ORDER-DISTANCE AND OTHER METRIC-LIKE FUNCTIONS ON JOINTLY DISTRIBUTED RANDOM VARIABLES

EHTIBAR N. DZHAFAROV AND JANNE V. KUJALA

Abstract. We construct a class of real-valued nonnegative binary functions on a set of jointly distributed random variables. These functions satisfy the triangle inequality and vanish at identical arguments (pseudo-quasi-metrics). We apply these functions to the problem of selective probabilistic causality encountered in behavioral sciences and in quantum physics. The problem reduces to that of ascertaining the existence of a joint distribution for a set of variables with known distributions of certain subsets of this set. Any violation of the triangle inequality by one of our functions when applied to such a set rules out the existence of the joint distribution. We focus on an especially versatile and widely applicable class of pseudo-quasi-metrics called order-distances. We show, in particular, that the Bell-CHSH-Fine inequalities of quantum physics follow from the triangle inequalities for appropriately defined order-distances.

1. Introduction

We show how certain metric-like functions on jointly distributed random variables (*pseudo-quasi-metrics* introduced in Section 2) can be used in dealing with the problem of selective probabilistic causality (introduced in Section 3), illustrating this on examples taken from behavioral sciences and quantum physics (Section 4). Although most of Section 3 applies to arbitrary pseudo-quasi-metrics on jointly distributed random variables, we single out one, termed *order-distance*, which is especially useful due to its versatility. We discuss examples of other pseudo-quasi-metrics and rules for their construction in Section 5.

2. Order p.q.-metrics

Random variables in this paper are understood in the broadest sense as measurable functions $X : V_s \to V$, with no restrictions being imposed on the sample spaces (V_s, Σ_s, μ_s) and the induced probability spaces (V, Σ, μ), with the usual meaning of the terms (sets of values V_s, V, sigma-algebras Σ_s, Σ, and probability measures μ_s, μ). In particular, any set X of jointly distributed random variables (functions on the same sample space) is a random variable, and its induced probability space (or, simply, *distribution*) $\overline{X} = (V, \Sigma, \mu)$ is referred to as the joint distribution of its elements.

2000 *Mathematics Subject Classification.* Primary 60B99, Secondary 81Q99, 91E45.

Key words and phrases. Bell-CHSH-Fine inequalities, Einstein-Podolsky-Rosen paradigm, probabilistic causality, pseudo-quasi-metrics on random variables, quantum entanglement, selective influences.

First author's work was supported by AFOSR grant FA9550-09-1-0252. Second author's work was supported by Academy of Finland grant 121855.

Given a class of random variables $\mathscr{X}$, not necessarily jointly distributed, let $\mathscr{X}^*$ be the class of distributions $\overline{X}$ for all $X \in \mathscr{X}$. For any class function $f^* : \mathscr{X}^* \to \mathbb{R}$ (reals), the function $f : \mathscr{X} \to \mathbb{R}$ defined by $f(X) = f^*\left(\overline{X}\right)$ is called *observable* (as it does not depend on sample spaces, typically unobservable). We will conveniently confuse f and f^* for observable functions, so that if f is defined on $\mathscr{X}$, then $f(Y)$, identified with $f^*\left(\overline{Y}\right)$, is also defined for any $Y \notin \mathscr{X}$ with $\overline{Y} \in \mathscr{X}^*$. (This convention is used in Section 3, when we apply a function defined on a set of random variables H to different but identically distributed sets of A-variables.)

For an arbitrary nonempty set Ω, let $H = \{H_\omega : \omega \in \Omega\}$ be a indexed set of jointly distributed random variables H_ω with distributions $\overline{H}_\omega = (V_\omega, \Sigma_\omega, \mu_\omega)$. For any $\alpha, \beta \in \Omega$, the ordered pair (H_α, H_β) is a random variable with distribution $(V_\alpha \times V_\beta, \Sigma_\alpha \times \Sigma_\beta, \mu_{\alpha,\beta})$, and $H \times H$ is a set of jointly distributed random variables (hence also a random variable).

Definition 2.1. We call an observable function $d : H \times H \to \mathbb{R}$ a pseudo-quasi-metric (*p.q.-metric*) on H if, for all $\alpha, \beta, \gamma \in \Omega$,

(i) $d(H_\alpha, H_\beta) \geq 0$,
(ii) $d(H_\alpha, H_\alpha) = 0$,
(iii) $d(H_\alpha, H_\gamma) \leq d(H_\alpha, H_\beta) + d(H_\beta, H_\gamma)$.

For terminological clarity, the conventional pseudometrics (also called semimetrics) obtain by adding the property $d(H_\alpha, H_\beta) = d(H_\beta, H_\alpha)$; the conventional quasimetrics are obtained by adding the property $\alpha \neq \beta \Rightarrow d(H_\alpha, H_\beta) > 0$. A conventional metric is both a pseudometric and a quasimetric. (See, e.g., [27] for a discussion of a variety of metrics and pseudometrics on random variables.)

By obvious argument we can generalize the triangle inequality (iii): for any $H_{\alpha_1}, \ldots, H_{\alpha_l} \in H$ $(l \geq 3)$,

$$d(H_{\alpha_1}, H_{\alpha_l}) \leq \sum_{i=2}^{l} d\left(H_{\alpha_{i-1}}, H_{\alpha_i}\right). \tag{2.1}$$

We refer to this inequality (which plays a central role in this paper) as the *chain inequality*.

