International Journal of Theoretical Physics, Vol. 34, No. 8, 1995

Uncertain Evidence and Nonlocality

D. Costantini[1] and U. Garibaldi[2]

Received October 13, 1994

After the work of Suppes and Zanotti it is clear that the proof of the impossibility of local theories is a probability argument. The notion of locality is essentially a principle of conditional statistical independence which is strictly tied to that of exchangeability. De Finetti's celebrated representation theorem makes the connection clear. The way in which Bell's experiment is performed suggests that the probability function which is more suitable to describe it is not exchangeable, but partially exchangeable. It is known that partially exchangeable probability functions show a nonlocal behavior. Working with these functions, it is possible to make use of observations regarding one stochastic process in order to change the distribution of another process. We enlarge to uncertain evidence a classical probability function we have used in deriving some quantum correlations. By means of this enlargement we give simple examples of a nonlocal probability function.

1. INTRODUCTION

After the work of Suppes and Zanotti it is quite clear that the proof of the impossibility of local theories is a probability argument. More specifically: while "the notion of locality of an objective hidden variable theory is essentially a principle of conditional statistical independence" (Suppes and Zanotti, 1980), the derivation of Bell's inequality rests "on the assumption that there is a hidden variable that renders the spin results conditional independent" (Suppes and Zanotti, 1984). On the other hand, the notion of conditional independence is strictly tied to that of exchangeability. De Finetti's celebrated representation theorem makes the connection clear. For these reasons it is quite natural that the work of Suppes and Zanotti proceeds from an exchangeable probability function. However, the way in which Bell's experiment is performed suggests that the probability function which is most suitable to

[1]Istituto di Statistica, Università di Genova, Genoa, Italy.
[2]CSFBT-CNR, Dipartimento di Fisica, Università di Genova, Genoa, Italy.

0020-7748/95/0800-1321$07.50/0

describe these experimental results is not exchangeable, but partially exchangeable. Moreover, it is known that partially exchangeable probability functions show a nonlocal behavior (Daboni and Wedlin, 1982). This means that working with such functions, it is possible to make use of observations regarding one stochastic process in order to change the distribution of another process. In the present paper, first, we enlarge to uncertain evidence a probability function we have already used in deriving some quantum correlations and, second, using this enlargement, we give simple examples of a nonlocal probability function.

2. FOURFOLD TABLES

It is normally taken for granted that the spin results of two parallel apparatuses *I* and *II* in a Bell experiment assume values of the same observable. More specifically, while the spin measurements of each apparatus can take the value $+1/2$ or $-1/2$, one supposes that the spin measurement of apparatus *I* and that of apparatus *II* are values belonging to the same observable. On the contrary, we are persuaded that measurements performed at *I* and those performed at *II* refer to different observables or, to use statistical jargon, to different variables. These measurements give rise to bivariate data which demand a bivariate analysis. As a consequence, the spin results cannot be described by a univariate distribution, but they are described by a bivariate one. Plainly said, in Bell's experiment we are faced with two variables (observables), i.e., measurement at apparatus *I* and measurement at apparatus *II*, each with two values (eigenvalues), i.e., $+1/2$ and $-1/2$. Hence, once we have supposed the exchangeability of each sequence of measurements, the description of the experiment is given by a 2×2 (fourfold) contingency table whose possible values are described as follows:

II \ *I*	+	−
+	++	−+
−	+−	−−

(I)

The head row, which refers to apparatus *I*, is + for $+1/2$ and − for $-1/2$, and the same holds for the head column, which refers to apparatus *II*. Hence ++ in the entry 11 refers to spin $+1/2$ for both measurements, −+ in the entry 12 refers to spin $-1/2$ for measurement at *I* and spin $+1/2$ for measurement at *II*, and so on.

When the directions of orientation are θ for *I* and ϕ for *II*, the distribution of probabilities, given from quantum mechanics, is described as follows:

$$
\begin{array}{cccc}
II \backslash I & + & - & \\
+ & \frac{1}{2}\sin^2\left(\frac{\theta-\phi}{2}\right) & \frac{1}{2}\cos^2\left(\frac{\theta-\phi}{2}\right) & \frac{1}{2} \\
- & \frac{1}{2}\cos^2\left(\frac{\theta-\phi}{2}\right) & \frac{1}{2}\sin^2\left(\frac{\theta-\phi}{2}\right) & \frac{1}{2} \\
 & \frac{1}{2} & \frac{1}{2} & 1
\end{array}
\qquad \text{(II)}
$$

