Joint distributions, quantum correlations, and commuting observables

Arthur Fine
University of Illinois at Chicago Circle, Chicago, Illinois 60680

(Received 25 August 1981; accepted for publication 26 February 1982)

We provide necessary and sufficient conditions for several observables to have a joint distribution. When applied to the bivalent observables of a quantum correlation experiment, we show that these conditions are equivalent to the Bell inequalities, and also to the existence of deterministic hidden variables. We connect the no-hidden-variables theorem of Kochen and Specker to these conditions for joint distributions. We conclude with a new theorem linking joint distributions and commuting observables, and show how violations of the Bell inequalities correspond to violations of commutativity, as in the theorem.

PACS numbers: 02.50. – r, 03.65.Bz

1. INTRODUCTION

The question of when joint probabilities exist in quantum mechanics is not entirely settled, although several results in the literature suggest that joint probabilities exist only for commuting observables.[1] We show here that the special question of whether observables can have a joint distribution in a given, fixed state lies at the center of recent investigations into hidden variables; in particular, it is the key to the Bell theorems[2] and the no-hidden-variables result of Kochen and Specker.[3] We conclude with a new theorem linking joint distributions and commuting observables.

2. STATISTICAL OBSERVABLES

In this section we establish a framework, and results on joint probabilities, to be applied below to quantum mechanics. We begin by defining a *statistical observable* (or, *observable*, for short) as a pair $\langle A, P_A \rangle$, where A is a real-valued function and P_A is a probability measure on the Borel subsets of the reals ($\mathbf{R}$). Intuitively, $P_A(S)$ gives the probability that A takes a value in S. Thus every random variable paired with its distribution function is a statistical observable. In quantum mechanics every self-adjoint operator $\tilde{A}$ gives rise to statistical observables $\langle A, P_A^\Psi \rangle$, where A maps a sequence of unit rays ("states") ϕ_n to a real number λ iff $\mathrm{Inf}\|\tilde{A}\phi_n - \lambda\phi_n\| = 0$, and where $P_A^\Psi(S) = \langle \chi_S(A)\rangle_\Psi$, for χ_S, the characteristic function of the set S and Ψ any state (i.e., unit ray). It is convenient to refer to the function A alone, in the pair $\langle A, P_A \rangle$, as the (statistical) observable, suppressing reference to the measure P_A. Using this convention, we define a *joint distribution* of statistical observables A_1, $A_2,\ldots,A_n$ as a probability measure $P_{A_1,\ldots,A_n}$ on the Borel subsets of $\mathbf{R}^n$ returning each measure associated with each observable as marginals; i.e., satisfying

$$P_{A_1,\ldots,A_i,\ldots,A_n}(\mathbf{R}\times\cdots\times S\times\cdots\times\mathbf{R}) = P_{A_i}(S),$$

where Borel set S occurs in the i^{th} place in the Cartesian product. It is trivial to show that observables always have a joint distribution, since the product measure

$$P_{A_1,\ldots,A_n} = P_{A_1}\cdots P_{A_n}$$

always suffices. If, however, one is given a set of observables with certain fixed joint distributions already defined for various tuples of observables in the set, then it is a nontrivial question whether there exists a joint distribution for all the observables that returns the fixed joints as marginals. If so, we shall say that there is a joint distribution *compatible with* the fixed joints. We now establish some results of this sort, having in mind an application to quantum correlation experiments. (Intuitively, below, think of the fixed joints as the ones quantum mechanics gives—in some state—for pairs of commuting observables.)

Theorem 1: Let observables $A_1, A_2, \ldots, A_n; B_1, \ldots, B_m$ be given together with joint distributions P_{A_i,B_j} for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. There exists a joint distribution for all $n + m$ observables compatible with the given joints if and only if there exists a joint distribution $P_{B_1,\ldots,B_m}$ and corresponding joint distributions $P_{A_i,B_1,\ldots,B_m}$, each of which is compatible with $P_{B_1,\ldots,B_m}$ and P_{A_i,B_j} for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$.

