IMA Journal of Mathematics Applied in Medicine & Biology (1986) **3**, 115–127

Measures of Niche Overlap, II

R. GERRARD* AND A. D. BARBOUR

Universität Zürich, Institut für Angewandte Mathematik, Rämistrasse 74, CH-8001 Zürich

[Received 20 March 1986 and in revised form 11 July 1986]

An axiomatic basis for the construction of measures of niche overlap is analysed, and its implications are discussed, with particular reference to the more commonly used measures. A method of establishing that a measure of overlap complies with the axioms is put forward, making possible the construction of families of new measures. Certain extremal measures of overlap are also identified.

1. Introduction

MANY different indices are currently being used to quantify the notion of niche overlap (Ricklefs & Lau, 1980). In Schatzmann, Gerrard, and Barbour (1986) [Part I], a number of properties that a reasonable measure of overlap might be expected to possess are proposed in the form of axioms, and their motivations and consequences are discussed from a biological standpoint, with particular reference to the more commonly used measures. In the present paper, the implications of the axioms are explored in a more mathematical framework. Certain results which are only stated in Part I are proved, and simple conditions are established which are sufficient to guarantee that a measure of overlap complies with all the axioms. This, in turn, makes it easy to construct new measures for applications in which existing measures may be inappropriate, and some examples of these are given. It is shown that four of the commonly used measures of overlap, those of Renkonen (1938), Matusita (1955), Horn (as modified in Ricklefs & Lau, 1980), and van Belle & Ahmad (1974), satisfy all the axioms.

This discussion is conducted in sufficient generality that its validity is not necessarily limited to biological applications. The current work is concerned essentially with comparison of distributions; other applications which might be considered include comparisions of political allegiance in various social classes, or quantification of the substitutability of two goods in an economy.

2. The axioms and their effects

The setting in Part I consists of a region $\mathscr{A}$ endowed with a finite measure θ and partitioned into homogeneous θ-measurable sub-regions $\mathscr{A}_1, \ldots, \mathscr{A}_J$, over which the distributions of two species of flora or fauna are to be compared. θ can, for instance, be taken to represent the distribution of resources over $\mathscr{A}$.

* Now at the Statistical Laboratory, The City University, London EC1V 0HB.

Abundances L_j and M_j of the two species on the sub-regions $\mathscr{A}_j$ $(1 \leq j \leq J)$ of $\mathscr{A}$ are converted into number densities with respect to θ by setting $\lambda_j = L_j/\theta(\mathscr{A}_j)$ and $\mu_i = M_i/\theta(\mathscr{A}_i)$, and finally into probability densities $f(x)$ and $g(x)$ with respect to the probability distribution $\nu = \theta/\theta(\mathscr{A})$ by writing $f(x) = \lambda_j/\ell$ and $g(x) = \mu_j/m$ whenever $x \in \mathscr{A}_j$, where

$$\ell = \sum_j \lambda_j \theta(\mathscr{A}_j)/\theta(\mathscr{A}), \qquad m = \sum_j \mu_j \theta(\mathscr{A}_j)/\theta(\mathscr{A}).$$

Thus

$$\int_{\mathscr{A}} f(x)\nu(\mathrm{d}x) = \int_{\mathscr{A}} g(x)\nu(\mathrm{d}x) = 1.$$

The problem is to find reasonable measures of the overlap between the distributions of the two species. It is, in particular, argued that overlap is a concept naturally associated with relative rather than absolute abundances, and hence that reasonable measures of overlap should be functions of f, g, and θ alone (Part I, Axiom 1); if f, g, and θ remain fixed, changing the overall average population densities ℓ and m should have no effect on overlap.

In the present paper we start essentially from this point, taking an arbitrary measurable space $(\mathscr{A}, \Sigma)$, on which is defined a finite measure θ and two probability distributions F and G which are absolutely continuous with respect to θ. The densities of F and G with respect to $\nu = \theta/\theta(\mathscr{A})$ are denoted by f and g respectively, and we wish to find measures of the similarity or overlap between F and G which are functions $o = o(f, g, \theta)$. Let $\tilde{\nu}$ denote the probability distribution on $\mathbb{R}^+ \times \mathbb{R}^+$ induced from ν by the mapping from $\mathscr{A}$ into $\mathbb{R}^+ \times \mathbb{R}^+$ defined by $x \mapsto (f(x), g(x))$; note that

$$\int u\tilde{\nu}(\mathrm{d}\xi) = \int v\tilde{\nu}(\mathrm{d}\xi) = 1, \tag{2.1}$$

where $\xi = (u, v) = (f(x), g(x))$.

Axiom 2 of Part I requires that, if a homogeneous sub-region $\mathscr{A}_j$ is instead considered as two sub-regions $\mathscr{A}_{j1}$ and $\mathscr{A}_{j2}$, overlap should remain unchanged, and Axiom 3 that, if investigation of two distinct regions $\mathscr{A}$ and $\mathscr{A}'$ yields relative distributions of identical pattern, the measure of overlap derived from the combined information should be the same as that from either region alone. These axioms have the effect of restricting attention to functions $\tilde{o}$ of $\tilde{\nu}$ alone. The effect of this is to prevent the structure and extent of the space $\mathscr{A}$ from influencing the degree of overlap, in that multiplication of θ by a scalar leaves ν unaffected, and only the ν-measure of sets of the form

$$\{x \in \mathscr{A} : (f(x), g(x)) \in \mathscr{B}\}$$

enters the calculation of the measure of overlap; the geometrical structure of such sets, or of $\mathscr{A}$, is thus unimportant.

