

Quantum conditional probabilities

Ignacio Pérez and Alfredo Luis

Departamento de Óptica, Facultad de Ciencias Físicas, Universidad Complutense,
28040 Madrid, Spain

E-mail: alluis@fis.ucm.es

April 2022

Abstract. We investigate the consistency of conditional quantum probabilities. This is whether there is compatibility between the Kolmogorov-Bayes conditional probabilities and the Born rule. We show that they are not compatible in the sense that there are situations where there is no legitimate density matrix that may reproduce the conditional statistics of the other observable via the Born rule. This is to say that the Gleason theorem does not apply to conditional probabilities. Moreover, we show that when this occurs the joint statistics is nonclassical. We show that conditional probabilities are not equivalent to state reduction, so these results do not affect the validity of the Lüders expression.

Keywords: Conditional probabilities, Gleason theorem, Nonclassical states

1. Introduction

Measurement statistics is at the hearth of the very foundations of the quantum theory. This is so that the real signature of most quantum effects manifests itself in the differences between quantum and classical behavior of probabilities. Examples of this are for example Bell tests [1, 2, 3], or plain nonclassicality expressed in terms of lack of a well-behaved distribution, as exemplified by the Glauber-Sudarshan distribution in quantum optics [4]. In many occasions quantum-classical discrepancies are actually expressed in terms of intended probability distributions taking negative values in the quantum realm [3, 4, 8, 9, 10, 5, 6, 7, 12, 11].

In this work we present one of these quantum-classical statistical discrepancies in terms of the inconsistency between the standard Kolmogorov-Bayes formulation of statistics and the Born rule. We consider a joint noisy observation of two incompatible observables leading to a true joint probability distribution $p(x, y)$. We focus on the conditional probability $p(x|y) = p(x, y)/p(y)$ for x given y . We show situations where there is no legitimate density matrix ρ_y that might reproduce the conditional probability $p(x|y)$ via the Born rule, $p(x|y) = \text{tr}[\rho_y \Delta(x)]$, being $\Delta(x)$ the corresponding positive

operator-valued measure (POVM) for the x variable. That is to say that quantum conditional probabilities do not follow Gleason's theorem [13]. Throughout the work we will focus on POVMs because it is the most general scenario possible. In particular, this includes the case of observables well represented by Hermitian operators. Moreover, the POVM framework is the best suited for noisy joint observation of incompatible observables.

Moreover, we show that when there is no legitimate density matrix ρ_y we have that the joint statistics $p(x, y)$ is nonclassical. This is because once we remove the quantum noise in $p(x, y)$ the so obtained noise-free joint distribution $P(x, y)$ takes negative values. By nonclassical we refer to properties that cannot be explained by the classical theory. Precisely, the most powerful and standard way of revealing nonclassical features is by the lack of a true joint distribution for complementary observables, as it may be the case of the Glauber-Sudarshan P distribution in quantum optics [4]. Usually this is in the form of observation processes that in classical physics would lead to a legitimate joint probability distribution, but that in quantum physics fail to provide it. This is the idea of the formalism that we will follow in this work, that has been presented and extensively discussed in Refs. [14, 15, 16, 17, 18, 19, 20]. It is worth noting also that our work is not about the definition or construction of exact joint probabilities $P(x, y)$, but about conditional probabilities once the noisy joint distribution $p(x, y)$ has been obtained by practical means. The exact joint distribution $P(x, y)$ is just invoked as evidence of nonclassical behavior.

We will show in detail later that this situation is different in general from the problem of quantum state reduction associated with measurement, as clearly illustrated by a trivial example in Sec. II and the practical case of Young interferometer and homodyne detection examined in Appendices A and B. Note that if the solution of the problem would be given by Lüders approach to state reduction, then ρ_y would be always a well defined density matrix, and we show that this is not always the case. So, this is not a debate about Lüders expression since our work would shows that conditional probabilities are not equivalent to state reduction [21, 22, 23].

The existence of joint probabilities is an old and interesting problem as far as probability for two noncommuting observables is actually the next important case in the quantum statistical connection after the probability of a single observable [3, 24, 25]. Moreover, it has a direct relation with some of the most celebrated theorems at the hearth of the very foundations of the quantum theory, such as the Bell and Kochen-Specker theorems, as well as with potential quantum generalizations of conventional probability theory [26, 27, 28, 29].