Let

$$R \subset \bigcup_{(\alpha,\beta)\in\Omega\times\Omega} V_\alpha \times V_\beta,$$

and write $a \preceq b$ to designate $(a, b) \in R$. Let R be a total order, that is, transitive, reflexive, and connected in the sense that for any $(a, b) \in \bigcup_{(\alpha,\beta)\in\Omega\times\Omega} V_\alpha \times V_\beta$, at least one of the relations $a \preceq b$ and $b \preceq a$ holds. We define the equivalence $a \sim b$ and strict order $a \prec b$ induced by $\preceq$ in the usual way. Finally, we assume that for any $(\alpha, \beta) \in \Omega \times \Omega$, the sets

$$\{(a, b) : a \in V_\alpha, b \in V_\beta, a \preceq b\}$$

are $\mu_{\alpha,\beta}$-measurable. This implies the $\mu_{\alpha,\beta}$-measurability of the sets

$$\{(a, b) : a \in V_\alpha, b \in V_\beta, a \prec b\}, \ \{(a, b) : a \in V_\alpha, b \in V_\beta, a \sim b\}.$$

Thus, if all V_ω are intervals of reals, $\preceq$ can be chosen to coincide with $\leq$, and (assuming the usual Borel sigma algebra) all the properties above are satisfied. Another example: for arbitrary V_ω, provided each Σ_ω contains at least $n > 1$ disjoint nonempty sets, one can partition V_ω as $\bigcup_{k=1}^{n} V_\omega^{(k)}$, with $V_\omega^{(k)} \in \Sigma_\omega$, and

put $a \preceq b$ if and only if $a \in V_\alpha^{(k)}, b \in V_\beta^{(l)}$ and $k \leq l$. Again, all properties above are clearly satisfied.

Definition 2.2. The function

$$\mathrm{D}\left(H_\alpha, H_\beta\right) = \Pr\left[H_\alpha \prec H_\beta\right] = \int_{a \prec b} \mathrm{d}\mu_{\alpha,\beta}\left(a, b\right)$$

is called an *order p.q.-metric*, or *order-distance*, on H.

That the definition is well-constructed follows from

Theorem 2.3. *Order-distance* D *is a p.q.-metric on* H.

Proof. Let $\alpha, \beta, \gamma \in \Omega$, $H_\alpha = A$, $H_\beta = B$, and $H_\gamma = X$. That $\mathrm{D}\left(A, B\right)$ is determined by the distribution of (A, B) is obvious from the definition. The properties $\mathrm{D}\left(A, B\right) \geq 0$ and $\mathrm{D}\left(A, A\right) = 0$ are obvious too. To prove the triangle inequality,

$$\begin{aligned}\mathrm{D}\left(A, B\right) = \Pr\left[A \prec B\right] = \Pr\left[A \prec B \prec X\right] + \Pr\left[A \prec B \sim X\right] \\ + \Pr\left[A \prec X \prec B\right] + \Pr\left[A \sim X \prec B\right] + \Pr\left[X \prec A \prec B\right],\end{aligned}$$

$$\begin{aligned}\mathrm{D}\left(A, X\right) = \Pr\left[A \prec X\right] = \Pr\left[A \prec X \prec B\right] + \Pr\left[A \prec B \sim X\right] \\ + \Pr\left[A \prec B \prec X\right] + \Pr\left[A \sim B \prec X\right] + \Pr\left[B \prec A \prec X\right],\end{aligned}$$

$$\begin{aligned}\mathrm{D}\left(X, B\right) = \Pr\left[X \prec B\right] = \Pr\left[X \prec B \prec A\right] + \Pr\left[X \prec A \sim B\right] \\ + \Pr\left[X \prec A \prec B\right] + \Pr\left[A \sim X \prec B\right] + \Pr\left[A \prec X \prec B\right].\end{aligned}$$

So

$$\begin{aligned}\mathrm{D}\left(A, X\right) + \mathrm{D}\left(X, B\right) - \mathrm{D}\left(A, B\right) = \Pr\left[B \prec A \prec X\right] + \Pr\left[A \sim B \prec X\right] \\ + \Pr\left[X \prec B \prec A\right] + \Pr\left[X \prec A \sim B\right] + \Pr\left[A \prec X \prec B\right] \geq 0.\end{aligned}$$

□

Since in the last expression all events are pairwise exclusive, we have

$$\mathrm{D}\left(A, X\right) + \mathrm{D}\left(X, B\right) - \mathrm{D}\left(A, B\right) \leq 1.$$

This may seem an attractive addition to the triangle inequality. The inequality is redundant, however, as it is subsumed by the triangle inequalities holding on $\{A, B, X\}$. Rewriting it as

$$\mathrm{D}\left(A, B\right) + 1 - \mathrm{D}\left(X, B\right) - \mathrm{D}\left(A, X\right) \geq 0,$$

this inequality immediately follows from

$$\mathrm{D}\left(A, B\right) + \mathrm{D}\left(B, X\right) - \mathrm{D}\left(A, X\right) \geq 0$$

and

$$\mathrm{D}\left(B, X\right) = \Pr\left[B \prec X\right] \leq 1 - \Pr\left[X \prec B\right] = 1 - \mathrm{D}\left(X, B\right).$$