Proof: Clearly, if there is a joint distribution for all the $n + m$ observables, compatible with the joints for the AB pairs, then the stated conditions hold. To establish the converse, notice that these conditions enable one to define density functions $\rho_i = dP_{A_i,B_1,\ldots,B_m}$ on $\mathbf{R}^{m+1}$ and a density $\beta = dP_{B_1,\ldots,B_m}$ on $\mathbf{R}^m$ such that for $\mathbf{y} = \langle y_1, \ldots, y_m \rangle$, $\int_{\mathbf{R}}\rho_i(x_i, \mathbf{y})dx_i = \beta(\mathbf{y})$ for $i = 1, \ldots, n$. Then for $\mathbf{x} = \langle x_1, \ldots, x_n \rangle$ we can define a probability density ρ on $\mathbf{R}^{n+m}$ by

$$\rho(\mathbf{x}, \mathbf{y}) = [\rho_1(x_1,\mathbf{y})\ldots\rho_n(x_n,\mathbf{y})]/\beta^{n-1}(\mathbf{y}). \tag{1}$$

(For $\beta = 0$, we can set the left-hand side to zero as well.) This is a proper density, for

$$\int_{\mathbf{R}^{n+m}}\rho(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} = \int_{R^m}\beta(\mathbf{y})d\mathbf{y} = 1.$$

Moreover, we get the given distributions $P_{A_i,B_1,\ldots,B_m}$ back as marginals because

$$\int_{\mathbf{R}^{n-1}}\rho(\mathbf{x},\mathbf{y})dx_1\ldots dx_{i-1}dx_{i+1}\cdots dx_n = \rho_i(x_i,\mathbf{y}).$$

for $i = 1, 2, \ldots, n$. Finally, since each ρ_i returns P_{A_i,B_j} $(j = 1, \ldots, m)$ as marginals, the probability measure on the Borel subsets of $\mathbf{R}^{n+m}$ corresponding to the density ρ is the required joint distribution. To apply the theorem it is

 0022-2488/82/071306-05$02.50

useful to state an immediate corollary.

Corollary: Necessary and sufficient for the existence of a joint distribution for observables $A_1, ..., A_n; B_1, B_2$, compatible with given joints P_{A_i,B_j} ($1 \leqslant i \leqslant k < n$ and $j = 1,2$), is the existence of a joint distribution P_{B_1,B_2} and of distributions P_{A_i,B_1,B_2}, each of the latter compatible with P_{B_1,B_2} and with the given P_{A_i,B_j}.

The corollary enables one to reduce the general problem to conditions on triples of obervables, which we now study in a special case.

Theorem 2: If A, B, B' are bivalent observables (each mapping into $\{x,y\}$) with given joint distributions $P_{A,B}, P_{A,B'}$ and $P_{B,B'}$, then necessary and sufficient for the existence of a joint distribution $P_{A,B,B'}$, compatible with the given joints for the pairs, is the satisfaction of the following system of inequalities:

$$P(A) + P(B) + P(B') \leqslant 1 + P(AB) + P(AB') + P(BB'), \tag{2a}$$

$$P(AB) + P(AB') \leqslant P(A) + P(BB'), \tag{2b}$$

$$P(AB) + P(BB') \leqslant P(B) + P(AB'), \tag{2c}$$

and

$$P(AB') + P(BB') \leqslant P(B') + P(AB), \tag{2d}$$

where we write $P(\;)$ for the probability that each enclosed observable takes the value x.[4]

Proof: Write $\bar{S}$ for the observable taking value y iff S takes value x, and let $\alpha = P(ABB')$. Then the terms in a distribution $P_{A,B,B'}$, if there were one compatible with the given joint distributions for pairs, would have to satisfy

$$P(AB\bar{B}') = P(AB) - \alpha, \tag{3a}$$

$$P(A\bar{B}B') = P(AB') - \alpha, \tag{3b}$$

$$P(A\bar{B}\bar{B}') = P(A) - P(AB) - P(AB') + \alpha, \tag{3c}$$

$$P(\bar{A}BB') = P(BB') - \alpha, \tag{3d}$$

$$P(\bar{A}B\bar{B}') = P(B) - P(AB) - P(BB') + \alpha, \tag{3e}$$

$$P(\bar{A}\bar{B}B') = P(B') - P(AB') - P(BB') + \alpha, \tag{3f}$$

$$P(\bar{A}\bar{B}\bar{B}') = 1 - P(A) - P(B) - P(B') + P(AB) + P(AB') + P(BB') - \alpha. \tag{3g}$$

Using $0 \leqslant \alpha \leqslant \min(P(AB), P(AB'), P(BB'))$, the condition that each term in (3) be non-negative produces the system (2). For example, requiring (3c) to be non-negative yields (2b). Conversely, if the system (2) is satisfied then choosing α as above insures that Eqs. (3) define the required distribution $P_{A,B,B'}$.