In fact, the choice of functions $\tilde{o}$ is narrowed still further, and only functions of the form

$$C_2:\ \tilde{\sigma}(\tilde{v}) = \int_{\mathbb{R}^+\times\mathbb{R}^+} \phi(u, v)\tilde{v}(d\xi) \tag{2.2}$$

are considered, where $\phi : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is some fixed symmetric function satisfying

$$\sup_{u,v\in\mathbb{R}^+} \phi(u, v)/(1 + u + v) < \infty, \tag{2.3}$$

the condition (2.3) ensuring that the expectation (2.2) exists for all $\tilde{v}$, in view of (2.1). This is because, in the setting of Part I, such measures are expressible in the form

$$\sum_{j=1}^{J} \theta(\mathcal{A}_j)\phi(f_j, g_j)/\theta(\mathcal{A}).$$

This enables $\phi(f_j, g_j)$ to be interpreted as the degree of overlap at any point in the sub-region $\mathcal{A}_j$, and the term $\theta(\mathcal{A}_j)\phi(f_j, g_j)/\theta(\mathcal{A})$ can then be viewed as the contribution to overlap from the sub-region $\mathcal{A}_j$, weighted according to the measure θ. Other possible types of function $\tilde{\sigma}$ are less intuitively appealing.

Axiom 4 in Part I states that, if a given region $\mathcal{A}'$ is adjoined to $\mathcal{A}$, where $F(\mathcal{A}') = G(\mathcal{A}') = 0$ and $\theta(\mathcal{A}') = \eta\theta(\mathcal{A})$ for some $\eta > 0$, then overlap as measured over $\mathcal{A} \cup \mathcal{A}'$ should be the same as that measured over $\mathcal{A}$ alone: areas empty of both species should have no effect on the measure of overlap. Let $\tilde{v}$ denote the probability distribution induced on $\mathbb{R}^+ \times \mathbb{R}^+$ from the measure $\theta/\theta(\mathcal{A})$ over $\mathcal{A}$, and $\tilde{v}_\eta$ the corresponding distribution induced from $\theta/\theta(\mathcal{A} \cup \mathcal{A}')$ over $\mathcal{A} \cup \mathcal{A}'$: that is,

$$\tilde{v}_\eta(\{(0, 0)\}) = [\eta + \tilde{v}(\{(0, 0)\})]/(1 + \eta),$$

$$\tilde{v}_\eta(\mathcal{B}) = \tilde{v}((1 + \eta)\mathcal{B})/(1 + \eta), \qquad (0, 0) \notin \mathcal{B} \in \mathfrak{B}(\mathbb{R}^+ \times \mathbb{R}^+),$$

where $\lambda\mathcal{B}$ denotes $\{(\lambda x, \lambda y) : (x, y) \in \mathcal{B}\}$. Then Axiom 4 is equivalently expressed by requiring that the function $\tilde{\sigma}$ must satisfy $\tilde{\sigma}(\tilde{v}) = \tilde{\sigma}(\tilde{v}_\eta)$ for all $\tilde{v}$ and for all $\eta \geq 0$. In terms of C_2 functions, this means that

$$\int_{\mathbb{R}^+\times\mathbb{R}^+} \phi(u, v)\tilde{v}(d\xi) = (1 + \eta)^{-1}\left(\eta\phi(0, 0) + \int_{\mathbb{R}^+\times\mathbb{R}^+} \phi(u(1 + \eta), v(1 + \eta))\tilde{v}(d\xi)\right) \tag{2.4}$$

for all $\eta \geq 0$ and for all $\tilde{v}$. The effect of this axiom is stated as a proposition.

PROPOSITION 2.1 *A function $\tilde{\sigma}(\tilde{v})$ of the form C_2 satisfies* (2.4) *if and only if*

$$\phi(u, v) = \tfrac{1}{2}(u + v)\psi\left(\frac{|u - v|}{u + v}\right) \quad (u,v \geq 0) \tag{2.5}$$

for some bounded $\psi : [0, 1] \to \mathbb{R}$.

Proof. Let $\tilde{\sigma}$ be of the form C_2 such that (2.4) is satisfied. Take $\tilde{v}$ to be a two-point distribution concentrating mass $\tfrac{1}{2}$ on $(2\alpha, 2 - 2\alpha)$ and $\tfrac{1}{2}$ on $(2 -$

$2\alpha, 2\alpha)$. Then

$$\tilde{\sigma}(\tilde{\nu}) = \phi(2\alpha, 2-2\alpha),$$

whereas

$$\tilde{\sigma}(\tilde{\nu}_\eta) = (1+\eta)^{-1}\big[\eta\phi(0,0) + \phi\big(2\alpha(1-\eta), 2(1-\alpha)(1+\eta)\big)\big].$$