3. Selective probabilistic causality

Consider an indexed set $W = \{W^\lambda : \lambda \in \Lambda\}$, with each W^λ being a set referred to as a (deterministic) *input*, with the elements of $\{\lambda\} \times W^\lambda$ called *input points*. Input points therefore are pairs of the form $x = (\lambda, w)$, with $w \in W^\lambda$, and should not be confused with *input values* w. A nonempty set $\Phi \subset \prod_{\lambda\in\Lambda} W^\lambda$ is called a set of (allowable) *treatments*. A treatment therefore is a function $\phi : \Lambda \to \bigcup_{\lambda\in\Lambda} W^\lambda$ such that $\phi(\lambda) \in W^\lambda$ for any $\lambda \in \Lambda$. Note that the symbol ϕ not followed by an argument always refers to the entire function, the set $\{(\lambda, \phi(\lambda) : \lambda \in \Lambda)\}$.

In the following we use two kinds of random variables: those indexed as A^λ_ϕ (each corresponding to a fixed index $\lambda \in \Lambda$ and a fixed function ϕ) and those indexed as H^λ_w (with $w \in W^\lambda$), corresponding to input points (λ, w).

Let there be a collection of sets of random variables, referred to as (random) *outputs*,

$$A_\phi = \left\{A^\lambda_\phi : \lambda \in \Lambda\right\},\ \phi \in \Phi,$$

such that the distribution of A_ϕ (i.e., the joint distribution of all A^λ_ϕ in A_ϕ) is known for every treatment ϕ. We define

$$A^\lambda = \left\{A^\lambda_\phi : \phi \in \Phi\right\},\ \lambda \in \Lambda,$$

with the understanding that A^λ *is not* a random variable (i.e., A^λ_ϕ for different ϕ are not jointly distributed). To illustrate the notation, let $\Lambda = \{1, 2, \ldots\}$ and W^λ be the set of reals for all $\lambda \in \Lambda$. A treatment ϕ then is a real-valued function (sequence) $\{(1, \phi(1)), (2, \phi(2)), \ldots\} = (\phi(1), \phi(2), \ldots)$, where $\phi(1) \in W^1, \phi(2) \in W^2$, etc. Let Φ be a nonempty set of such sequences. Fixing one of them, $\phi = (w_1, w_2, \ldots)$,

$$A_\phi = A_{(w_1,w_2,\ldots)} = \left\{A^1_{(w_1,w_2,\ldots)}, A^2_{(w_1,w_2,\ldots)}, \ldots\right\};$$

fixing, say, $\lambda = 2$ and allowing $(w_1, w_2, \ldots)$ range over Φ,

$$A^\lambda = A^2 = \left\{A^2_{(w_1,w_2,\ldots)} : (w_1, w_2, \ldots) \in \Phi\right\}.$$

The following problem is encountered in a wide variety of contexts [6, 7, 15]. We say that the dependence of random outputs A^λ_ϕ on the deterministic inputs W^λ is (canonically) *selective* if, for any distinct $\lambda, \lambda' \in \Lambda$ and any $\phi \in \Phi$, the output A^λ_ϕ is "not influenced" by $\phi(\lambda')$. The question is how one should define this selectivity of "influences" rigorously and how one can determine whether this selectivity holds. This problem was introduced to behavioral sciences by Sternberg [18] and Townsend [22]. In quantum physics, using different terminology, it was introduced by Bell [3] and elaborated by Fine [10, 11]. The definition can be given in several equivalent forms, of which we present the one focal for the present context.

Definition 3.1. The dependence of outputs $\left\{A^\lambda : \lambda \in \Lambda\right\}$ on inputs $\left\{W^\lambda : \lambda \in \Lambda\right\}$ (or the "influence" of the latter on the former) is (canonically) selective if there is a set of jointly distributed random variables

$$H = \left\{H^\lambda_w : w \in W^\lambda, \lambda \in \Lambda\right\}$$

(one random variable for every value of every input) such that, for any treatment $\phi \in \Phi$,

$$\overline{H}_\phi = \overline{A}_\phi,$$

where

$$H_\phi = \left\{ H^\lambda_{\phi(\lambda)} : \lambda \in \Lambda \right\}$$

and

$$A_\phi = \left\{ A^\lambda_\phi : \lambda \in \Lambda \right\}$$

(the corresponding elements of H_ϕ and A_ϕ being those sharing the same λ).

This definition is known as the *Joint Distribution Criterion* (JDC) for selectivity of influences, and the set H satisfying this definition is referred to as a (hypothetical) JDC-set. Specialized forms of this criterion in quantum physics can be found in [19] and [10, 11]; in the behavioral context and in complete generality this criterion is given (derived from an equivalent definition) in [8].

Remark 3.2. The adjective "canonical" in the definition refers to the one-to-one correspondence between W^λ and A^λ sharing the same λ. A seemingly more general scheme in which different A^λ are selectively influenced by different (possibly overlapping) subsets of $\left\{ W^\lambda : \lambda \in \Lambda \right\}$ is always reducible to the canonical form by considering, for every A^λ, the Cartesian product of the inputs influencing it to be a single input and redefining correspondingly the sets of input points and the set of allowable treatments.