If we combine Theorem 2 with the corollary to Theorem 1, we get a good working condition for when bivalent observables $A_1, ..., A_n, B_1, B_2$ with preassigned joints P_{A_i,B_j} for $1 \leqslant i \leqslant k < n$ and $j = 1,2$, have a compatible joint distribution; namely, when there exist joint distributions $P_{A_l B_j}$ (for $k < l \leqslant n$ and $j = 1,2$) such that the system (2) of inequalities is simultaneously satisfiable for $A = A_i$, $B = B_1$, and $B' = B_2$, $1 \leqslant i \leqslant n$. In special cases these inequalities form an especially tractable system.

Theorem 3: If A_1, A_2, B_1, B_2 are bivalent observables with joint distributions P_{A_i,B_j} (for $i = 1,2$ and $j = 1,2$), then necessary and sufficient for there to exist a joint distribution P_{A_1,A_2,B_1,B_2} compatible with the given joints is that the following system of inequalities is satisfied:

$$-1 \leqslant P(A_iB_j) + P(A_i,B_{j'}) + P(A_{i'}B_{j'}) - P(A_{i'}B_j) - P(A_i) - P(B_{j'}) \leqslant 0, \tag{4}$$

for $i \neq i' = 1,2$ and $j \neq j' = 1,2$.

Proof: To show necessity note that, assuming the distribution P_{A_1,A_2,B_1,B_2}, for $i \neq i' = 1,2$, and $j \neq j' = 1,2$,

$$P(A_iB_jB_{j'}) = P(A_iA_{i'}B_jB_{j'}) + P(A_i\bar{A}_{i'}B_jB_{j'}) \leqslant P(A_{i'}B_j) + P(B_{j'}) - P(A_{i'}B_{j'}) \tag{5}$$

and

$$P(\bar{A}_iB_jB_{j'}) = P(\bar{A}_iA_{i'}B_jB_{j'}) + P(\bar{A}_i\bar{A}_{i'}B_jB_{j'}) \leqslant P(A_{i'}B_{j'}) + P(B_j) - P(A_{i'}B_j). \tag{6}$$

Also

$$0 \leqslant P(A_i\bar{B}_j\bar{B}_{j'}) = P(A_i) - P(A_iB_j) - P(A_iB_{j'}) + P(A_iB_jB_{j'}), \tag{7}$$

and

$$0 \leqslant P(\bar{A}_i\bar{B}_j\bar{B}_{j'}) = 1 - P(A_i) - P(B_j) - P(B_{j'}) + P(A_iB_j) + P(A_iB_{j'}) + P(\bar{A}_iB_jB_{j'}). \tag{8}$$

Then (5) with (7) yields the right-hand side of (4), and (6) with (8) yields the left-hand side of (4). In order to show sufficiency, consider inequalities (2), first for $B = B_1$, $B' = B_2$, and $A = A_1$ and then, similarly, for $A = A_2$. If these eight inequalities hold simultaneously for one and the same $P(B_1B_2)$ then, by Theorem 2 and the corollary to Theorem 1, we have the required P_{A_1,A_2,B_1,B_2}. To show that inequalities (4) guarantee all this, let $n = 1,2$ and $m \neq k = 1, 2$; set

$$\gamma = \min(P(A_nB_m) + P(B_k) - P(A_nB_k), P(B_m), P(B_k)) \tag{9}$$

and define $P(B_1B_2) = \gamma$. We can fill out the rest of the distribution P_{B_1,B_2} by letting $P(\bar{B}_1B_2) = P(B_1) - \gamma$, $P(B_1\bar{B}_2) = P(B_2) - \gamma$ and $P(\bar{B}_1\bar{B}_2) = 1 - P(B_1) - P(B_2) + \gamma$. Then (9) and the left-hand side of (4) imply that $P(A_i) + P(B_1) + P(B_2) \leqslant 1 + P(A_iB_1) + P(A_iB_2) + P(B_1B_2)$ for $i = 1, 2$. Similarly, (9) and the right-hand side of (4) imply the remaining six inequalities corresponding to (2b), (2c), and (2d) for the successive $A = A_1, A_2$; $B = B_1$, and $B' = B_2$.