Other formulations of the problem of quantum conditional probabilities can be found in Refs. [29, 30], via Kirkwood-type relations. This is to say that if $\Delta_X(x)$ and $\Delta_Y(y)$ are projectors determining the probability that two observables X, Y take the

values x, y , respectively, then a joint distribution for both with correct marginals is constructed via their product in the form $\Delta_{X,Y}(x, y) \propto \Delta_X(x)\Delta_Y(y)$, which in general is neither Hermitian nor positive semidefinite [31, 32, 33]. Nevertheless, Kirkwood-type expressions have demonstrated very interesting properties in quantum nonclassical statistics [12]. Other proposal within the context of quantum information focus on the relation between completely positive maps and compound states [34, 35]. A further different notion of quantum conditional probability has been addressed in [36], based on the conditional von Neumann entropy, while in [37] a novel form of quantum conditional probability is introduced to define new measures of quantum information in a dynamical context, revealing a failure of the Bayes theorem.

In Sec. II we provide the main settings and definitions required. Then, for the sake of simplicity, throughout the work we focus on a qubit system. In Sec. III we address the solution to the problem posed, showing that for every POVM there are states such that the Kolmogorov-Bayes conditional probabilities and the Born rule are incompatible. Then in Sec. IV we relate these facts with the quantum nature of the corresponding joint statistics. Finally, in the Appendices A and B we provide a real physical implementation of the analysis carried out in the rest of the paper in terms of fully measurable quantities. In Appendix A this is a Young interferometer, where the abstract observables X, Y become the classic complementary observables of interferometric path and interference. This provides physical intuition in terms of the usual quantum linear optics with a photon passing through a Young interferometer, phase plates and linear polarizers. In the Appendix B we consider the case of double homodyne detection implemented via beam splitting and interference. Moreover, the Appendices also serve to show that conventional state reduction in a Neumark scenario [38, 39, 40], also spelled Naimark in the literature, does not solve the question addressed here. We can recall that every POVM can be represented by a family of orthogonal projectors on an enlarged Hilbert space, that includes ancillary degrees of freedom where partial information about the observed system is transferred in the measurement process. Beyond qubit systems, in Appendix B we provide the example of double homodyne detection in the infinite-dimensional space of a single-mode field, proving the validity of the main results of this work.

2. Settings

Let us consider the joint measurement of two observables, which we call X and Y , of a given physical system, S . The measurement should admit a joint probability distribution $p(x, y)$, that tells us the probability that the property X takes the value x and property Y takes the value y . This is to say that X and Y must be compatible in the sense of being jointly measurable [41], also referred to as commensurable [42] or coexistent [43], which in general is not equivalent to commuting. Throughout this work we consider that

X and Y actually provide the noisy joint observation of two incompatible observables, say the Pauli matrices σ_X and σ_Y for a qubit system.

According to the Born rule and the Gleason theorem, every $p(x, y)$ is linearly determined from the density-matrix of the system ρ via a positive operator-valued measure $\Delta(x, y)$ as

$$p(x, y) = \text{tr} [\rho \Delta(x, y)], \quad (1)$$

where the statistical interpretation demands that

$$\Delta^\dagger(x, y) = \Delta(x, y), \quad \Delta(x, y) \geq 0, \quad \sum_{x, y} \Delta(x, y) = I, \quad (2)$$

and I is the identity. We will assume a discrete character for x, y just for the sake of simplicity. The corresponding marginal POVMs for the observables X and Y are, respectively,

$$\Delta_X(x) = \sum_y \Delta(x, y), \quad \Delta_Y(y) = \sum_x \Delta(x, y), \quad (3)$$

so that the marginal probabilities $p_W(w)$ that each observable $W = X, Y$ takes the value $w = x, y$ are

$$p_W(w) = \text{tr} [\rho \Delta_W(w)]. \quad (4)$$

A Kolmogorovian-Bayesian formulation of plain statistics lead us to define the conditional probability distribution $p(x|y)$ of getting x provided that Y takes the value y as

$$p(x, y) = p(x|y)p_Y(y). \quad (5)$$

The main questions we address in this work is whether this conditional probability determines a legitimate and unique density matrix ρ_y such that the conditional probability $p(x|y)$ may be obtained via the Born rule, that is

$$p(x|y) = \text{tr} [\rho_y \Delta_X(x)]. \quad (6)$$

No further assumptions are made beyond the basic relations (1), (4), (5) and (6). We even avoid to refer to ρ_y as the reduced state associated with the measurement of Y , nor impose any form for the relation between ρ_y and ρ , such as complete positivity of $\rho \rightarrow \rho_y$ mapping, or any other physical requirement. This is because we let the formalism to answer these points by itself once the solution is found. In particular, note that (5) is independent of whether the measurement of X precedes the measurement of Y or the other way round, so there is no point here regarding causal ordering. This will be clearly illustrated by the two examples presented in the Appendices.