whose marginal column is

$$\frac{1}{2} = \frac{1}{2}\sin^2\left(\frac{\theta-\phi}{2}\right) + \frac{1}{2}\cos^2\left(\frac{\theta-\phi}{2}\right),$$

$$\frac{1}{2} = \frac{1}{2}\cos^2\left(\frac{\theta-\phi}{2}\right) + \frac{1}{2}\sin^2\left(\frac{\theta-\phi}{2}\right)$$

and whose marginal row is

$$\frac{1}{2} = \frac{1}{2}\sin^2\left(\frac{\theta-\phi}{2}\right) + \frac{1}{2}\cos^2\left(\frac{\theta-\phi}{2}\right)$$

$$\frac{1}{2} = \frac{1}{2}\cos^2\left(\frac{\theta-\phi}{2}\right) + \frac{1}{2}\sin^2\left(\frac{\theta-\phi}{2}\right)$$

Putting $\omega = \theta - \phi$, we have the following results:

For $\omega = 0$

$$\underline{\mathbf{T}}(0) = \frac{1}{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

i.e., the *contragraduation* table.

For $\omega = \pi/2$

$$\underline{\mathbf{T}}\left(\frac{\pi}{2}\right) = \frac{1}{2}\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

i.e., the *independence* table.

For $\omega = \pi$

$$\underline{\mathbf{T}}(\pi) = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

i.e., the *cograduation* table.

At the outset we must say that our work is not intended to derive table (II) in an original way. For this reason we limit ourselves to some considerations which could be useful in a future statistical derivation. First, table (II) is a special type of 2×2 table because its margin distributions are uniform. Second, it is a mixture of the *extreme* $\underline{\mathbf{T}}(0)$ and $\underline{\mathbf{T}}(\pi)$ with *weights* $w_0(\omega) = \cos^2(\omega/2)$ and $w_\pi(\omega) = \sin^2(\omega/2)$. Finally, a nontrivial derivation of table (II) should be based on probability conditions justifying the weights of the mixture. The way in which probability conditions can justify probability allotment will be made clear in the following sections. As a first step in the direction of making these conditions explicit, we describe a statistical method, working on extreme tables of the type just seen, which exhibits nonlocal features.

3. TWO-DIMENSIONAL STOCHASTIC PROCESSES

As we have seen, in order to give a probability description of Bell's experiment we must consider bivariate distributions. Hence, first of all, we give a brief description of a way in which it is possible to build up a predictive bivariate distribution.

Consider a two-dimensional stochastic process $\mathbf{Y} := Y_1, Y_2, \ldots, Y_n, \ldots, Y_n = (I_n, II_n)$ whose subprocesses are $\mathbf{I} := I_1, I_2, \ldots, I_n, \ldots$ and $\mathbf{II} := II_1, II_2, \ldots, II_n, \ldots$. The first subprocess refers to variable I whose values are $1, \ldots, j, \ldots, k$; the second one to variable II whose values are $1, \ldots, i \ldots, h$. The evidence of these processes is given by $\mathbf{Y}^n := (I_1 = j_1, II_1 = i_1), \ldots, (I_n = j_n, II_n = i_n), j_n = 1, \ldots, k, i_n = 1, \ldots, h$, for $\mathbf{Y}$; $\mathbf{I}^n := I_1 = j_1, \ldots, I_n = j_n$ for $\mathbf{I}$; $\mathbf{II}^n := II_1 = i_1, \ldots, II_n = i_n$ for $\mathbf{II}$. The $k \times h$ table corresponding to $\mathbf{Y}^n$ is

II \ I	1	2	⋯	k	
1	n_{11}	n_{12}	⋯	n_{1k}	$n_{1.}$
2	n_{21}	n_{22}	⋯	n_{2k}	$n_{2.}$
⋯	⋯	⋯	⋯	⋯	⋯
h	n_{h1}	n_{h2}	⋯	n_{hk}	$n_{h.}$
	$n_{.1}$	$n_{.2}$	⋯	$n_{.k}$	n

(III)

The vectors corresponding to $\mathbf{I}^n$ and $\mathbf{II}^n$ are, respectively, $\mathbf{n}_{.\mathbf{I}} := (n_{.1}, \ldots, n_{.k})$ and $\mathbf{n}_{\mathbf{II}.} := (n_{1.}, \ldots, n_{h.})$.