3. CORRELATION EXPERIMENTS AND HIDDEN VARIABLES

We apply the preceding results to quantum correlation experiments. These involve distinct measurements of two noncommuting, bivalent obvservables (with values ± 1) A_1, A_2 in spacetime region R_1 and of two noncommuting, bivalent observables B_1, B_2 (values ± 1) in region R_2. Ideally, R_1 and R_2 would be spacelike separated. In any case, we assume that each A_i commutes with each B_j. Each measurement is performed on one of a correlated pair of particles, for example, on one of pairs of photons emitted in the singlet state from an atomic cascade (see Ref. 2). Various sets of assumptions about the workings of the experiment have been shown to lead to the probabilities of the experiment (i.e., the observed distributions for A_i,B_j and for the commuting pairs A_i, B_j) being constrained by the system of inequalities (4).

Let us refer to these, collectively, as *the Bell/CH inequalities*. It follows from Theorem 3, that the Bell/CH inequalities hold for the probabilities of a quantum correlation experiment if and only if there exists a joint distribution P_{A_1,A_2,B_1,B_2} for the observables of the experiment that is compatible with the observed distributions for the singles A_i and B_j and the commuting pairs A_i, B_j. We now show, in turn, that the existence of such a joint distribution function is equivalent to the existence of a deterministic hidden variables theory for the experiment. Such a theory is defined as follows. Let $A_1,\ldots,A_n$,... be observables of a quantum system, in a given state Ψ. A *deterministic hidden variables theory* for these observables (in that state) consists of a classical probability space $\Omega = \langle \Lambda, \sigma(\Lambda), P\rangle$, where Λ is a nonempty set (the "hidden variables" = "complete states" of the system), $\sigma(\Lambda)$ is a σ-algebra of subsets of Λ and P is a probability measure on $\sigma(\Lambda)$. We require that there is a mapping $A \rightarrow A(\;)$ from the observables $A = A_i$ to random variables on Ω, where the range of $A(\;)$ is the spectrum of A and satisfying

$$P_A^{\Psi} = P_{A(\;)}, \tag{D$_1$}$$

for each given observable $A = A_i$, and

$$P_{A,B}^{\Psi} = P_{A(\;),B(\;)} \tag{D$_2$}$$

for all commuting pairs, A, B among the given observables. {In (D$_2$) the left-hand side is the quantum joint distribution, determined by

$$P_{A,B}^{\Psi}(S\times T) = \langle \chi_S(A)\chi_T(B)\rangle_{\Psi}. \tag{10}$$

On the right-hand side of (D$_2$),

$$P_{A(\;),B(\;)}(S\times T) = P\,[A^{-1}(S)\cap B^{-1}(T)] \tag{11}$$

is the joint distribution of the random variables $A(\;)$, $B(\;)$.}

It is straightforward to see that there exists such a deterministic hidden variables theory for $A_1, A_2,\ldots$ if and only if there is a joint distribution for $A_1, A_2,\ldots$ compatible with the quantum mechanical distributions $P_{A_i}^{\Psi}$ and P_{A_i,B_j}^{Ψ}. For given such a hidden variables theory we can define the distribution for $A_1, A_2,\ldots$ by

$$P_{A_1,\ldots,A_n,\ldots} = P_{A_1(\;)\cdots A_n(\;)\cdots}; \text{ i.e.,}$$

as the usual product measure. Conversely, suppose we have a joint distribution $P_{A_1,\ldots,A_n,\ldots}$ compatible with the quantum single and joint probabilities (for commuting pairs), then let Λ consist of all sequences $\langle a_1, a_2, \ldots\rangle$, where $a_i \in$ *spectrum of* A_i. Let $\sigma(\Lambda)$ consist of all the infinite-dimensional Borel subsets of Λ^{ω}, and define P by

$$P(S_1\times\cdots\times S_n\times\cdots) = P_{A_1,\ldots,A_n,\ldots}(S_1\times\cdots\times S_n\times\cdots).$$