The simplest consequence of JDC is that the selectivity of influences implies *marginal selectivity* [6, 24], defined as follows. For any $\Lambda' \subset \Lambda$ we can uniquely present any $\phi \in \Phi$ as $\phi' \cup \overline{\phi'}$, where $\phi' \in \prod_{\lambda \in \Lambda'} W^\lambda$ and $\overline{\phi'} \in \prod_{\lambda \in \Lambda - \Lambda'} W^\lambda$. Then, if JDC is satisfied, the joint distribution of $\left\{ A^\lambda_{\phi' \cup \overline{\phi'}} : \lambda \in \Lambda' \right\}$ does not depend on $\overline{\phi'}$.

Remark 3.3. In the following we always assume that marginal selectivity is satisfied.

The relevance of the order-distance and other p.q.-metrics on the sets of jointly distributed random variables to the problem of selectivity lies in the general test (necessary condition) for selectivity of influences, formulated after the following definition.

Definition 3.4. We call a sequence of input points

$$x_1 = (\alpha_1, w_1), \ldots, x_l = (\alpha_l, w_l)$$

(where $w_i \in W^{\alpha_i}$ for $i = 1, \ldots, l \geq 3$) *treatment-realizable* if there are treatments $\phi^1, \ldots, \phi^l \in \Phi$ (not necessarily pairwise distinct) such that

$$\{x_1, x_l\} \subset \phi^1 \text{ and } \{x_{i-1}, x_i\} \subset \phi^i \text{ for } i = 2, \ldots, l.$$

If a JDC-set H exists, then for any p.q.-metric d on H we should have

$$d\left(H^{\alpha_1}_{w_1}, H^{\alpha_l}_{w_l}\right) = d\left(A^{\alpha_1}_{\phi^1}, A^{\alpha_l}_{\phi^1}\right)$$

and

$$d\left(H^{\alpha_{i-1}}_{w_{i-1}}, H^{\alpha_i}_{w_i}\right) = d\left(A^{\alpha_{i-1}}_{\phi^i}, A^{\alpha_i}_{\phi^i}\right)$$

for $i = 2, \ldots, l$, whence

$$d\left(A^{\alpha_1}_{\phi^1}, A^{\alpha_l}_{\phi^1}\right) \leq \sum_{i=2}^{l} d\left(A^{\alpha_{i-1}}_{\phi^i}, A^{\alpha_i}_{\phi^i}\right). \tag{3.1}$$

This chain inequality, written entirely in terms of observable probabilities, is referred to as a *p.q.-metric test* for selectivity of influences. If this inequality is violated for at least one treatment-realizable sequence of input points, no JDC-set H exists, and the selectivity is ruled out. Note: if the sequence $\phi^{(1)},\dots,\phi^{(l)}\in\Phi$ for a given $x_1,\dots,x_l$ can be chosen in more than one way, the observable quantities $d\left(A^{\alpha_1}_{\phi^{(1)}},A^{\alpha_l}_{\phi^{(1)}}\right)$ and $d\left(A^{\alpha_{i-1}}_{\phi^{(i-1)}},A^{\alpha_i}_{\phi^{(i)}}\right)$ remain invariant due to the (tacitly assumed) marginal selectivity.

As an example, let $\Lambda=\{1,2\}$, $W^1=[0,1]$, $W^2=[0,1]$, $\Phi=W^1\times W^2$. Let $\left\{A^1_\phi,A^2_\phi\right\}$ for any treatment ϕ have a bivariate normal distribution with zero means, unit variances, and correlation $\rho=\min\left(1,w_1+w_2\right)$, where $w_1=\phi\left(1\right),w_2=\phi\left(2\right)$. Marginal selectivity is trivially satisfied. Do $\left\{W^1,W^2\right\}$ influence $\left\{A^1,A^2\right\}$ selectively? For any bivariate normally distributed (A,B), let us define $A\prec B$ iff $A<0,B\geq 0$. Then the corresponding order-distance on the hypothetical JDC-set H is

$$\mathrm{D}\left(H^1_{w_1},H^2_{w_2}\right)=\frac{\arccos\left(\min\left(1,w_1+w_2\right)\right)}{2\pi}.$$

The sequence of input points $\left(1,0\right),\left(2,1\right),\left(1,1\right),\left(2,0\right)$ is treatment-realizable, so if H exists, we should have

$$\mathrm{D}\left(H^1_0,H^2_0\right)\leq\mathrm{D}\left(H^1_0,H^2_1\right)+\mathrm{D}\left(H^2_1,H^1_1\right)+\mathrm{D}\left(H^1_1,H^2_0\right).$$

The numerical substitutions yield, however,

$$\frac{1}{4}\leq 0+0+0,$$

and as this is false, the hypothesis that $\left\{W^1,W^2\right\}$ influence $\left\{A^1,A^2\right\}$ selectively is rejected.

The theorem below and its corollary show that one only needs to check the chain inequality for a special subset of all possible treatment-realizable sequences $x_1,\dots,x_l$.