As a suitable point of comparison let us consider the straightforward case when the joint POVM factorizes as the product of independent POVMs defined in different Hilbert spaces.

Theorem 1.– If the POVM factorizes, this is

$$\Delta(x, y) = \Delta_X(x) \otimes \Delta_Y(y), \quad (7)$$

then we have the following form for ρ_y :

$$\rho_y = \frac{\text{tr}_Y [\rho \Delta_Y(y)]}{\text{tr}_{X \otimes Y} [\rho \Delta_Y(y)]}, \quad (8)$$

where for clarity the subscripts on the traces indicate the space where the trace is performed.

Proof.– This is demonstrated by a simple ordering of the trace in (1). It can be easily seen that ρ_y is a legitimate density matrix as it can be easily checked, including positive semidefiniteness, that we will represent as $\rho_y \succeq 0$. This is because for any vector $|\psi\rangle$ in the X space we have

$$\langle \psi | \rho_y | \psi \rangle \propto \text{tr}_{X \otimes Y} [\rho | \psi \rangle \langle \psi | \otimes \Delta_Y(y)] \geq 0, \quad (9)$$

since $\rho \succeq 0$, $|\psi\rangle\langle\psi| \succeq 0$ and $\Delta_Y(y) \succeq 0$ ■.

This solution in the trivial case points out that we are not in a simple scenario of standard state reduction. In particular, notice that the state ρ_y belongs to the Hilbert space of the X measurement instead of the Hilbert space of the measurement Y .

2.1. Qubit system

Next we particularize this POVM framework to a qubit system. The chain of thought is the following. We first construct the most general POVM $\Delta(x, y)$ that represents a noisy joint observation of the two incompatible observables σ_X and σ_Y . Once we have $\Delta(x, y)$ we derive all the probabilities involved in (5) for the most general state of the system. Then, in Sec. 3 we will solve Eq. (5) to obtain ρ_y .

The most simple form for such $\Delta(x, y)$ involves just two dichotomic variables x, y , that may be referred to as the compatible X, Y observables taking the values x, y , respectively. By dichotomic we mean that $x, y = \pm 1$. In these conditions, the most general form for $\Delta(x, y)$ is

$$\Delta(x, y) = \frac{1}{4} (\Delta_0 + x\Delta_X + y\Delta_Y + xy\Delta_{XY}), \quad (10)$$

where the operators $\Delta_{X,Y,XY}$ no longer depend on x, y , being linear functions of the Pauli matrices $\boldsymbol{\sigma}$ and the 2×2 identity matrix σ_0 . From the third condition in (2) we get $\Delta_0 = \sigma_0$. Moreover, we just add a unbiasedness requirement on $\Delta(x, y)$ [44], that is that the maximally mixed state carries no information, which means that its statistics $p(x, y)$ does not depend on x, y :

$$\rho = \frac{1}{2}\sigma_0 \rightarrow p(x, y) = \frac{1}{4}. \quad (11)$$

This means that the $\Delta_{X,Y,XY}$ have no σ_0 component, so that the final form for our POVM is

$$\Delta(x, y) = \frac{1}{4} [\sigma_0 + \mathbf{S}(x, y) \cdot \boldsymbol{\sigma}], \quad (12)$$

with

$$\mathbf{S}(x, y) = x\gamma_X \mathbf{S}_X + y\gamma_Y \mathbf{S}_Y + xy\gamma_{XY} \mathbf{S}_{XY}, \quad (13)$$

where \mathbf{S}_X , \mathbf{S}_Y , and \mathbf{S}_{XY} are three-dimensional unit-modulus real vectors. The γ factors are real and positive expressing the accuracy of the observation as explained below. In any case all these elements are constrained so that $|\mathbf{S}(x, y)| \leq 1$, so that conditions (2) are satisfied. For definiteness let us consider that \mathbf{S}_X , \mathbf{S}_Y and \mathbf{S}_{XY} are mutually orthogonal

$$\mathbf{S}_X = (1, 0, 0), \quad \mathbf{S}_Y = (0, 1, 0), \quad \mathbf{S}_{XY} = (0, 0, 1), \quad (14)$$