Then (D$_1$) and (D$_2$) follow from the compatibility requirements on $P_{A_1,\ldots,A_n,\ldots}$ if we associate with observable A_i the random variable $A_i(\;)$ defined by

$$A_i(\lambda) = a_i \quad \text{for } \lambda = \langle a_1,\ldots,a_i,\ldots\rangle \in \Lambda.$$

Clearly, this same equivalence between hidden variables and joint distributions obtains if we replace the left-hand side of (D$_1$) and (D$_2$) by any given distributions. In the case of the quantum correlation experiments, the weight of evidence suggest that the observed distributions are those of quantum mechanics (see Ref. 2). But even if this were not so, we could ask about the possibility for a hidden variables theory returning the experimentally observed probabilities, whatever they are, on the left-hand side of (D$_1$) and (D$_2$). We summarize the bearing of our results on this question in the following theorem.

Theorem 4: For a correlation experiment with observables $A_1\,A_2$, B_1, B_2 and with exactly the four pairs A_i, B_j ($i = 1,2; j = 1,2$) commuting, the following statements are mutually equivalent: (1) The Bell/CH inequalities hold for the single and double probabilities of the experiment; (2) there is a joint distribution P_{A_1,A_2,B_1,B_2} compatible with the observed single and double distributions; (3) there is a deterministic hidden variables theory for A_1, A_2, B_1, B_2 returning the observed single and double distributions; and (4) there is a well-defined joint distribution (for the noncommuting pair) P_{B_1,B_2} and joint distributions P_{A_1,B_1,B_2} and P_{A_2,B_1,B_2}, each of the latter compatible with P_{B_1,B_2} and with the observed single and double distributions.[5]

4. OTHER HIDDEN VARIABLES

There are observables whose quantum mechanical probabilities for certain states of correlated quantum systems violate the Bell/CH inequalities. Likewise, in most of the correlation experiments the observed probabilities also violate these inequalities. Thus both theoretically and experimentally we have a refutation of the possibility of deterministic hidden variables. Before the investigations initiated by Bell on correlated systems, however, there were other no-hidden-variables results. The strongest recent one is due to Kochen and Specker (Ref. 3). We show here the connection between their work and our investigation of joint probabilities and deterministic hidden variables.

Kochen and Specker begin by defining a hidden variables theory, for a set **0** of observables of a quantum system in state Ψ, exactly as in our definition in the preceding section for such a deterministic hidden variables theory, including (D$_1$) for every $A\in\mathbf{0}$, but not requiring (D$_2$) for commuting pairs. Let us refer to this as a *weak hidden variables theory*. They then suggest that a reasonable-looking formal requirement, in addition, would be to have the algebra of operators mirrored by the algebra of random variables. Thus they add the requirement

$$f(A)(\lambda) = f[A(\lambda)] \tag{KS}$$

for all $\lambda\in\Lambda$ and for every Borel function f (and for all $A\in\mathbf{0}$).

Our first result here is to show that if the set **0** is large enough, then (KS) is equivalent to (D$_2$). Specifically, define a set of observables **0** to be *large enough* if (1) whenever $A\in\mathbf{0}$ and $B\in\mathbf{0}$ and $AB = BA$, then $AB\in\mathbf{0}$, and also there is some observable $C\in\mathbf{0}$ such that $A = f(C)$ and $B = g(C)$ for Borel functions f and g; and (2) whenever $A\in\mathbf{0}$ and S is a Borel set, then $\chi_S(A)\in\mathbf{0}$.

Lemma: If **0** is large enough, then for $A\in\mathbf{0}$, $B\in\mathbf{0}$ and $AB = BA$, (KS) implies

$$AB(\lambda) = A(\lambda)B(\lambda). \tag{PR}$$

Proof: We have that $A = f(C)$ and $B = g(C)$ for $C\in\mathbf{0}$. By (KS), $A(\lambda) = f[C(\lambda)]$ and $B(\lambda) = g[C(\lambda)]$. But $AB = fg(C)$. So by (KS), $AB(\lambda) = fg(C)(\lambda) = fg[C(\lambda)]$ $= f[C(\lambda)]g[C(\lambda)] = A(\lambda)B(\lambda)$.