Definition 3.5. A treatment-realizable sequence $x_1,\dots,x_l$ is called *irreducible* if $x_1\neq x_l$, and the only subsequences $\{x_{i_1},\dots,x_{i_k}\}$ with $k>1$ that are subsets of treatments are pairs $\{x_1,x_l\}$ and $\{x_{i-1},x_i\}$, for $i=2,\dots,l$. Otherwise the sequence is *reducible*.

Theorem 3.6. *Given a p.q.-metric d on the hypothetical JDC-set H, inequality (3.1) is satisfied for all treatment-realizable sequences if and only if this inequality holds for all irreducible sequences.*

Proof. We prove this theorem by showing that if (3.1) is violated for some reducible sequence $x_1,\dots,x_l$, then it is violated for some proper subsequence thereof. Clearly, $x_1\neq x_l$ because otherwise (3.1) is not violated. For $l=3$, x_1,x_2,x_3 is reducible only if it is contained in a treatment, but then (3.1) would be satisfied. So $l>3$, and the reducibility of $x_1,\dots,x_l$ means that there is a pair $\{x_p,x_q\}$ belonging to a treatment, with $(p,q)\neq(1,l)$ and $q>p+1$. But then (3.1) must be violated for either $x_p,\dots,x_q$ or $x_1,\dots,x_p,x_q,\dots,x_l$ (allowing for $p=1$ or $q=l$ but not both). □

If $\Phi=\prod_{\lambda\in\Lambda}W^\lambda$ (all logically possible treatments are allowable), then any subsequence $x_{i_1},\dots,x_{i_k}$ of input points with pairwise distinct $\alpha_{i_1},\dots,\alpha_{i_k}$ belongs to some treatment. Therefore an irreducible sequence cannot contain points of more

than two inputs, and it is easy to see then that it must be a sequence of pairwise distinct $x_1 \in \{\alpha\} \times W^\alpha, x_2 \in \{\beta\} \times W^\beta, ..., x_{2m-1} \in \{\alpha\} \times W^\alpha, x_{2m} \in \{\beta\} \times W^\beta$ ($\alpha \neq \beta$). It is also easy to see that if $m > 2$, each of the subsets $\{x_1, x_4\}$ and $\{x_2, x_5\}$ will belong to a treatment. Hence $m = 2$ is the only possibility for an irreducible sequence.

Corollary 3.7. *If $\Phi = \prod_{\lambda\in\Lambda} W^\lambda$, then inequality (3.1) is satisfied for all treatment-realizable sequences if and only if this inequality holds for all tetradic sequences of the form x, y, s, t, with $x, s \in \{\alpha\} \times W^\alpha$, $y, t \in \{\beta\} \times W^\beta$, $x \neq s$, $y \neq t$, $\alpha \neq \beta$.*

Remark 3.8. This formulation is given in [8], although there it is unnecessarily confined to metrics of a special kind.

4. An application

The four tables below represent results of an experiment with a 2×2 factorial design, $\{x, x'\} \times \{y, y'\}$, and two binary responses, A and B. In relation to our general notation, here we have $\Lambda = \{1, 2\}$, $W^1 = \{x, x'\}$, $W^2 = \{y, y'\}$, and four treatments $(x, y), \ldots, (x', y')$. For every treatment ϕ, the random outputs A^1_ϕ and A^2_ϕ are represented by, respectively, A_ϕ and B_ϕ, each having two possible values, arbitrarily labeled. This design is arguably the simplest possible, and it is ubiquitous in science. In a psychological double-detection experiment (see, e.g., [23]), the input values may represent presence (x and y) or absence (x' and y') of a designated signal in two stimuli labeled 1 and 2, presented side-by-side. The participant in such an experiment is asked to indicate whether the signal was present or absent in stimulus 1 and in stimulus 2. The output values $A = \circ$ and $B = \sqcap$ may indicate either that the response was "signal present" or that the response was correct, and analogously for $A = \bullet$ and $B = \sqcup$ (either "signal absent" or an incorrect response). The entries p_{ij}, q_{ij}, etc., represent joint probabilities of the corresponding outcomes and $a_{i\cdot}, a'_{i\cdot}$, etc., represent marginal probabilities. The question to be answered is: does the response to a given stimulus (A to 1 and B to 2) selectively depend on that stimulus alone (despite A and B being stochastically dependent for every treatment), or is A or B influenced by both 1 and 2?

$\phi = (x, y)$	$B_{xy} = \sqcup$	$B_{xy} = \sqcap$		$\phi = (x, y')$	$B_{xy'} = \sqcup$	$B_{xy'} = \sqcap$	
$A_{xy} = \bullet$	p_{11}	p_{12}	$a_{1\cdot}$	$A_{xy'} = \bullet$	q_{11}	q_{12}	$a_{1\cdot}$
$A_{xy} = \circ$	p_{21}	p_{22}	$a_{2\cdot}$	$A_{xy'} = \circ$	q_{21}	q_{22}	$a_{2\cdot}$
	$b_{\cdot 1}$	$b_{\cdot 2}$			$b'_{\cdot 1}$	$b'_{\cdot 2}$	