so that $|\mathbf{S}(x, y)| \leq 1$ holds provided that

$$\gamma_X^2 + \gamma_Y^2 + \gamma_{XY}^2 \leq 1. \quad (15)$$

This choice is motivated by our understanding of $\Delta(x, y)$ as a noisy joint measurement of σ_X and σ_Y . In the Appendix A we provide a detailed derivation of this POVM in the particular but significant enough case of path-interference complementarity in a Young interferometer. Incidentally, note that the choice $\Delta_{XY} = \sigma_Z$ is rather natural if we consider that all information about the observable σ_W is in Δ_W for both $W = X, Y$. Moreover, since Δ_{XY} represents correlations between X and Y we can take into account that the most similar relation in terms of σ_X and σ_Y says that $\sigma_X \sigma_Y = i\sigma_Z$.

The joint POVM $\Delta(x, y)$ leads to the marginals for X and Y

$$\Delta_W(w) = \frac{1}{2} (1 + w\gamma_W \sigma_W), \quad (16)$$

with $W = X, Y$, $w = x, y$.

The system state can be expressed as

$$\rho = \frac{1}{2} (\sigma_0 + \mathbf{s} \cdot \boldsymbol{\sigma}), \quad (17)$$

where

$$\mathbf{s} = \text{tr}(\rho \boldsymbol{\sigma}) = (s_X, s_Y, s_Z), \quad (18)$$

is a three-dimensional real vector with $|\mathbf{s}| \leq 1$. Then we have the following joint distribution

$$p(x, y) = \frac{1}{4} (1 + x\gamma_X s_X + y\gamma_Y s_Y + xy\gamma_{XY} s_Z), \quad (19)$$

with marginals

$$p_X(x) = \frac{1}{2} (1 + x\gamma_X s_X), \quad p_Y(y) = \frac{1}{2} (1 + y\gamma_Y s_Y). \quad (20)$$

3. Solution

Combining (5), (19) and (20) we readily get

$$p(x|y) = \frac{1}{2} \left(1 + x\gamma_X \frac{s_X + y s_Z \gamma_{XY} / \gamma_X}{1 + y\gamma_Y s_Y} \right). \quad (21)$$

Expressing the the desired density matrix ρ_y in the most general form

$$\rho_y = \frac{1}{2} (\sigma_0 + \mathbf{t} \cdot \boldsymbol{\sigma}), \quad (22)$$

and using (16) we get

$$p(x|y) = \text{tr} [\rho_y \Delta_X(x)] = \frac{1}{2} (1 + x\gamma_X t_X). \quad (23)$$

Then, equating (21) and (23) we obtain

$$t_X = \frac{s_X + y s_Z \gamma_{XY} / \gamma_X}{1 + y\gamma_Y s_Y}. \quad (24)$$

This is the only condition on \mathbf{t} that we can get, so the other components t_Y and t_Z are free parameters, always respecting the condition $|\mathbf{t}| \leq 1$ for ρ_y being positive semidifinite.

Common quantum intuition may suggest that ρ_y is the reduced state associated with the measurement of Y , typically as $\sqrt{\Delta_Y(y)}\rho\sqrt{\Delta_Y(y)}$ according to the Lüders rule, taking always into account that the square root of a matrix is not unique reflecting the fact that POVM alone does not fully determine the reduced state . But in our case this is enough to show that there is no simple relation between ρ and ρ_y implemented by a quantum instrument in the typical form of a completely positive linear map. In particular, $\Delta_Y(y)$ does not depend on γ_{XY} while ρ_y actually does. Also, the factorized case already analyzed around (8) shows that such intuition fails regarding the particular problem we are addressing here.

The most important point is that there are situations in which there is no legitimate state ρ_y . This occurs when (24) is incompatible with the condition $|\mathbf{t}| \leq 1$ required so that $\rho_y \succeq 0$. We will refer to the case when ρ_y is not positive semidifinite as $\rho_y \not\succeq 0$. For example this is the case of $s_X = s_Y = 0$ provided that

$$|t_X| = \left| y s_Z \frac{\gamma_{XY}}{\gamma_X} \right| > 1, \quad (25)$$

that holds provided that $\gamma_{XY} \neq 0$ and γ_X is small enough, a choice that is always possible. For details regarding these conditions from the point of view of measurable quantities in a real physical scenario we refer to the Appendix A. Finally, let us prove the following theorem:

Theorem 2.– For all POVMs (12) with the choice (13) there is an state ρ such that the corresponding ρ_y is not a legitimate density matrix since $\rho_y \not\succeq 0$.