Theorem 5: If **0** is large enough, then a weak hidden variables theory for **0** satisfies (KS) only if it satisfies (D_2) for all commuting pairs A, B in **0**.

Proof: It follows from (D_1) and the lemma, that

$$P^{\Psi}_{A,B}(S\times T) = \langle \chi_S(A)\chi_T(B)\rangle_{\Psi} = P^{\Psi}_{\chi_S(A)\chi_T(B)}(1)$$
$$= P_{\chi_S[A(\ )]\chi_T[B(\ )]}(1).$$

By (KS), this yields

$$P^{\Psi}_{A,B}(S\times T) = P\left[\{\lambda\,|\chi_S(A(\lambda)) = \chi_T(B(\lambda)) = 1\}\right]$$
$$= P\left[A^{-1}(S)\cap B^{-1}(T)\right] = P_{A,B}(S\times T).$$

There is nearly a converse to this theorem, as follows.

Theorem 6: If **0** is large enough, then the following are equivalent. (1) There is a deterministic hidden variables theory for **0**; (2) there is a weak hidden variables theory for **0** satisfying (KS) almost everywhere; (3) there is a weak hidden variable theory for **0** satisfying (PR) almost everywhere.

Proof: We show that (1) implies (2), that (2) implies (3), and that (3) implies (1). To show that (1) implies (2), suppose we have (D_2) for all commuting pairs A, B in **0**. We want to show that $f(A)(\lambda)^{\text{a.e.}} = f[A(\lambda)]$; i.e., that $P[\{\lambda\,|f(A)(\lambda)\neq f[A(\lambda)]\}] = 0$. Let y be any number in the spectrum of $f(A)$, and let $S = \{\lambda\,|\,f[A(\lambda)] = y\}$ and $T = \{\lambda\,|\,f[A(\lambda)] = y\}$. We want $P(\bar{S}\cap T) = P(S\cap \bar{T}) = 0$. This will follow if we have $P(S) = P(T) = P(S\cap T)$. From (D_1) and the usual rules for functions of observables, we have $P(T) = P[\{\lambda\,|A(\lambda)\in f^{-1}(y)\}] = P^{\Psi}_{A}(f^{-1}(y)) = P^{\Psi}_{f(A)}(y) = P(S)$. $= P(S)$.) Using the spectral representation of A, it follows that $\chi_D(A)\chi_{f(D)}(f(A)) = \chi_D(A)$ for any set D, where $f(D) = \{f(x)|x\in D\}$. Hence, $P^{\Psi}_{Af(A)}(D\times f(D)) = \langle\chi_D(A)\rangle_{\Psi} = P^{\Psi}_{A}(D)$. In particular, $P^{\Psi}_{Af(A)}(f^{-1}(y)\times\{y\}) = P^{\Psi}_{A}(f^{-1}(y)) = P(S) = P(T)$. But, $P_{Af(A)}(f^{-1}(y)\times\{y\}) = P(S\cap T)$. The conclusion now follows from (D_2). That (2) implies (3) is a consequence of the lemma. Finally, the derivation of (1) from (3) has already been carried out elsewhere[6] and, since it involves no new principles, need not be repeated here.

This theorem has an immediate corollary that applies to the correlation experiments.

Corollary: If **0** is large enough, then a necessary condition for there to exist a weak hidden variables theory for **0** that satisfies (KS) [or (PR)] is that there exists a joint distribution for every finite subset of **0**, one compatible with all the well-defined quantum mechanical single and joint probabilities in that subset.

If we consider the observables A_1, A_2, B_1, and B_2 for a correlation experiment, then clearly there is a finite, large enough set **0** containing them all. According to the corollary above, and Theorem 4, the failure of the Bell/CH inequalities for particular correlated systems implies that there is no weak hidden variables theory satisfying (KS) for any finite large enough set of observables of such a system. It was just the tying down of the no-hidden-variables results to such finite systems of observables that was the central concern of the Kochen and Specker results. Our work in this section and the previous one shows that the Bell/CH inequalities for the correlation experiments achieve the same end.