$\phi = (x', y)$	$B_{x'y} = \sqcup$	$B_{x'y} = \sqcap$		$\phi = (x', y')$	$B_{x'y'} = \sqcup$	$B_{x'y'} = \sqcap$	
$A_{x'y} = \bullet$	r_{11}	r_{12}	$a'_{1\cdot}$	$A_{x'y'} = \bullet$	s_{11}	s_{12}	$a'_{1\cdot}$
$A_{x'y} = \circ$	r_{21}	r_{22}	$a'_{2\cdot}$	$A_{x'y'} = \circ$	s_{21}	s_{22}	$a'_{2\cdot}$
	$b_{\cdot 1}$	$b_{\cdot 2}$			$b'_{\cdot 1}$	$b'_{\cdot 2}$	

Another important situation in which we encounter formally the same problem is the Einstein-Podolsky-Rosen (EPR) paradigm. Two particles are emitted from a common source in such a way that they remain *entangled* (have highly correlated properties, such as momenta or spins) as they run away from each other [1, 16]. An experiment may consist, e.g., in measuring the spin of electron 1 along one of two axes, x or x', and (in another location but simultaneously in some inertial frame of reference) measuring the spin of electron 2 along one of two axes, y or

y'. The outcome A of a measurement on electron 1 is a random variable with two possible values, "up" or "down", and the same holds for B, the outcome of a measurement on electron 2. The question here is: do the measurements on electrons 1 and 2 selectively affect, respectively, A and B (even though generally A and B are not independent at any of the four combinations of spin axes)? If the answer is negative, then the measurement of one electron affects the outcome of the measurement of another electron even though no signal can be exchanged between two distant events that are simultaneous in some frame of reference. What makes this situation formally identical to the double-detection example described above is that the measurements performed along different axes on the same particle, x and x' or y and y', are *noncommuting*, i.e., they cannot be performed simultaneously. This makes it possible to consider such measurements as mutually exclusive values of an input.

Theorem 4.1. *(Fine* [10, 11]*) A JDC-set* $H = \left\{H_x^1, H_{x'}^1, H_y^2, H_{y'}^2\right\}$ *satisfying*

$$\overline{\left\{H_x^1, H_y^2\right\}} = \overline{\{A_{xy}, B_{xy}\}}, \quad \overline{\left\{H_x^1, H_{y'}^2\right\}} = \overline{\{A_{xy'}, B_{xy'}\}},$$

$$\overline{\left\{H_{x'}^1, H_y^2\right\}} = \overline{\{A_{x'y}, B_{x'y}\}}, \quad \overline{\left\{H_{x'}^1, H_{y'}^2\right\}} = \overline{\{A_{x'y'}, B_{x'y'}\}}$$

exists if and only if the following eight inequalities are satisfied:

$$\begin{array}{l} -1 \leq p_{11} + r_{11} + s_{11} - q_{11} - a'_{1\cdot} - b_{\cdot 1} \leq 0, \\ -1 \leq q_{11} + s_{11} + r_{11} - p_{11} - a'_{1\cdot} - b'_{\cdot 1} \leq 0, \\ -1 \leq r_{11} + p_{11} + q_{11} - s_{11} - a_{1\cdot} - b_{\cdot 1} \leq 0, \\ -1 \leq s_{11} + q_{11} + p_{11} - r_{11} - a_{1\cdot} - b'_{\cdot 1} \leq 0. \end{array} \tag{4.1}$$

We refer to (4.1) as *Bell-CHSH-Fine inequalities*, where CHSH abbreviates Clauser, Horne, Shimony, and Holt [4]. In this work Bell's [3] approach was developed into a special version of (4.1).

Remark 4.2. The proof given in [10, 11] that (4.1) is both necessary and sufficient (under marginal selectivity) for the existence of a JDC-set can be conceptually simplified: the Bell-CHSH-Fine inequalities can be algebraically shown to be the criterion for the existence of a vector Q with 16 probabilities,

$$\Pr\left[H_x^1 = \bullet, H_{x'}^1 = \bullet, H_x^1 = \sqcup, H_x^1 = \sqcup\right], \ldots,$$
$$\Pr\left[H_x^1 = \circ, H_{x'}^1 = \circ, H_x^1 = \sqcap, H_x^1 = \sqcap\right],$$

that sum to one and whose appropriately chosen partial sums yield the eight observable probabilities

$$p_{11}, q_{11}, r_{11}, s_{11}, a_{1\cdot}, b_{\cdot 1}, a'_{1\cdot}, b'_{\cdot 1}$$

(other probabilities being determined due to marginal selectivity). This is a simple linear programming task, and the Bell-CHSH-Fine inequalities can be derived "mechanically" by a facet enumeration algorithm (see [25, 26] and [2]). For extensions of the Bell-CHSH-Fine inequalities to multiple particles, multiple spin axes, and multiple random outputs, see [9] and [17]. For modern accounts of mathematical and interpretational aspects of the entanglement problem in quantum physics, see [12, 13, 14].

The point of interest in the present context is that the Bell-CHSH-Fine inequalities, whose rather obscure structure does not seem to fit their fundamental importance, turn out to be interpretable as the triangle inequalities for appropriately chosen order-distances.