Proof.— Let the state ρ be of the form in (17) with

$$\mathbf{s} = \left(\sqrt{1 - \gamma_Y^2}, \gamma_Y, 0 \right), \quad (26)$$

and let it be $y = -1$. Then after (24)

$$t_X = \frac{1}{\sqrt{1 - \gamma_Y^2}} > 1 \rightarrow \rho_y \not\leq 0. \quad \blacksquare \quad (27)$$

4. Nonexistence and nonclassicality

Let us show here that the nonexistence of ρ_y in (24) implies the nonclassical character of ρ as next specified briefly recalling the formalism already presented in [14, 15, 16, 17, 18, 19, 20].

4.1. Nonclassicality

In a classical-physics scenario, the joint statistics of two observables is of the form, as extensively discussed when looking for test for quantum nonlocality and noncontextuality [45, 46, 47, 48, 49, 50, 51],

$$p(x, y) = \sum_{\lambda} \nu_X(x|\lambda) \nu_Y(y|\lambda) P(\lambda), \quad (28)$$

where $P(\lambda)$ is a well-behaved probability distribution over the states of the system, denoted by λ , which typically corresponds to its phase space, and $\nu_X(x|\lambda)$, $\nu_Y(y|\lambda)$ are suitable conditional probability distributions defined by the observation process, so they can be known in practice independently of the system state, in our case they will be determined later in (29). We say that $p(x, y)$ is separable when it admits the form (28) for a suitable probability distribution $P(\lambda)$, and two conditional probabilities $\nu_X(x|\lambda)$, $\nu_Y(y|\lambda)$, consistent with the measurement at hand. Otherwise, we say that $p(x, y)$ is nonseparable. It is relevant to stress that this separability condition holds in classical physics respectively of whether the observation is exact or noisy. But in quantum physics the situation is utterly different [20]. Thus we may say that if $p(x, y)$ fails to be separable then it is nonclassical.

The idea is that the POVM $\Delta(x, y)$ in (12) provides a simultaneous noisy observation of the two incompatible observables σ_X and σ_Y . Because of this in Eq. (28) we are not addressing a full hidden-variable theory valid for arbitrary measuring settings, but just a classical-like model for the particular noisy measuring procedure leading to the POVM (12). In this spirit we consider the four combinations $\lambda = (x, y)$ with $x, y = \pm 1$ to represent the exact values of the observables σ_X and σ_Y , so that $P(x, y)$ in Eq. (28) represents the exact noiseless joint distribution for these observables. In this spirit we may also naturally assume that the conditional probabilities take the form $\nu_X(x|\lambda) = \nu_X(x|x')$ and $\nu_Y(y|\lambda) = \nu_Y(y|y')$, so that the statistics of X is derived

exclusively in terms of the statistics of σ_X , and equivalently, the statistics of Y is derived exclusively in terms of the statistics of σ_Y . Moreover, we assume that the observed marginals $p_X(x)$ and $p_Y(y)$ in (20) contain complete statistical information about σ_X and σ_Y in the state ρ . To provide suitable expressions for the conditional probabilities, $\nu_X(x|x')$ and $\nu_Y(y|y')$ we focus on the observed marginal statistics in (20) considering the case of eigenstates of σ_X and σ_Y , so we may conclude that

$$\nu_W(w|w') = \frac{1}{2}(1 + \gamma_W w w'), \quad (29)$$

for $W = X, Y$ and $w = x, y$, $w' = x', y'$. Therefore, in a classical scenario the observed joint statistics satisfies a separability condition of the form

$$p(x, y) = \sum_{x', y' = \pm 1} \nu_X(x|x') \nu_Y(y|y') P(x', y'). \quad (30)$$

A direct way to test separability is to invert relation (30) to obtain $P(x, y)$ in terms of the observed statistics $p(x, y)$: if $P(x, y) \not\geq 0$ then $p(x, y)$ is not separable. This can be done in the form

$$P(x, y) = \sum_{x', y' = \pm 1} \mu_X(x, x') \mu_Y(y, y') p(x', y'), \quad (31)$$

where $\mu_W(w, w')$ are the inverses of $\nu_W(w|w')$

$$\mu_W(w, w') = \frac{1}{2} \left(1 + \frac{w w'}{\gamma_W} \right), \quad (32)$$

leading to

$$P(x, y) = \frac{1}{4} \left(1 + x s_X + y s_Y + x y s_Z \frac{\gamma_{XY}}{\gamma_X \gamma_Y} \right). \quad (33)$$

Thus we may refer to the case $P(x, y) \not\geq 0$ as evidence of nonclassical behavior. For a parallel inversion analysis in a slightly different context see for example [11].