5. COMMUTING OBSERVABLES

Our investigations suggest that what the different hidden variables programs have in common, and the common source of their difficulties, is the provision of joint distributions in those cases where quantum mechanics denies them. In this section, we formulate an intuitive criterion for a joint distribution, and show that its satisfaction in quantum mechanics leads to the usual connection between joint distributions and commuting operators.

If A and B are random variables over a common probability space with measure P, then for any two-place Borel function f and any Borel set S, the joint distribution $P_{A,B}$ is well defined, as is the random variable $f(A,B)$, and they satisfy the condition that $P_{A,B}(f^{-1}(S)) = P_{f(A,B)}(S)$. We now propose, essentially, the same condition as a criterion for when several observables of a quantum system have a joint distribution, as follows.

We shall say that observables $A_1, \ldots, A_n$ of a quantum system satisfy *the joint distribution condition* [briefly, (**jd**)] just in case, corresponding to every n-place Borel function f, there is an observable of the system with operator $f(A_1, \ldots, A_n)$, and corresponding to every state Ψ of the system there is probability measure $\mu_{\Psi,A_1,\ldots,A_n}$ on the Borel sets of $\mathbf{R}^n$ that returns the quantum single distributions $P^{\Psi}_{A_i}$ as marginals, such that

$$\mu_{\Psi,A_1,\ldots,A_n}(f^{-1}(S)) = P^{\Psi}_{f(A_1,\ldots,A_n)}(S) \tag{12}$$

for every state Ψ and Borel set S of reals.

Theorem 7: Observables $A_1, \ldots, A_n$ satisfy (**jd**) if and only if all pairs commute.[7]

Proof: If $A_1, \ldots, A_n$ form a commuting set then $f(A_1, \ldots, A_n)$ is well defined for every n-place Borel function f, and the usual joint distribution determined by $\mu_{\Psi,A_1,\ldots,A_n}(S_1\times\cdots\times S_n) = \langle\chi_{S_1}(A_1)\cdots\chi_{S_n}(A_n)\rangle_{\Psi}$ satisfies (**jd**) for all states Ψ. To show the converse we will show that if (**jd**) holds and $A = A_i$, $B = A_j$ then the spectral projections $\chi_S(A)$, $\chi_T(B)$ commute for any Borel sets S, T of reals. So suppose that i, j are fixed and S, T are given Borel sets. Then there are n-place Borel functions f and Borel sets of reals S', T' such that

$$\mathbf{R}\times\cdots\times S\times\cdots\times\mathbf{R} = f^{-1}(S') \tag{13}$$

and

$$\mathbf{R}\times\cdots\times T\times\cdots\times\mathbf{R} = f^{-1}(T'), \tag{14}$$

where S occurs in the ith place in (13), and T in the jth place in (14). For example, we can define a Borel function f by $f(x_1, \ldots, x_n) = 0$ for $x_i\notin S$ and $x_j\notin T$, $f(x_1, \ldots, x_n) = 1$ for $x_i\in S$ and $x_j\in T$, $f(x_1, \ldots, x_n) = 2$ for $x_i\in S$ and $x_j\notin T$, and $f(x_1, \ldots, x_n) = 3$ for $x_i\notin S$ and $x_j\in T$. If $S' = \{1, 2\}$ and $T' = \{1, 3\}$ then (13) and (14) hold. For such an f, we have from (**jd**) that

$$\begin{aligned} P^{\Psi}_{A}(S) &= \mu_{\Psi,A_1,\ldots,A_n}(\mathbf{R}\times\cdots\times S\times\cdots\times\mathbf{R}) \\ &= \mu_{\Psi,A_1,\ldots,A_n}(f^{-1}(S')) \\ &= P^{\Psi}_{f(A_1,\ldots,A_n)}(S'). \end{aligned} \tag{15}$$

Since (15) holds for all states Ψ, it follows that $\chi_S(A) = \chi_{S'}(f(A_1, \ldots, A_n))$. Similarly,

$\chi_T(B) = \chi_{T'}(f(A_1, ..., A_n))$. Hence, $\chi_S(A)$ commutes with $\chi_T(B)$.