Since these two expressions must be equal, by (2.4), set

$$\phi(u, v) = \phi(0,0) + \tfrac{1}{2}(u+v)\chi(u, v)$$

to get, for each $\zeta = (1-\eta)^{-1} \geqslant 1$,

$$\chi(2\alpha, 2-2\alpha) = \chi\big(2\zeta\alpha, 2\zeta(1-\alpha)\big).$$

Thus there is a function ψ such that

$$\chi(u, v) = \psi\left(\frac{|u-v|}{u+v}\right) \quad (u+v \geqslant 2). \tag{2.6}$$

Now repeat the argument, but this time with $\tilde{\nu}$ as the two-point distribution assigning mass $\frac{1}{2}$ to each of $(2\alpha, 2\beta)$ and $(2-2\alpha, 2-2\beta)$, where $0 < \alpha+\beta < 1$, thus obtaining

$$\begin{aligned}(\alpha+\beta)\chi(2\alpha, 2\beta) &+ (2-\alpha-\beta)\chi(2-2\alpha, 2-2\beta)\\ &= (\alpha+\beta)\chi(2\zeta\alpha, 2\zeta\beta) + (2-\alpha-\beta)\chi\big(2\zeta(1-\alpha), 2\zeta(1-\beta)\big),\end{aligned}$$

for all $\zeta \geqslant 1$. By taking $\zeta = (\alpha+\beta)^{-1}$ and using (2.6), we obtain

$$\chi(u, v) = \psi\left(\frac{|u-v|}{u+v}\right)$$

for all $u, v \geqslant 0$ such that $0 < u+v < 2$; the boundedness of ψ follows from (2.3). $\phi(0,0)$ may be taken to be zero since, in view of (2.1), its effect can be reproduced by taking $\psi(\cdot) + \phi(0,0)$ in place of $\psi(\cdot)$.

The converse is immediate. □

Axiom 4 thus gives rise to a new form of overlap index,

$$C_3\colon\ \tilde{\sigma}(\tilde{\nu}) = \int_{\mathbf{R}^+\times\mathbf{R}^+} \tfrac{1}{2}(u+v)\psi\left(\frac{|u-v|}{u+v}\right)\tilde{\nu}(d\xi), \tag{2.7}$$

or, in terms of f, g, and θ,

$$\sigma(f, g, \theta) = \theta(\mathscr{A})^{-1}\int_{\mathscr{A}} \tfrac{1}{2}(f+g)\psi\left(\frac{|f-g|}{f+g}\right)\theta(dx). \tag{2.8}$$

This may also be rewritten as

$$\int_{\mathscr{A}} \psi\left(\left|\frac{d\big(\frac{1}{2}(F-G)\big)}{d\big(\frac{1}{2}(F+G)\big)}\right|\right) d\big(\tfrac{1}{2}(F+G)\big), \tag{2.9}$$

exhibiting that C_3 measures of overlap are in fact intrinsic measures, independent of θ, being functions of the probability distribution on $[0,1]$ induced from $\frac{1}{2}(F+G)$ by the map $x \mapsto |d\frac{1}{2}(F-G)/d\frac{1}{2}(F+G)|(x)$.

In Part I the class C_3 of measures of overlap is restricted in three further ways. First, it is assumed that ψ is a decreasing function satisfying $\psi(0)=1$, $\psi(1)=0$, and $0<\psi(x)<1$ for $0<x<1$. That ψ should be decreasing follows because, given two pairs (u_1, v_1) and (u_2, v_2) with $u_1+v_1=u_2+v_2$, the contribution to the overlap (2.7) should be greater from the pair with the smaller value of $|u-v|$. The remaining assumptions on ψ characterize perfect overlap as $o(f, g, \theta)=1$ and no overlap as $o(f, g, \theta)=0$. Secondly, it is required that any such function o should be simply related to a distance, in that there should exist a continuous decreasing function

$$\tau:[0,1]\rightarrow[0,1]$$

such that $\tau(1)=0$, $\tau(0)=1$ and $\tau(o(f, g, \theta))$ is a metric on the space of distributions absolutely continuous with respect to θ. This latter requirement is designed to make the concept of overlap interpretable, in the natural sense that distance and overlap, or dissimilarity and similarity, should be complementary concepts. Clearly, the function $\tau(x)=1-x$ would be the best, but it is shown in the next section that this places too strong a constraint on the class of possible measures. The function $\tau(x)=(1-x)^{\frac{1}{2}}$, for instance, allows much greater freedom of choice: we shall restrict ourselves to considering functions τ from the family $(1-x)^{1/p}$, $p\geq 1$. Lastly, there are biological reasons for supposing that the function u defined by

$$u(w)=[1-\psi(w)]/w \tag{2.10}$$

should be non-decreasing. Since this assumption is only relevant in what follows to Corollary 3.2, we refer discussion of it to Part I. Note that $u(1)=1$, because $\psi(1)=0$.