4.2. Lack of density matrix ρ_y implies nonclassicality

Theorem 3.– If ρ_y is not positive semidefinite then $P(x, y)$ cannot be a probability distribution, this is $\rho_y \not\geq 0 \rightarrow P(x, y) \not\geq 0$.

Proof.– Let us begin with considering $t_x > 0$ so that $\rho_y \not\geq 0$ holds when $t_x > 1$. After (24) we have that $t_x > 1$ implies

$$1 - s_X + y \gamma_Y s_Y - y s_Z \frac{\gamma_{XY}}{\gamma_X} < 0. \quad (34)$$

Since always $1 - s_X \geq 0$ then

$$y \gamma_Y s_Y - y s_Z \frac{\gamma_{XY}}{\gamma_X} < 0, \quad (35)$$

and then, since $1 \geq \gamma_Y > 0$

$$y s_Y - y s_Z \frac{\gamma_{XY}}{\gamma_X \gamma_Y} \leq y \gamma_Y s_Y - y s_Z \frac{\gamma_{XY}}{\gamma_X}, \quad (36)$$

which implies

$$1 - s_X + y s_Y - s_Z \frac{\gamma_{XY}}{\gamma_X \gamma_Y} < 0. \quad (37)$$

This readily means after (33) that $P(x = -1, y) < 0$. Likewise $t_X < -1$ implies that $P(x = 1, y) < 0$. ■

This is to say that the incompatibility of the Kolmogorov-Bayes conditional probability (5) and the Born rule is a nonclassical phenomenon. This might be a signature of the intrinsic nonclassical nature of the measurement processes, as already pointed out in Ref.[52], implying a failure of the classical-like models such as (28).

The other way round there are situations with $P(x, y) \not\equiv 0$ but $|t_x| \leq 1$. For example this is the case of $s_X = s_Y = 0$ provided that

$$\frac{1}{\gamma_Y} \geq \left| s_Z \frac{\gamma_{XY}}{\gamma_X \gamma_Y} \right| > 1. \quad (38)$$

We may nevertheless derive a complete equivalence between $\rho_y \not\equiv 0$ and $P(x, y) \not\equiv 0$ provided we examine the same problem of compatibility between conditional probabilities and Born rule after data inversion. This is replacing (5) by

$$P(x, y) = P(x|y)P_Y(y), \quad (39)$$

where

$$P_Y(y) = \sum_x P(x, y) = \text{tr} [\rho \Omega_Y(y)], \quad (40)$$

and

$$\Omega_W(w) = \frac{1}{2} (1 + w \sigma_W), \quad (41)$$

are the actual orthogonal projectors on the eigenstates of σ_W , $W = X, Y$, $w = x, y$ so that

$$P_W(w) = \frac{1}{2} (1 + w s_W). \quad (42)$$

Then we look for the density matrix ρ_y defined by the Gleason-type relation

$$P(x|y) = \text{tr} [\rho_y \Omega_X(x)]. \quad (43)$$

Theorem 4. – There is a complete equivalence between ρ_y not being positive semidefinite and $P(x, y)$ not being a probability distribution, this is $\rho_y \not\equiv 0$ if and only if $P(x, y) \not\equiv 0$.

Proof. – After (33) and (42) we get

$$P(x|y) = \frac{1}{2} \left(1 + x \frac{s_X + y s_Z \frac{\gamma_{XY}}{\gamma_X \gamma_Y}}{1 + y s_Y} \right) \quad (44)$$

so that asking for the ρ_y density matrix in (43) using the same notation in (22), we get

$$t_X = \frac{s_X + y s_Z \frac{\gamma_{XY}}{\gamma_X \gamma_Y}}{1 + y s_Y}, \quad (45)$$

so that in this case we readily get that $|t_X| > 1$ holds if and only if $P(x = 1, y) < 0$ or $P(x = -1, y) < 0$, so there is no legitimate ρ_y if and only if the joint statistics is nonclassical ■.

Roughly speaking the equivalence is clear after (39) since $\Omega(x) \succeq 0$ and $P_Y(y) \geq 0$, so that the negativity of the inverted distribution $P(x, y) \not\geq 0$ must come from that $\rho_y \not\geq 0$.