The method we are considering is a way to construct predictive bivariate distributions or, what is the same thing, to allot predictive probabilities to each cell of a $k \times h$ table. But before doing this, let us recall some results about predictive univariate distributions, that is, the case of a unique variable.

4. ONE-DIMENSIONAL STOCHASTIC PROCESSES

Given a stochastic process $\mathbf{X} := X_1, X_2, \ldots, X_n, \ldots$ referring to a unique variable whose values are $1, \ldots, d$, we consider the *predictive* probability

$$P\{X_{n+1} = j | \mathbf{X}^n\} \tag{1}$$

where $\mathbf{X}^n := X_1 = j_1, \ldots, X_n = j_n, j_n = 1, \ldots, d$, is the evidence whose d-tuple is $\mathbf{n} := (n_1, \ldots, n_d)$, $\Sigma_{j=1}^d n_j = n$. When (1) is *exchangeable* and *invariant* and $m > n$ we have[3]

$$P\{X_{n+1} = j | \mathbf{X}^n\} = P\{X_m = j | \mathbf{X}^n\} =: P\{j|\mathbf{n}\} = \frac{\lambda p_j + n_j}{\lambda + n} \tag{2}$$

$j = 1, \ldots, d$, where $p_j := P\{X_m = j\}$ is the *initial* distribution,

$$\lambda := \frac{\eta}{1 - \eta} \qquad \text{and} \qquad \eta := \frac{P\{X_m = j | X_n = g\}}{p_j p_g}$$

We call (2) the *final* distribution (given $\mathbf{X}^n$ or $\mathbf{n}$).

Considering *weights* $w_j := \lambda p_j$, $\mathbf{w} := (w_1, \ldots, w_d)$, $\Sigma_{j=1}^d w_j = w$, we have the following form of (2),

$$P\{j|\mathbf{n}\} = \frac{w_j + n_j}{w + n}$$

In vectorial form the final distribution is

$$\mathbf{p}(\mathbf{w}, \mathbf{n}) = \frac{1}{w + n} (\mathbf{w} + \mathbf{n}) \tag{3}$$

Considering the vertices $\mathbf{v}_i = (0, \ldots, 1, \ldots, 0)$, $i = 1, \ldots, d$, of a $(d - 1)$-dimensional simplex as the extreme (deterministic) distributions $\mathbf{v}_i(j) = \delta_{ij}$ for each $j = 1, \ldots, d$, then the predictive distribution is a convex combination of these distributions with weights $w_j + n_j$, that is, (3) becomes

$$\mathbf{p}(\mathbf{w}, \mathbf{n}) = \frac{1}{w + n} \sum_{i=1}^{d} (w_i + n_i)\mathbf{v}_i$$

The method we have just sketched establishes a 1–1 correspondence between the possible outcomes and the vertices of a simplex. An observation amounts to an (empirical) instantiation of a vertex of the simplex. The intuitive formulation of an exchangeable, invariant predictive inference referring to a one-dimensional stochastic process is as follows: any step of the process instanti-

[3] For a detailed discussion of this derivation see Costantini and Garibaldi (1989, 1991).

ates an extreme point; the weight of the instantiated vertex is increased by 1; the final predictive distribution is determined by renormalization.

5. UNCERTAIN EVIDENCE

We call an outcome that is biunivocally associated with an extreme point *pure*. A pure outcome can be thought of as coming from a mass whose members share the same value of a variable. The frequency distribution of a mass of this type concentrates all members on one value and none on the others. The law ruling this mass is a degenerate probability distribution which allots 1 to one value of its domain. We call laws of this type *deterministic*. Deterministic laws can also be used as extreme laws giving a statistical description of a mass with respect to two variables. This can be done by melting the two variables into a new one whose values are the Cartesian product of the given two. In this case an inference can be performed in the way we have described in the previous section.

But we can consider also outcomes which are not pure. An outcome of this type is not an instance of a single vertex, but involves more than one vertex. We call these outcomes *mixed*. A number of reasons, both subjective and objective in character, may give rise to mixed observations. A rather trivial case occurs when on throwing a die we can observe only whether the occurring number is even or odd, while the probabilistic dynamics depends on the number which actually occurred. A more interesting one is given by statistical physics: observed values are mixed because they refer to observable macroprocesses, while the vertices refer to unobservable microprocesses. For this reason we shall also speak of observable outcomes of the macroprocess and unobservable instantiations of the vertices of the microprocess. The case we are going to take into account arises by admitting conditions on vertices which forbid deterministic laws. This happens quite naturally when we consider a mass with respect to two variables. In this case, instead of considering extreme laws which concentrate the probability on a cell of the $k \times h$ table, we suppose that extreme laws join each value of one variable to one and only one of the other. As a consequence extreme laws are no longer deterministic, but statistical. If this is the case, no outcome can instantiate an extreme law; on the contrary, each law is consistent with more than one outcome.