These considerations suggest that reasonable measures of overlap should have the form

$$C_4\text{: } o(f, g, \theta)=1-\theta(\mathscr{A})^{-1}\int_{\mathscr{A}}\tfrac{1}{2}|f-g|\,u\left(\frac{|f-g|}{f+g}\right)\theta(\mathrm{d}x), \tag{2.11}$$

where $u:[0,1]\rightarrow[0,1]$ is non-decreasing with $u(1)=1$, and where $(1-o)^{1/p}$ is a metric for some $p\geq 1$. Of those mentioned in the literature, the following four are of the required form:

1. The Renkonen index (Renkonen, 1938), $o=\int\min\{f, g\}\nu(\mathrm{d}x)$, corresponding to $u\equiv 1$.

2. The Matusita index (Matusita, 1955), $o=\int(fg)^{\frac{1}{2}}\nu(\mathrm{d}x)$, corresponding to $u(w)=w^{-1}[1-(1-w^2)^{\frac{1}{2}}]$.

3. van Belle and Ahmad's index (van Belle and Ahmad, 1974),

$$o=2\int[fg/(f+g)]\nu(\mathrm{d}x),\text{ corresponding to } u(w)=w.$$

4. Modified form of Horn's index (Ricklefs and Lau, 1980),

$$o=(2\log 2)^{-1}\int[(f+g)\log(f+g)-f\log f-g\log g]\nu(\mathrm{d}x),$$

so that

$$u(w) = (2w \log 2)^{-1}[(1+w)\log(1+w) + (1-w)\log(1-w)].$$

In addition to these, the following family of indices, which are convex combinations of (1) and (3) above, is proposed in Part I:

5. [SGB] $o = \int [\alpha \min\{f, g\} + 2(1-\alpha)fg/(f+g)]\nu(dx)$, giving $u(w) = \alpha + (1-\alpha)w$ (with $0 < \alpha \leq 1$).

It is shown in the next section that $(1 - o)$ is a metric for the Renkonen index, and that $(1 - o)^{\frac{1}{2}}$ is a metric for each of the others.

3. Construction of C_4 measures of overlap

In this section, we address the problem of determining for which functions u the quantity $(1 - o)^{1/p}$ is a metric for some $p \geq 1$, where o is given by (2.11). This is equivalent to asking when

$$d(F, G) = \left[\int \tfrac{1}{2}\,|f-g|\, u\left(\frac{|f-g|}{f+g}\right)\nu(dx)\right]^{1/p} \tag{3.1}$$

is a metric on the space of measures absolutely continuous with respect to ν, where

$$f = \frac{dF}{d\nu}, \qquad g = \frac{dG}{d\nu},$$

and u is non-decreasing on $[0, 1]$, with $u(1) = 1$.

Two general observations may be made in connection with (3.1). The first is that, if (3.1) is a metric for a given value $p = p_0$, then it is also a metric for any $p > p_0$, as a concave function of a metric is still a metric. The second observation concerns combining two metrics, or rather two u functions. If d_0 and d_1 are metrics of the form (3.1) with corresponding functions u_0 and u_1 then, for any $\alpha \in (0, 1)$, the choice of $u_\alpha = \alpha u_1 + (1-\alpha)u_0$ in (3.1) gives rise to another metric d_α. This is because

$$d_\alpha(F, G) = [\alpha d_0^p(F, G) + (1-\alpha)d_1^p(F, G)]^{1/p} = \|d.(F, G)\|_p,$$

say, where $\|f(\cdot)\|_p$ denotes the L_p-norm of f with respect to the measure putting mass $1 - \alpha$ on zero and α on one. Thus

$$\begin{aligned} d_\alpha(F, G) + d_\alpha(F, H) &= \|d.(F, G)\|_p + \|d.(F, H)\|_p \\ &\geq \|d.(F, G) + d.(F, H)\|_p \geq \|d.(G, H)\|_p = d_\alpha(G, H), \end{aligned}$$

where the second inequality holds because d_i is a metric for $i = 0, 1$. This result clearly extends to general convex combinations.

An illustration of the application of these observations is exhibited in Example 3.5, where a function u_0 corresponding to a $p = 1$ metric and a function u_1 corresponding to a $p = 2$ metric are combined to generate a family of functions u_α corresponding to $p = 2$ metrics.

If

$$\delta(a, b) = \left[\tfrac{1}{2}\,|a-b|\, u\left(\frac{|a-b|}{a+b}\right)\right]^{1/p} \quad (a, b \in \mathbb{R}^+) \tag{3.2}$$

is a metric on $\mathbb{R}^+$, then d is automatically a metric, since it is symmetric, takes the value zero when $F = G$, and satisfies

$$\begin{aligned} d(F, G) + d(F, H) &= \|\delta(f(\cdot), g(\cdot))\|_p + \|\delta(f(\cdot), h(\cdot))\|_p \\ &\geq \|\delta(f(\cdot), g(\cdot)) + \delta(f(\cdot), h(\cdot))\|_p \\ &\geq \|\delta(g(\cdot), h(\cdot))\|_p \\ &= d(G, H), \end{aligned}$$

where $\|\cdot\|_p$ now denotes the $L_p(\nu)$-norm.

At least in the case $p = 1$, the converse is also true.