It might seem strange to invoke negative distributions in this statistical context, but we might recall that there are many formalisms where the peculiarities of quantum mechanics are expressed in this way, or even in terms of complex distributions [3, 4, 8, 9, 10, 5, 6, 7, 11], in particular regarding conditional probabilities [12].

5. Conclusions

We have examined the compatibility between classical-like Kolmogorov-Bayes conditional probabilities, Born rule and Gleason theorem for POVMs representing noisy joint observations of incompatible observables. We have shown that there are situations where they are incompatible. We have found that this incompatibility has a clear and definite relation with the nonclassicality of the statistics defined by the very same POVM, manifesting in a nonseparable joint statistics.

Acknowledgments

We acknowledge financial support from Spanish Ministerio de Economía y Competitividad Project No. FIS2016-75199-P. We thank Dr. G. García Moreno for continuous support and helpful comments.

Appendix A. Young interferometer realization

Path-interference duality is a classic example of complementarity at work in a two-dimensional space, so it is a natural arena for a nontrivial joint observation of the kind we are addressing in this work. To this end we will follow the scheme in [15]. To fix ideas let us assume that we are dealing with a single-photon state of light.

Let us represent the state at the apertures of a Young interferometer by the two orthogonal kets $|\pm\rangle$ that represent the presence of the photon at the corresponding aperture. They may be always regarded as the eigenvectors of the Pauli matrix σ_Y , so let it be $Y \propto \sigma_Y$ as the path observable.

Interference occurs by the coherent superposition of $|\pm\rangle$, so we may represent interference by the observable $X \propto \sigma_X$. We will consider that the observed interference

X is directly measured on the system space S by projection on the eigenstates $|x\rangle$ of σ_X .

Path information, this is information about Y , will be transferred from the system space to an auxiliary space A , that will be the polarization of the photon at each aperture. The path information can be imprinted in the polarization state by a different phase plate placed on each aperture. When the apertures are illuminated by right-handed circularly polarized light, represented by the vector $|\circ\rangle$ in the polarization space A , the phase plates produce the following aperture-dependent polarization transformation

$$|\pm\rangle \otimes |\circ\rangle \rightarrow |\pm\rangle \otimes |\pm\theta\rangle, \quad (\text{A.1})$$

being

$$|\pm\theta\rangle = \cos\frac{\theta}{2}|\circ\rangle \pm \sin\frac{\theta}{2}|\circ\rangle. \quad (\text{A.2})$$

This corresponds to the action of the unitary operator

$$U|\pm\rangle \otimes |\circ\rangle = |\pm\rangle \otimes |\pm\theta\rangle, \quad (\text{A.3})$$

with

$$U = e^{-i\frac{\theta}{2}\sigma_Y \otimes \Sigma_2} = \cos\frac{\theta}{2}\sigma_0 \otimes \Sigma_0 - i\sin\frac{\theta}{2}\sigma_Y \otimes \Sigma_2, \quad (\text{A.4})$$

where for the sake of clarity we denote by capital Σ the Pauli matrices in the auxiliary polarization space A , being $|\circ\rangle$ and $|\circ\rangle$ the eigenvectors of Σ_3 , with eigenvalues $+1$ and -1 respectively.

Then, the path information is retrieved by measuring any combination of the observables represented by the Pauli matrices Σ_1 and Σ_2 in the auxiliary polarization space, say

$$\Sigma_\varphi = \cos\varphi\Sigma_1 + \sin\varphi\Sigma_2. \quad (\text{A.5})$$

This polarization measurement can be very easily achieved in practice with a linear polarizer, where φ represents the orientation of its axis. We denote as $|y\rangle$ the eigenvectors of Σ_φ with eigenvalue $y = \pm 1$, so the photon passing through the polarizer is represented by the vector $|y = 1\rangle$ while the photon being stopped by the polarizer is represented by the vector $|y = -1\rangle$.