Bearing this in mind, let us consider the set of *observable outcomes* $O = \{1, \ldots, i, \ldots, s\}$ and the set of *unobservable vertices* $U = \{1, \ldots, j, \ldots, t\}$ and let $P\{i|j\}$ be the (hypothetical) probability of the (observable) outcome i given the (unobservable) vertex j. It is worth noting that $P\{i|j\}$, as a function of i, is the hypothetical probability of this outcome given the jth vertex, while, as a function of j, it is the likelihood of this vertex given the ith outcome. In statistical physics and in the above-considered case of the die, O is a partition of U, that is, $P\{i|j\}$, as function of i, is deterministic. When between O and U there is a 1–1 correspondence, $P\{i|j\}$ is a delta function and we are in the case considered in the previous section. We now see what happens when none of these cases hold.

For the predictive probability of an observable (but not yet observed) outcome we have

$$P\{i\} = \sum_{j \in U(i)} P\{i|j\}P\{j\}$$

where $U(i)$ is the set of vertices compatible with i and $P\{i|j\}$ is the likelihood of j. In general, given a microevidence whose k-tuple is $\mathbf{n}$, we have

$$P\{i|\mathbf{n}\} = \sum_{j \in U(i)} P\{i|j\}P\{j|\mathbf{n}\} \tag{4}$$

where $P\{j|\mathbf{n}\}$ is the microprobability of j. According to (4), in order to determine the probability of a macroprocess, we must know that of the microprocess. To do this, we suppose that the microprocess is exchangeable and invariant, i.e., that the final distribution of the microprocess is given by (2). Hence, given the macroevidence $\mathbf{N} = (N_1, \ldots, N_s)$, we have

$$P\{j|\mathbf{N}\} = \sum_{\mathbf{n} \in \mathsf{N}} P\{j|\mathbf{n}\}P\{\mathbf{n}|\mathbf{N}\} = \frac{1}{w+n} \sum_{\mathbf{n} \in \mathsf{N}} (w_j + n_j) \frac{P\{\mathbf{N}|\mathbf{n}\}\ P\{\mathbf{n}\}}{\sum_{\mathbf{n} \in \mathsf{N}} P\{\mathbf{N}|\mathbf{n}\}\ P\{\mathbf{n}\}}$$

where N is the set of all microvectors $\mathbf{n}$ consistent with the macroevidence $\mathbf{N}$.

6. TWO EXAMPLES

Now we give two examples of application of the method described in the previous section considering the most simple case, that in which the variables are both dichotomic, i.e., for the considered variables $k = h = 2$. Hence the four possible outcomes are the entries of table (I).

In the first example we take as extreme laws the tables of cograduation and contragraduation, i.e., the matrices

$$\underline{\mathbf{M}}_1 = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \underline{\mathbf{M}}_2 = \frac{1}{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and allot them the weights of the pair $\mathbf{w} = (P\{\underline{\mathbf{M}}_1\} = w_1, P\{\underline{\mathbf{M}}_2\} = w_2)$, $w_1 + w_2 = w$. Doing this, we assign to each entry the initial probabilities of the following bivariate distribution:

$$\underline{\mathbf{p}}(\mathbf{w}, \underline{\mathbf{0}}) = \frac{1}{2w}\begin{bmatrix} w_1 & w_2 \\ w_2 & w_1 \end{bmatrix}, \qquad \text{where } \underline{\mathbf{0}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

For the sake of simplicity we consider a unique outcome that we suppose to be

$$-- := \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

The final distribution is

$$\underline{\mathbf{p}}(\mathbf{w}, --) = \frac{1}{2(w+1)} \begin{bmatrix} w_1 + 1 & w_2 \\ w_2 & w_1 + 1 \end{bmatrix} \tag{5}$$

which shows the nonlocal features of the considered probability function.