PROPOSITION 3.1 *For $p = 1$, d is a metric if and only if δ is a metric.*

Proof. If $a = 0$, it is trivial that $\delta(a, b) + \delta(a, c) \geq \delta(b, c)$ for all $b,c \geq 0$. If $a > 0$, take any $b,c \geq 0$ and partition $\mathscr{A}$ into seven subsets $\mathscr{A}_1, \ldots, \mathscr{A}_7$. Define densities f, g, h on $\mathscr{A}$ as follows:

i	1	2	3	4	5	6	7
$\nu(\mathscr{A}_i)$	α	α	β	β	γ	γ	$1 - 2(\alpha + \beta + \gamma)$
$f\|_{\mathscr{A}_i}$	0	0	λa	λa	λa	λa	0
$g\|_{\mathscr{A}_i}$	λa	0	λa	0	λb	λc	0
$h\|_{\mathscr{A}_i}$	0	λa	0	λa	λc	λb	0

Here λ is some positive constant, and, since ν is a probability distribution, $\alpha \geq 0$, $\beta \geq 0$, $\gamma \geq 0$, and $2(\alpha + \beta + \gamma) \leq 1$. Since f, g, and h are densities with respect to ν, we must choose α, β, γ, and λ so that

$$2\lambda a(\beta + \gamma) = 1 \quad \text{and} \quad \lambda a(\alpha + \beta) + \lambda\gamma(b + c) = 1. \tag{3.3}$$

Provided that (3.3) can be satisfied whatever the values of a, b and c, the triangle inequality for d ensures that

$$\begin{aligned} d(f, g) + d(f, h) &= 2\left[\alpha\lambda a + \beta\lambda a + \gamma\lambda\,|a - b|\,u\left(\frac{|a - b|}{a + b}\right) + \gamma\lambda\,|a - c|\,u\left(\frac{|a - c|}{a + c}\right)\right] \\ &\geq 2\alpha\lambda a + 2\beta\lambda a + 2\gamma\lambda\,|b - c|\,u\left(\frac{|b - c|}{b + c}\right) = d(g, h), \end{aligned}$$

which reduces to $\delta(a, b) + \delta(a, c) \geq \delta(b, c)$, proving that δ is a metric.

Now, to satisfy (3.3), we need

$$\gamma \geq 0, \qquad 0 \leq \beta = \frac{1}{2\lambda a} - \gamma, \qquad 0 \leq \alpha = \frac{1}{2\lambda a} - \gamma\left(\frac{b + c - a}{a}\right),$$

$$1 \geq 2(\alpha + \beta + \gamma) = \frac{2}{\lambda a} - 2\gamma\left(\frac{b + c - a}{a}\right),$$

which can be re-expressed as

$$0 \leq \gamma \leq \frac{1}{2\lambda a}, \qquad (b + c - a)\gamma \leq \frac{1}{2\lambda}, \qquad \gamma(a - b - c) \leq \tfrac{1}{2}a - \frac{1}{\lambda}. \tag{3.4}$$

Picking $\lambda > 2/a$ allows us to choose γ small enough to satisfy all three conditions. □

From this we derive

COROLLARY 3.2 *The unique metric d of the form* (3.1) *for* $p = 1$ *is given by* $\delta(a, b) = |a - b|$, *and corresponds to the Renkonen index.*

Proof. The triangle inequality (3.4) gives us, for $b = 0$, that

$$a + |a - c|\, u\left(\frac{|a - c|}{a + c}\right) \geq c.$$

Taking $c > a$, this implies that $u(|a - c|/(a + c)) \geq 1$. But u is non-decreasing and $u(1) = 1$, so the only possibility is $u \equiv 1$.

Thus the Renkonen index is the only C_4 measure of overlap for which $(1 - o)$ is a metric. However, if $p > 1$ there are many C_4 measures of overlap for which $(1 - o)^{1/p}$ is a metric. The following theorem presents an extra condition on the function u which is sufficient to ensure that δ, and hence d, is a metric, for a given value of $p > 1$. The functions u corresponding to the five indices mentioned in Section 2 all satisfy the condition, as is shown subsequently.

THEOREM 3.3 *If*

$$\eta(w) = \frac{[2wu(w)]^{1/p}}{(1 + w)^{1/p} - (1 - w)^{1/p}}$$

is a non-increasing function on $[0, 1]$, *then* δ *is a metric.*

Proof. We need to verify the triangle inequality for δ; that is,

$$|a - b|^{1/p} u^{1/p}\left(\frac{|a - b|}{a + b}\right) + |a - c|^{1/p} u^{1/p}\left(\frac{|a - c|}{a + c}\right) \geq |b - c|^{1/p} u^{1/p}\left(\frac{|b - c|}{b + c}\right), \quad (3.5)$$

or, equivalently,

$$|a^{1/p} - b^{1/p}|\, \eta\left(\frac{|a - b|}{a + b}\right) + |a^{1/p} - c^{1/p}|\, \eta\left(\frac{|a - c|}{a + c}\right) \geq |b^{1/p} - c^{1/p}|\, \eta\left(\frac{|b - c|}{b + c}\right). \quad (3.6)$$

The proof is divided into two groups of cases. First, suppose that $a \geq b \geq c$. Then

$$\frac{a - c}{a + c} \geq \frac{b - c}{b + c},$$

so that, since u is non-decreasing,

$$|a - c|^{1/p} u^{1/p}\left(\frac{|a - c|}{a + c}\right) \geq |b - c|^{1/p} u^{1/p}\left(\frac{|b - c|}{b + c}\right).$$

The same holds for $c \geq b \geq a$ and, by the symmetry of b and c in (3.6), also for $a \geq c \geq b$ and $b \geq c \geq a$.