The criterion (**jd**) and Theorem 7 help us to understand the significance of the violations of the Bell inequalities for the correlation experiments, for the observables A_1, A_2, B_1, B_2 of the experiments (with values ± 1) do not form a commuting set. Hence, by Theorem 7, if $f(x_1, x_2, y_1, y_2) = x_1y_1 + x_1y_2 + x_2y_2 - x_2y_1$ and we try the correspondence rule $f(A_1, A_2, B_1, B_2) = A_1B_1 + A_1B_2 + A_2B_2 - A_2B_1$ then Eq. (12) will fail for some set S and state Ψ. In particular, if S is the closed interval from -2 to $+2$, then $f^{-1}(S) \supseteq \{-1, 1\}^4$ and the left side of (12) must be 1 for any measure. But in certain singlet states Ψ (namely, those for which the Bell/CH inequalities fail) the quantum mechanical probability on the right side of (12) will differ from 1. Thus violations of the Bell/CH inequalities are particular cases where (**jd**) fails, as Theorem 7 tells us it somewhere must, for observables not all pairs of which commute. [Of course, it is Bell's important and lasting contribution to have found cases especially simple, and also experimentally tractable, where (**jd**) does fail.]

It seems natural to take (**jd**) as a criterion for when observables have a joint distribution. It is a coarse-grained criterion, not sensitive to the particular state of a system. As we have seen in the preceding sections, more finely grained criteria (and hidden variables are among them) are equivalent to constraints (like the Bell/CH inequalities) that some quantum systems violate in certain states. These violations have been experimentally confirmed. Perhaps, then, we ought to accept the straight-line induction; that where (**jd**) fails, and quantum mechanics does not give a well-defined joint distribution, neither would experiments. After all, if we hold that probabilities (including joint probabilities) are real properties, then some observables may simply not have them.

ACKNOWLEDGMENTS

Work for this study was supported by a grant from the National Science Foundation. My research assistant for this project, Dana Fine, provided the insight for the general formulation of Theorem 1 and, more especially, of its corollary. Paul Teller suggested the term "statistical observable" to replace the term "statistical variable" that I had used in previous publications. Patrick Suppes and Isaac Levi have been helpful correspondents.

[1]See K. Urbanik, Studia Math. **21**, 117 (1961); L. Cohen, J. Math. Phys. **7**, 781 (1966); E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton, U.P., Princeton, 1967), pp. 117–119; S. Gudder, J. Math. Mech. **18**, 325 (1968); V. V. Kuryshkin, in *The Uncertainly Principle and Foundations of Quantum Mechanics*, edited by W. Price and S. Chissick (Wiley, New York, 1977), pp. 61–83; S. Bugajski, Z. Naturforsch. Teil A **33**, 1383 (1978).

[2]See the survey of this literature by J. Clauser and A. Shimony, Rep. Prog. Phys. **41**, 1881 (1978).

[3]S. Kochen and E. Specker, J. Math. Mech. **17**, 59 (1967).

[4]For the special case where the range of the observables is ± 1 and where $\frac{1}{2} = P(A) = P(B) = P(B')$, a system of inequalities equivalent to (2) was first discovered by P. Suppes and M. Zanotti, Synthese (to be published). I want to thank Suppes and Zanotti for sharing their work with me, and for some very stimulating exchanges of ideas.

[5]We can add a fifth equivalent statement to this list; namely, (5) there exists a factorizable (so-called "local") stochastic hidden variables theory for A_1, A_2, B_1, B_2 returning the observed single and double distributions. See J. Clauser and M. Horne, Phys. Rev. D **10**, 526 (1974), for the definitions here. It is well known that (3) implies (5), and Clauser and Horne show that (5) implies (1). Thus the equivalence follows from our proof that (1) implies (3). It is easier, however, indeed trivial, to show that (5) implies (2), and to get the equivalence from that of (2) to (3). See my "Hidden variables, joint probability and the Bell inequalities," Phys. Rev. Lett. **48**, 291 (1982), which also contains another derivation of (3) from (1).

[6]A. Fine, Synthese **29**, 257 (1974). This is reprinted, with a relevant correction to the proof, in *Logic and Probability in Quantum Mechanics*, edited by P. Suppes (Reidel, Dordrecht, 1976), pp. 249–281.

[7]For pairs of discrete observables a computational proof is contained in A. Fine, Brit. J. Philos. Sci. **24**, 1 (1973). I want to thank Robert Latzer for correspondence that helped me find the simple, general derivation below.