rho^*_j \, D_{j\rho} = 2 \sum_k \rho_k \, (\hat{g}_k - 1) \, (\hat{h}_k - 1)$$

where $\hat{g}_k := \sum_j \rho^*_j \, g(q_{jk})$ and $\hat{h}_k := \sum_j \rho^*_j \, h(q_{jk})$. Then the weak's Huygens principle holds iff $\hat{g}_k = 1$ or $\hat{h}_k = 1$ for any quotient profiles; but the only rule to be followed by all profiles is (2), which shows (with $h(1) = g(1) = 1$) that $g(q) = q$ or $h(q) = q$, or equivalently that D is a g-dissimilarity.

To prove the second part of the third point, suppose the product dissimilarity D to obey Huygens' strong principle, and therefore the weak principle as well. One finds Huygens' strong principle to hold iff $\sum_{jk} \rho_k \rho^*_j g(q_{jk})(g(a_k) - a_k) = 0$ for any quotient profiles q_{jk} and a_k, which implies $g(a_k) = a_k$, that is $D = D^{\chi}$. □

3 Examples and comments

a) Product dissimilarities do not exhaust the class of "interesting" quotient dissimilarities: as a counter-example, consider for instance the "symmetrized generalized power quotient dissimilarities"

$$D^{\lambda}_{jj'} := \frac{1}{2\lambda(\lambda+1)} \sum_k \rho_k \; (q_{jk} \; [(\frac{q_{jk}}{q_{j'k}})^{\lambda} - 1] + q_{j'k} \; [(\frac{q_{j'k}}{q_{jk}})^{\lambda} - 1]) \qquad \text{for } \lambda \text{ real}$$

whose name refers to similar functionals studied by Cressie and Read (1984) in the context of model selection.

b) For $s > 0$, g-dissimilarities with $g(q) := (q^s + s - 1)/s$ constitute the so-called "type s" dissimilarities, identified in Bavaud (2000) as an aggregation-invariant class of *weight* dissimilarities obeying Huygens's weak principle. Theorem 3 permits to show that this class is unique. The particular cases $g(q) = q$ (chi-square dissimilarity), $g(q) = \ln q + 1$ (logarithmic dissimilarity) and $g(q) = 2 - 1/q$ (ratio dissimilarity) obtain for $s = 1, 0, -1$ respectively.

c) Let us list some f-dissimilarities with promising properties: they are flexible, intuitively transparent regarding the non-linear distortion of departures of the independence, and thus potentially well-adapted and finely tunable for particular needs:

1. the *power dissimilarities* with $f(q) := (q^{\beta} + \beta - 1)/\beta$, restoring the chi-square dissimilarity for $\beta = 1$. The limit $\beta \to 0$ yields the logarithm $f(q) = \ln q + 1$. The case $\beta = 1/2$ gives the so-called Hellinger distance, whose aggregation invariance properties have been noticed and investigated by Escofier (1978). The limit of large positive β produces "caricature" dissimilarities in that object j tends to be characterized exclusively by its dominant feature k_0 (such that $q_{jk_0} \geq q_{jk}$). Inversely, low positive values of β overweight the effect of rare features.
2. the *presence-absence* dissimilarity with $f(q) := I(q > 0)$, where $I(A)$ is the indicator function for the event A. The resulting dissimilarity is aggregation invariant by construction (theorem 1), and so is the *weighted simple matching similarity*

$$S^{\text{simple, weighted}}_{jj'} := \sum_{k \in K} \rho_k \; [1 - (I(q_{jk} > 0) - I(q_{j'k} > 0))^2]$$

By contrast, the usual (unweighted or uniform) simple matching similarity

$$S^{\text{simple, unweighted}}_{jj'} := \sum_{k \in K} [1 - (I(q_{jk} > 0) - I(q_{j'k} > 0))^2]/|K|$$

(see e.g. Joly and Le Calvé (1994)) is not aggregation invariant. The same remarks apply to the unweighted and weighted dissimilarity of Jaccard defined as

$$S_{jj'}^{\text{Jaccard, unweighted}} = \frac{\text{number of features common to } j \text{ and } j'}{\text{number of features common to } j \text{ or } j'}$$

$$S_{jj'}^{\text{Jaccard, weighted}} := \frac{\sum_{k\in K} \rho_k \; I(q_{jk} > 0) \; I(q_{j'k} > 0)}{\sum_{k\in K} \rho_k \; [1 - I(q_{jk} = 0) \; I(q_{j'k} = 0)]}$$