Now suppose that $b \geq a \geq c$. In this case

$$\frac{|b - c|}{b + c} \geq \max\left\{\frac{|a - b|}{a + b}, \frac{|a - c|}{a + c}\right\},$$

and we use the fact that η is non-increasing to show that the LHS of (3.6) is at least as large as

$$(|a^{1/p} - b^{1/p}| + |a^{1/p} - c^{1/p}|)\eta\left(\frac{|b-c|}{b+c}\right),$$

and (3.6) is established by the triangle inequality for $|\cdot|$. The case $c \geq a \geq b$ is treated symmetrically. □

To illustrate the uses of Theorem 3.3, we now construct a number of examples of indices, based on choosing functions u for which the condition of the theorem holds. The first five are familiar from Section 2, those following are presented here for the first time.

EXAMPLE 3.1 (Renkonen index) $p = 1$, $u(w) = 1$, $\eta(w) = 1$. Since u is non-decreasing and η is non-increasing, the Renkonen index is a C_4 measure of overlap with $p = 1$.

EXAMPLE 3.2 (Matusita index) $p = 2$, $u(w) = w^{-1}[1-(1-w^2)^{\frac{1}{2}}]$; $\eta(w) = 1$. The substitution $w = \sin\theta$ yields $u(w) = \tan\frac{1}{2}\theta$, showing that u is an increasing function of w for $0 \leq w \leq 1$, and hence that the Matusita index is a C_4 measure of overlap, with $p = 2$.

EXAMPLE 3.3 (Modified Horn index) $p = 2$,

$$u(w) = \frac{(1+w)\log(1+w) + (1-w)\log(1-w)}{2w\log 2};$$

$$\eta(w) = \frac{[(1+w)\log(1+w) + (1-w)\log(1-w)]^{\frac{1}{2}}}{(\log 2)^{\frac{1}{2}}[(1+w)^{\frac{1}{2}} - (1-w)^{\frac{1}{2}}]}.$$

To check that u is increasing and η decreasing, we introduce the substitution $w = \tanh\alpha = (1-e^{-2\alpha})/(1+e^{-2\alpha})$ $(\alpha \geq 0)$. Then

$$\begin{aligned} u &= \frac{(1+\tanh\alpha)\log(1+\tanh\alpha) + (1-\tanh\alpha)\log(1-\tanh\alpha)}{(2\log 2)\tanh\alpha} \\ &= (\log 2)^{-1}(\alpha - \log\cosh\alpha/\tanh\alpha), \end{aligned}$$

so that

$$\frac{du}{dw} = \left(\frac{dw}{d\alpha}\right)^{-1}\frac{du}{d\alpha} = \frac{\cosh^2\alpha\log\cosh\alpha}{(\log 2)\sinh^2\alpha} \geq 0.$$

Now $\eta(\tanh\alpha) = c\ (\alpha\sinh\alpha - \cosh\alpha\log\cosh\alpha)^{\frac{1}{2}}/\sinh\frac{1}{2}\alpha$, c being a positive constant, so $\eta^2 = 2c^2(\alpha\sinh\alpha - \cosh\alpha\log\cosh\alpha)/(\cosh\alpha - 1)$. Since η is non-negative, it suffices to show that η^2 is decreasing, and

$$\frac{d}{d\alpha}(\eta^2) = K(\alpha)\{\log\cosh\alpha - \alpha\tanh\tfrac{1}{2}\alpha\},$$

where K is a non-negative function of $\alpha \geq 0$. The function in braces takes the value zero at $\alpha = 0$, and has non-positive derivative: hence the derivative of η^2 is non-positive, and η is therefore non-increasing, as required.

EXAMPLE 3.4 (van Belle and Ahmad's index) $p=2$ and $u(w)=w$;

$$\eta(w)=\frac{2^{\frac{1}{2}}w}{(1+w)^{\frac{1}{2}}-(1-w)^{\frac{1}{2}}}.$$

u is increasing, and η can be rewritten as $2^{-\frac{1}{2}}[(1+w)^{\frac{1}{2}}+(1-w)^{\frac{1}{2}}]$, which has derivative $2^{-\frac{3}{2}}[(1+w)^{-\frac{1}{2}}-(1-w)^{-\frac{1}{2}}]\leqslant 0$: thus, η is non-increasing.

EXAMPLE 3.5 [Part I] $p=2$ and $u(w)=\alpha+(1-\alpha)w$ $(0<\alpha\leqslant 1)$; this generates a metric for each $0\leqslant\alpha\leqslant 1$, as observed following (3.1). It is not necessary to take $p=2$ in all cases; the larger α is, the smaller is the minimal value of p for which η is decreasing. The condition $\alpha=1$ corresponds to the Renkonen index, $\alpha=0$ to van Belle and Ahmad's index.

EXAMPLE 3.6.

$$p\geqslant 1,\qquad u(w)=\frac{[(1+w)^{(1+\beta)/p}-(1-w)^{(1+\beta)/p}]^p}{2w(1+w)^{\beta}},\qquad 0\leqslant\beta\leqslant p-1;$$

$$\eta(w)=\frac{1-[(1-w)/(1+w)]^{(1+\beta)/p}}{1-[(1-w)/(1+w)]^{1/p}}.$$

This time we use the substitution $v=(1-w)/(1+w)$, giving

$$u(w)=u_0(v)=\frac{(1-v^{(1+\beta)/p})^p}{1-v},\qquad \eta(w)=\eta_0(v)=\frac{1-v^{(1+\beta)/p}}{1-v^{1/p}}:$$

since v is a decreasing function of w, we require that u_0 should not increase and η_0 not decrease with v.