Consider the chain inequalities for the order-distance D_1 obtained by putting $\bullet = \sqcup = 1$, $\circ = \sqcap = 2$ and identifying $\preceq$ with $\leq$:

(4.2)

$$\begin{aligned}
q_{12} = \mathrm{D}_1(H^1_x,H^2_{y'}) &\leq \mathrm{D}_1(H^1_x,H^2_y)+\mathrm{D}_1(H^2_y,H^1_{x'})+\mathrm{D}_1(H^1_{x'},H^2_{y'}) = p_{12}+r_{21}+s_{12},\\
p_{12} = \mathrm{D}_1(H^1_x,H^2_y) &\leq \mathrm{D}_1(H^1_x,H^2_{y'})+\mathrm{D}_1(H^2_{y'},H^1_{x'})+\mathrm{D}_1(H^1_{x'},H^2_y) = q_{12}+s_{21}+r_{12},\\
s_{12} = \mathrm{D}_1(H^1_{x'},H^2_{y'}) &\leq \mathrm{D}_1(H^1_{x'},H^2_y)+\mathrm{D}_1(H^2_y,H^1_x)+\mathrm{D}_1(H^1_x,H^2_{y'}) = r_{12}+p_{21}+q_{12},\\
r_{12} = \mathrm{D}_1(H^1_{x'},H^2_y) &\leq \mathrm{D}_1(H^1_{x'},H^2_{y'})+\mathrm{D}_1(H^2_{y'},H^1_x)+\mathrm{D}_1(H^1_x,H^2_y) = s_{12}+q_{21}+p_{12}.
\end{aligned}$$

Also consider the inequalities for the order-distance D_2 obtained by putting $\bullet = \sqcap = 1$, $\circ = \sqcup = 2$ and identifying $\preceq$ with $\leq$:

(4.3)

$$\begin{aligned}
q_{11} = \mathrm{D}_2(H^1_x,H^2_{y'}) &\leq \mathrm{D}_2(H^1_x,H^2_y)+\mathrm{D}_2(H^2_y,H^1_{x'})+\mathrm{D}_2(H^1_{x'},H^2_{y'}) = p_{11}+r_{22}+s_{11},\\
p_{11} = \mathrm{D}_2(H^1_x,H^2_y) &\leq \mathrm{D}_2(H^1_x,H^2_{y'})+\mathrm{D}_2(H^2_{y'},H^1_{x'})+\mathrm{D}_2(H^1_{x'},H^2_y) = q_{11}+s_{22}+r_{11},\\
s_{11} = \mathrm{D}_2(H^1_{x'},H^2_{y'}) &\leq \mathrm{D}_2(H^1_{x'},H^2_y)+\mathrm{D}_2(H^2_y,H^1_x)+\mathrm{D}_2(H^1_x,H^2_{y'}) = r_{11}+p_{22}+q_{11},\\
r_{11} = \mathrm{D}_2(H^1_{x'},H^2_y) &\leq \mathrm{D}_2(H^1_{x'},H^2_{y'})+\mathrm{D}_2(H^2_{y'},H^1_x)+\mathrm{D}_2(H^1_x,H^2_y) = s_{11}+q_{22}+p_{11}.
\end{aligned}$$

Theorem 4.3. *Each right-hand Bell-CHSH-Fine inequality is equivalent to the corresponding chain inequality in (4.2) for the order-distance* D_1*. Each left-hand Bell-CHSH-Fine inequality is equivalent to the corresponding chain inequality in (4.3) for the order-distance* D_2*.*

Proof. We show the proof for the first of the Bell-CHSH-Fine double-inequalities. The equivalence of

$$p_{11} + r_{11} + s_{11} - q_{11} - a'_{1.} - b_{.1} \leq 0$$

to

$$q_{12} \leq p_{12} + r_{21} + s_{12}$$

is obtained by using the identities

$$\begin{aligned}
q_{12} &= a_{1.} - q_{11},\\
p_{12} &= a_{1.} - p_{11},\\
r_{21} &= b_{.1} - r_{11},\\
s_{12} &= a'_{1.} - s_{11}.
\end{aligned}$$

The equivalence of

$$p_{11} + r_{11} + s_{11} - q_{11} - a'_{1.} - b_{.1} \geq -1$$

to

$$q_{11} \leq p_{11} + r_{22} + s_{11}$$

follows from the identity

$$r_{22} = 1 + r_{11} - a'_{1.} - b_{.1}.$$

□

5. Concluding remarks

The order-distances are versatile and have a broad sphere of applicability because order relations on the domains of any given set of random variables can always be defined in many different ways. If no other structure is available, this can always be done by the partitioning of the domains mentioned in Section 2 and used in the example with bivariate normal distributions in Section 3 as well as for the binary variables of the previous section: $V_\omega = \bigcup_{k=1}^{n} V_\omega^{(k)}$, $V_\omega^{(k)} \in \Sigma_\omega$, $\omega \in \Omega$, putting $a \preceq b$ if and only if $a \in V_\alpha^{(k)}, b \in V_\beta^{(l)}$ and $k \leq l$. Due to its universality and convenience of use, it deserves a special name, *classification distance*.