A slightly more complex case arises when we take as extreme laws the following three matrices[4]:

$$\underline{\mathbf{M}}_1 = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \underline{\mathbf{M}}_2 = \frac{1}{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \underline{\mathbf{M}}_3 = \frac{1}{2}\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$

and the weights $\mathbf{w} = (P\{\underline{\mathbf{M}}_1\} = w_1, P\{\underline{\mathbf{M}}_2\} = w_2, P\{\underline{\mathbf{M}}_3\} = w_3)$, $w_1 + w_2 + w_3 = w$. In this case we take as initial distribution the matrix

$$\underline{\mathbf{p}}(\mathbf{w}, \underline{\mathbf{0}}) = \frac{1}{2w}\begin{bmatrix} w_1 & w_2 + w_3 \\ w_2 & w_1 + w_3 \end{bmatrix}$$

whose margins are not uniform. Considering the same outcome as before, we have

$$\underline{\mathbf{p}}(\mathbf{w}, --) = \frac{1}{2(w+1)}\begin{bmatrix} w_1 + \dfrac{w_1}{w_1 + w_3} & w_2 + w_3 + \dfrac{w_3}{w_1 + w_3} \\ w_2 & w_1 + w_3 + 1 \end{bmatrix}$$

As can easily be imagined, taking into account a great number of extreme laws has the result of emphasizing the nonlocal feature of the probability function.

7. CONCLUSION

In conclusion we comment on the probability function described in the present paper.

First, we point out the difference between the extreme laws of inferences based on pure and mixed evidence. In the first case the extreme laws are deterministic and each outcome is compatible with only one law. In the second case the extreme laws are statistical and each outcome is compatible

[4]This case was first examined by Daboni and Wedlin (1982). They referred to de Finetti's theorem taking a Dirichlet distribution as prior.

with more than one law. Moreover, these laws are statistical but of a special order. Considering a law of this type, we can restore determinism about one variable by choosing a definite value of the other. For example, considering $\underline{\mathbf{M}}_2$ of the previous section and limiting our attention to the value $-$ of variable I, we have a deterministic law for variable II, that is, all members of this submass have value $+$.

Second we notice the difference between the two types of dependences of the two methods we have considered. In the method sketched in Section 4 we take into account the first sort of dependence. More specifically, supposing that the statistical behavior of the members of a mass is described by (2), the dependence ensues from the fact that a member of a mass bears a value of the same variable or not. Considering elementary particles, in the simpler case, we have

$$P\{X_m = j\} \neq P\{X_m = j | X_n = j\}$$

In fact, once we put $p_j = k^{-1}$, the dependence characterizing these particles is captured by the parameter λ:

Putting $\lambda = k$, i.e., considering bosons, we have

$$P\{X_m = j\} = \frac{1}{k} \neq \frac{1 + 1}{k + 1} = P\{X_m = j | X_n = j\}$$

Putting $\lambda = -k$, i.e., considering fermions, we have

$$P\{X_m = j\} = \frac{1}{k} \neq \frac{1 - 1}{k + 1} = P\{X_m = j | X_n = j\}$$

Also, the method described in Section 5 shows, with respect to the microprocess of vertices, this type of dependence. But it shows another type of dependence, namely that existing between the two subprocesses. Considering the first example of the previous section, the final distribution (5) attests that, having observed a member of the submass marked by $II-$ with the value $-$ of the variable I has changed the probability of being $I-$ for a member of the same submass, as is obvious, but also the probability of being $I+$ for a member of the other submass marked by $II+$, and this is not so obvious. The frequency distribution resulting from Bell's experiment shows this type of dependence. We have presented a statistical method that accounts for the same type of dependence.

REFERENCES

Costantini, D., and Garibaldi, U. (1989). Classical and quantum statistics as finite random processes, *Foundations of Physics,* **19**, 743–754.

Costantini, D., and Garibaldi, U. (1991). Una formulazione probabilistica del principio di esclusione di Pauli nel contesto delle inferenze predittive, *Statistica,* **LI**, 21–34.

Daboni, L., and Wedlin, A. (1982). *Statistica. Un'introduzione all'impostazione neo-bayesiana,* Utet Turin, Chapter VII.

Suppes, P., and Zanotti, M. (1980). A new proof of the impossibility of hidden variables using the principles of exchangeability and identity of conditional distributions, in *Studies in the Foundations of Quantum Mechanics,* P. Suppes, ed., East Lansing, Michigan.

Suppes, P., and Zanotti, M. (1984). Causality and symmetry, *The Wave–Particle Dualism,* S. Diner *et al.*, eds., Reidel, Dordrecht.