Consider the function $(1-x^b)^a/(1-x)$ $(0\leqslant x<1)$. Its derivative is

$$\frac{(1-x^b)^{a-1}}{(1-x)^2}[1-abx^{b-1}+(ab-1)x^b].$$

This is non-negative for all $0\leqslant x<1$ if $b\geqslant 1\geqslant a>0$ and is non-positive for all $0\leqslant x<1$ if $0<b\leqslant 1\leqslant a$. Writing $x=v$, $a=p$, and $b=(1+\beta)/p$ shows that u_0 is non-increasing; writing $x=v^{1/p}$, $a=1$, and $b=1+\beta$ shows that η_0 is non-decreasing.

Note that $p=1$ and $\beta=0$ yields the Renkonen index, while $p=2$ and $\beta=0$ yields the Matusita index. Note also that, if $\beta>p-1$, then $u(w)>1$ for w sufficiently close to 1.

EXAMPLE 3.7 $p\geqslant 2$;

$$u(w)=\frac{[(1+w)^{2/p}-(1-w)^{2/p}]^p}{2w[(1+w)^{2/p}+(1-w)^{2/p}]^{p/2}};\qquad \eta(w)=\frac{(1+w)^{1/p}+(1-w)^{1/p}}{[(1+w)^{2/p}+(1-w)^{2/p}]^{1/2}}.$$

The substitution $w=\tanh\alpha$ gives $\eta^2=2\cosh^2(\alpha/p)/\cosh(2\alpha/p)=1+\operatorname{sech}(2\alpha/p)$, which is decreasing, and $u(\tanh\alpha)=c\sinh^p(2\alpha/p)/\sinh\alpha\cosh^{p/2}(2\alpha/p)$, where c is a positive constant. Now

$$\frac{d}{d\alpha}\log u(\tanh\alpha)=2\coth\frac{2\alpha}{p}-\coth\alpha-\tanh\frac{2\alpha}{p};\tag{3.7}$$

but $\coth(2\alpha/p) > 1 > \tanh(2\alpha/p)$, and $p \geq 2$ implies that $\coth(2\alpha/p) \geq \coth \alpha$. Hence $\log u$, and therefore u, is increasing.

If $p < 2$, then $u(w) > 1$ for all w sufficiently close to 1.

Note that $p = 2$ yields van Belle and Ahmad's index.

The various examples listed above offer a wide range of possible measures of overlap. The functions u for Examples 3.1–3.4 and that with $\alpha = 0{\cdot}2$ in Example 3.5 are depicted in Fig. 1 of Part I, and some members of the family of Example 3.7 are shown in Fig. 1. In the family of Example 3.6,

$$u(z) \asymp z^{p-1} \quad \text{as } z \to 0, \qquad 1 - u(z) \asymp (1-z)^{(1+\beta)/p} \quad \text{as } z \to 1,$$

with $u(z) \sim 1 - \frac{1}{2}(p-1)(1-z)$ as $z \to 1$ when $\beta = p - 1$. Thus many possible combinations of the behaviour of u near 0 and 1 can be achieved within this family. The choice $\beta = 0$ depicted in Figure 2 is particularly interesting, in that it always gives rise to an extremal measure of overlap, in the following sense.

THEOREM 3.4 *If δ is a metric, then, for all $w \in [0, 1]$,*

$$u(w) \geq [(1+w)^{1/p} - (1-w)^{1/p}]^p / 2w$$

or, equivalently, $\eta(w) \geq 1$.

Proof. Write $a = \lambda e^{2\alpha p}$, $b = \lambda e^{2(\alpha+\beta)p}$, $c = \lambda$ $(\lambda, \alpha, \beta \geq 0)$, $\hat{\eta}(\gamma) = \eta(\tanh p\gamma)$ $(\gamma > 0)$. The triangle inequality for δ, as expressed in (3.6), reduces to

$$e^{2\alpha}(e^{2\beta} - 1)\hat{\eta}(\beta) + (e^{2\alpha} - 1)\hat{\eta}(\alpha) \geq (e^{2(\alpha+\beta)} - 1)\hat{\eta}(\alpha + \beta).$$