3. the *major-minor* dissimilarity with $f(q) := I(q \geq 1)$. This dissimilarity only distinguishes whether quotients are above or below average (whence its name). Major-minor and presence-absence dissimilarities are particular cases of the *threshold* dissimilarities $f(q) := I(q \geq \theta) + I(\theta > 1)$ (obeying $f(1) = 1$ for any $\theta > 0$).
4. the *entropic* dissimilarity with $f(q) := 1+\text{sgn}(q-1)\sqrt{2}\sqrt{q(\ln q - 1) + 1}$. Calculus shows $f(q)$ to be increasing with $f(1) = 0$ and $f'(1) = 1$. The resulting central (half-)inertia is

$$\frac{1}{2}\sum_j \rho_j^* \; D_{j\rho} = \frac{1}{2}\sum_{jk} \rho_j^* \; \rho_k \; (f(q_{jk}) - 1)^2 = \sum_{jk} \rho_j^* \; \rho_k \; (q_{jk}[\ln q_{jk} - 1] + 1) =$$
$$\sum_{jk} \rho_j^* \; \rho_k \; q_{jk} \ln q_{jk} = \sum_{jk} \rho_j^* \; w_{jk} \; \ln \frac{w_{jk}}{\rho_k} = \sum_{jk} \rho_j^* \; w_{jk} \; \ln w_{jk} -$$
$$\sum_k \rho_k \; \ln \rho_k = -H(K|J) + H(K) = I(J:K)$$

where $I(J:K) := H(J) + H(K) - H(J,K) \geq 0$ is the mutual information between rows $j \in J$ and columns $k \in K$, null iff J and K are independent, that is iff $q_{jk} \equiv 1$. The non-linear function $f(q)$ thus allows an *exact Euclidean representation for mutual information*, without having to expand the logarithm to the second order: mutual information $H(J) + H(K) - H(J,K)$ can be visualized as a particular instantiation of the central inertia, thus providing a direct link between Data Analysis and Information Theory.

d) The Euclidean embeddability property enjoyed by f-dissimilarities is obvious when considering the transformation $q_{jk} \to x_{jk} := \sqrt{\rho_k}\; f(q_{jk})$, transforming the *profile* $\{q_{jk}\}_{k\in K}$ of row j into *coordinates* $\{x_{jk}\}_{k\in K}$, since $D_{jj'} = \sum_{k\in K}(x_{jk} - x_{j'k})^2$ from definition 7 [4]. In this paper, we have defined the coordinates of the average profile by first averaging over the row profiles and then applying the above transformation. This coincides with the direct averaging of the row coordinates iff $f(q)$ is linear, which is the chi-square case. Distinguishing clearly between those two definitions for average profiles is therefore crucial: in terms of average coordinates (rather than average profiles) f-dissimilarities *do* indeed trivially satisfy the weak Huygens' principle.

[4] or equivalently $D_{jj'} = \sum_{k\in K}(x^c_{jk} - x^c_{j'k})^2$ in the centered version with $x^c_{jk} := \sqrt{\rho_k}\,(f(q_{jk}) - 1))$.

4 Conclusion

The Euclidean embeddability condition is of course essential for validating visualization techniques in Data Analysis. On the other hand, the weak Huygens' principle justifies the definition of a local dissimilarity by splitting the double summation into a summation restricted on pairs satisfying a given relation (such as a contiguity relation in local variance formulations; see e.g. Lebart (1969)) and its complementary. Thus the theoretical results obtained in this paper should invite to reconsider a few practical aspects in Visualization, Classification and Factor Analysis, when dealing with generalized, non chi-square dissimilarities. In particular:

- usual (dis-)similarities indices should be modified into their aggregation-invariant versions (section 3.c.2).
- the distinction between representing clusters by averaging profiles *or* coordinates should be carefully addressed (section 3.d).
- to consider dissimilarities between binary profiles as particular cases of general categorical dissimilarities (sections 3.c.2 and 3.c.3) is more direct than the current practice, which operates the other way round by first dichotomizing categorical variables.

References

BAVAUD, F. (2000): On a class of Aggregation-invariant Dissimilarities obeying the weak Huygens' principle. In H.A.L. Kiers and al. (Eds.): *Data Analysis, Classification and Related Methods.* Springer, New York, 131-136.

CRESSIE, N. and READ, T.R.C. (1984): Multinomial goodness-of-fit tests. *J.R.Statist.Soc.B*, *46*, 440-464.

CRITCHLEY, F. and FICHET, B. (1994): The partial order by inclusion of the classes and dissimilarity on a finite set, and some of their basic properties. In B. Van Custem (Ed.): *Classification and Dissimilarity Analysis.* Lecture Notes in Statistics, Springer, New York, 5-65.

ESCOFIER, B. (1978): Analyse factorielle et distances répondant au principe d'équivalence distributionnelle. *Revue de Statistique Appliquée*, *26*, 29-37

GOWER, J.C. (1982): Euclidean distance geometry. *The Mathematical Scientist*, *7*, 1-14

JOLY, S. and LE CALVE, G. (1994): Similarity functions. In B. Van Custem (Ed.): *Classification and Dissimilarity Analysis.* Lecture Notes in Statistics, Springer, New York, 67-86.

LEBART, L. (1969): L'analyse statistique de la contiguïté. *Publications de l'ISUP*, *XVIII*, 81-112

SCHOENBERG, I.J. (1935): Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espaces vectoriels distancés applicables vectoriellement sur l'espace de Hilbert". *Annals of Mathematics*, *36*, 724-732