There are numerous ways of creating new p.q.-metrics from the ones already constructed, including those taken from outside probabilistic context. Thus, if d is a p.q.-metric on a set S, then, for any set H of jointly distributed random variables taking their values in S,

$$D(A,B) = \mathrm{E}\left[d(A,B)\right], \quad A,B \in H,$$

is a p.q.-metric on H. This follows from the fact that expectation E preserves inequalities and equalities identically satisfied for all possible realizations of the arguments. Another example: given any family of p.q.-metrics $\{d_\upsilon : \upsilon \in \Upsilon\}$, their average with respect to a random variable U with a probability measure m,

$$d(A,B) = \int_{\upsilon\in\Upsilon} d_\upsilon(A,B)\,\mathrm{d}m(\upsilon),$$

is a p.q.-metric. As a special case, consider a classification distance with binary partitions: the domain V_ω of every H_ω in H is partitioned into two (measurable) subsets, $W_{\omega,\upsilon}^{(1)}$ and $W_{\omega,\upsilon}^{(2)}$. Making these partitions random, i.e., allowing the index υ to randomly vary in any way whatsoever, we get a new p.q.-metric. In the special case when all random variables in H take their values in the set of real numbers and $W_{\omega,\upsilon}^{(1)}$ is defined by $z \leq \upsilon$ ($z \in V_\omega \subset \mathbb{R}$, $\upsilon \in \mathbb{R}$), the randomization of the partitions reduces to that of the separation point υ. The p.q.-metric then becomes

$$d_S(A,B) = \Pr\left[A \leq U < B\right],$$

where U is some random variable. An additively symmetrized (i.e., pseudometric) version of this p.q.-metric, $d_S(A,B) + d_S(B,A)$, was introduced in [20, 21] under the name "separation (pseudo)metric" and shown to be a conventional metric if U is chosen stochastically independent of all random variables in H.

References

[1] Aspect, A. (1999). Bell's inequality tests : More ideal than ever. *Nature* **398** 189-190.

[2] Basoalto, R.M., & Percival, I.C. (2003). BellTest and CHSH experiments with more than two settings. *J. Phys. A* **36** 7411–7423.

[3] Bell, J. (1964). On the Einstein-Podolsky-Rosen paradox. *Physics* **1** 195-200.

[4] Clauser, J.F. , Horne, M.A. , Shimony, A., & Holt, R.A. (1969). Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23** 880-884.

[5] Cover, T.M. & Thomas, J.A. (1990). *Elements of Information Theory*. New York: Wiley.

[6] Dzhafarov, E.N. (2003). Selective influence through conditional independence. *Psychometrika* **68** 7–26.

[7] Dzhafarov, E.N., & Gluhovsky, I. (2006). Notes on selective influence, probabilistic causality, and probabilistic dimensionality. *J. Math. Psych.* **50** 390–401.

[8] Dzhafarov, E.N., & Kujala, J.V. (2010). The Joint Distribution Criterion and the Distance Tests for selective probabilistic causality. *Frontiers in Quantitative Psychology and Measurement* **1**:151 doi: 10.3389/fpsyg.2010.00151.

[9] Dzhafarov, E.N., & Kujala, J.V. (2010). Selectivity in probabilistic causality: Where psychology runs into quantum physics. *J. Math. Psych.* **56**: 54-63.
[10] Fine, A. (1982). Joint distributions, quantum correlations, and commuting observables. *J. Math. Phys.* **23** 1306-1310.
[11] Fine, A. (1982). Hidden variables, joint probability, and the Bell inequalities. *Phys. Rev. Lett.* **48** 291-295.
[12] Gudder, S. (2010). Finite quantum measure spaces. Amer. Math. Mon., **117** 512-527
[13] Khrennikov, A. (2008). EPR–Bohm experiment and Bell's inequality: Quantum physics meets probability theory. *Theor. Math. Phys.* **157** 1448–1460.
[14] Khrennikov, A. (2009). *Contextual Approach to Quantum Formalism*. Berlin: Springer.
[15] Kujala, J.V., & Dzhafarov, E.N. (2008). Testing for selectivity in the dependence of random variables on external factors. *J. Math. Psych.* **52** 128–144.
[16] Mermin, N.D. (1985). Is the moon there when nobody looks? Reality and the quantum theory. *Physics Today* **38** 38-47.
[17] Peres, A. (1999). All the Bell inequalities. *Found. Phys.* **29** 589-614.
[18] Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), Attention and Performance II. *Acta Psychologica* **30** 276–315. Amsterdam: North Holland
[19] Suppes, P., & Zanotti, M. (1981). When are probabilistic explanations possible? *Synthese* **48** 191-199.
[20] Taylor, M.D. (1984). Separation metrics for real-valued random variables. *Int. J. Math. Math. Sci.* **7** 407-408.
[21] Taylor, M.D. (1985). New metrics for weak convergence of distribution functions. *Stochastica* **9** 5-17.
[22] Townsend, J.T. (1984). Uncovering mental processes with factorial experiments. *J. Math. Psych.* **28** 363–400.
[23] Townsend, J.T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *J. Math. Psych.* **39** 321–359.
[24] Townsend, J.T., & Schweickert, R. (1989). Toward the trichotomy method of reaction times: Laying the foundation of stochastic mental networks. *J. Math. Psych.* **33** 309–327.
[25] Werner, R.F., & Wolf, M.M. (2001). All multipartite Bell correlation inequalities for two dichotomic observables per site. arXiv:quant-ph/0102024v1.
[26] Werner, R.F., & Wolf, M.M. (2001). Bell inequalities and entanglement. arXiv:quant-ph/0107093v2.
[27] Zolotarev, V.M. (1976). Metric distances in spaces of random variables and their distributions. *Mathematics of the USSR-Sbornik* **30** 373-401.