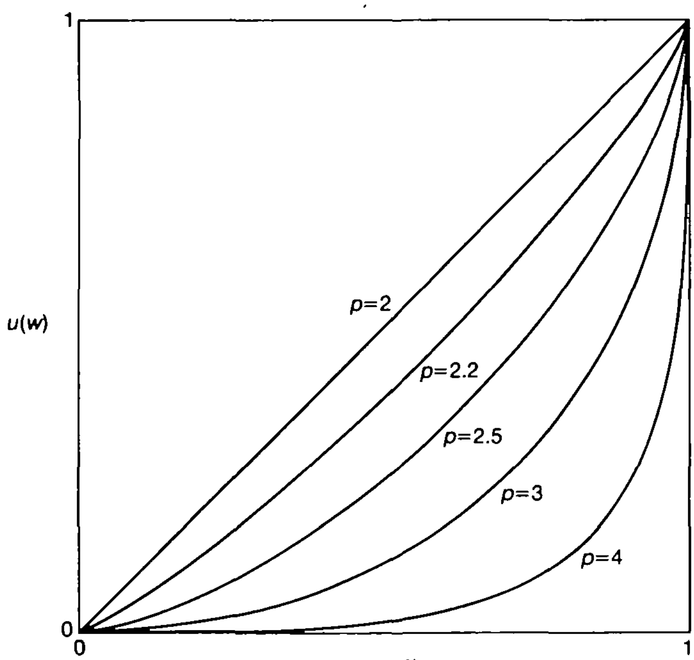

FIG. 1. The functions $u(w)$ for the C_4 indices of Example 3.7 with values of p as shown.

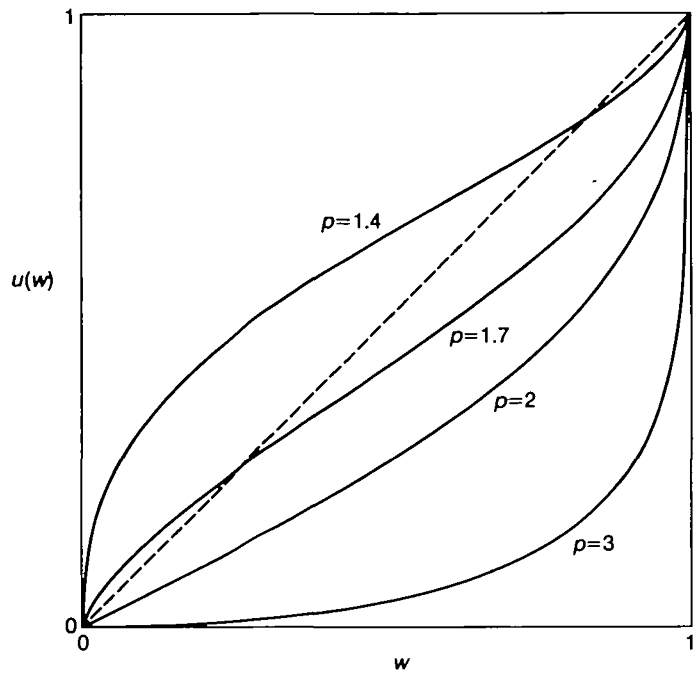

FIG. 2. The functions $u(w)$ for the C_4 indices of Example 3.6 with $\beta = 0$ and values of p as shown.

$\hat{\eta}(\alpha+\beta)$ is thus less than or equal to a convex combination of $\hat{\eta}(\alpha)$ and $\hat{\eta}(\beta)$. It follows by induction that $\hat{\eta}(n\alpha) \leq \hat{\eta}(\alpha)$ for all positive integers n and all $\alpha > 0$, and therefore that

$$\inf_{0<w\leq 1} \eta(w) = \inf_{\gamma>0} \hat{\eta}(\gamma) = \liminf_{\gamma\to\infty} \hat{\eta}(\gamma) = \liminf_{w\to 1} \eta(w) = \liminf_{w\to 1} u(w)^{1/p}.$$

It is now enough to prove that u is left-continuous at 1.

Using the form (3.5) of the triangle inequality for δ, with $c = 0$, $b = 1$ and $a < 1$, we obtain

$$(1-a)u\left(\frac{1-a}{1+a}\right) \geq (1-a^{1/p})^p.$$

Thus, letting $a \to 0$ and remembering that $u(w) \leq 1$, it follows that $\lim_{w\to 1} u(w) = 1$. □

Thus, taking $p = 2$ and $\beta = 0$, the Matusita index has everywhere smaller values of $u(w)$ than any other C_4 measure of overlap for $p = 2$. The Renkonen index, $p = 1$ and $\beta = 0$, has already been shown in Corollary 3.1 to be the unique C_4 measure for $p = 1$. Of course, the condition $u(w) \leq 1$ and the remark following (3.1) ensure that the Renkonen index is the C_4 measure with everywhere greatest values of $u(w)$, for all $p \geq 1$.

References

MATUSITA, K. 1955 Decision rules based on the distance, for problems of fit, two samples, and estimation. *Ann. Math. Stat.* **26,** 631–640.

RENKONEN, O. 1938 Statistisch ökologische Untersuchungen über die terrestrische Käferwelt der finnischen Bruchmoore. *Ann. Zool. Soc. Zool.-Bot. Fenn. Vanamo* **6,** 1–231.

RICKLEFS, R. E., & LAU, M. 1980 Bias and dispersion of overlap indices: results of some Monte-Carlo simulations. *Ecology* **61,** 1019–1024.

SCHATZMANN, E., GERRARD, R., & BARBOUR, A. D. 1986 Measures of niche overlap, I. *IMA J. Math. Appl. Med. Biol.* **3,** 99–113 (this issue).

VAN BELLE, G. & AHMAD, I. 1974 Measuring affinity of distributions. Pp 651–668 in Proschan, F. & Serfling, R. J. (Eds), *Reliability and Biometry.* SIAM Publications, Philadelphia, Pennsylvania.