Let us denote by $\tilde{\Delta}(x, y)$ the actual projection-valued measure in the total space $S \otimes A$

$$\tilde{\Delta}(x, y) = |x\rangle\langle x| \otimes |y\rangle\langle y|, \quad (\text{A.6})$$

leading to the joint statistics

$$p(x, y) = \text{tr}_{S \otimes A} \left[U\rho \otimes \rho_A U^\dagger \tilde{\Delta}(x, y) \right] = \text{tr}_S [\rho \Delta(x, y)], \quad (\text{A.7})$$

where

$$\rho_A = |\circ\rangle\langle\circ|, \quad (\text{A.8})$$

and the subscripts A and S refers in each case to the auxiliary polarization space or the system space, and we use no subscript S in ρ and $\Delta(x, y)$ to match the notation in the main body of the text. Then the actual POVM $\Delta(x, y)$ just defined in the system space S is

$$\Delta(x, y) = \text{tr}_A \left[\rho_A U^\dagger \tilde{\Delta}(x, y) U \right], \quad (\text{A.9})$$

which exactly of the form in (12) and (13) with

$$\gamma_X = \cos \theta, \quad \gamma_Y = \sin \theta \cos \varphi, \quad \gamma_{XY} = \sin \theta \sin \varphi, \quad (\text{A.10})$$

that are actually points on the surface of a unit sphere,

$$\gamma_X^2 + \gamma_Y^2 + \gamma_{XY}^2 = 1. \quad (\text{A.11})$$

Let us investigate the reduced states $\tilde{\rho}_y$ associated with the Y measurement, with a significant proviso: This is not what we were looking for in the main body of the paper, i. e. $\rho_y \neq \tilde{\rho}_y$, mainly because

$$\Delta_W(w) \neq |w\rangle\langle w|, \quad (\text{A.12})$$

for $W = X, Y$, $w = x, y$. To determine $\tilde{\rho}_y$ we follow exactly the same steps mentioned in the case of a factorized POVM since X and Y have their own independent Hilbert spaces,

$$\tilde{\rho}_y = \frac{\text{tr}_A [U \rho \otimes \rho_A U^\dagger |y\rangle\langle y|]}{\text{tr}_{S \otimes A} [U \rho \otimes \rho_A U^\dagger |y\rangle\langle y|]}, \quad (\text{A.13})$$

leading to

$$\tilde{\rho}_y = \frac{\Lambda_y \rho \Lambda_y^\dagger}{\text{tr}_S [\Lambda_y^\dagger \Lambda_y \rho]}, \quad (\text{A.14})$$

with

$$\Lambda_y = \langle y|U|\circ\rangle = \frac{1}{\sqrt{2}} \left(y e^{i\varphi} \cos \frac{\theta}{2} \sigma_0 + \sin \frac{\theta}{2} \sigma_Y \right), \quad (\text{A.15})$$

and

$$\Lambda_y^\dagger \Lambda_y = \Delta_Y(y) = \frac{1}{2} (\sigma_0 + y \gamma_Y \sigma_Y). \quad (\text{A.16})$$

Then it holds

$$p(x|y) = \text{tr}_S [\tilde{\rho}_y |x\rangle\langle x|], \quad (\text{A.17})$$

which is different from the relation we were looking for

$$p(x|y) = \text{tr}_S [\rho_y \Delta_X(x)]. \quad (\text{A.18})$$

Appendix B. Number states in double homodyne detection

Let us present an example beyond the qubit system that may illustrate the application of our approach in system spaces with higher dimension. This can be the case of double homodyne detection in an electromagnetic field mode of complex-amplitude operator a . The double homodyne detection may be considered as a joint simultaneous measurement of two field quadratures X and Y [53, 54]

$$X = \frac{1}{2}(a + a^\dagger), \quad Y = \frac{i}{2}(a^\dagger - a), \quad (\text{B.1})$$

whose statistics is given essentially, up to dilation factors, by the Husimi Q function [4]. So, the POVM in this case is

$$\Delta(x, y) = \frac{1}{\pi}|\alpha\rangle\langle\alpha|, \quad \alpha = x + iy, \quad (\text{B.2})$$

where $|\alpha\rangle$ are the Glauber-Sudarshan coherent states, and here x, y are unbounded continuous Cartesian variables. Let the mode be in a one-photon state, so that

$$p(x, y) = \frac{1}{\pi}(x^2 + y^2)e^{-x^2 - y^2}, \quad p_Y(y) = \frac{1}{2\sqrt{\pi}}(1 + 2y^2)e^{-y^2}, \quad (\text{B.3})$$

and the conditioned probability is

$$p(x|y) = \frac{2}{\sqrt{\pi}} \frac{x^2 + y^2}{1 + 2y^2} e^{-x^2}. \quad (\text{B.4})$$

Let us show that we can find values of y such that $\rho_y \not\leq 0$. To this end we note that

$$\Delta_X(x) = \sqrt{\frac{2}{\pi}} \int_{-\infty}^{\infty} dx' e^{-2(x-x')^2} |x'\rangle\langle x'|, \quad (\text{B.5})$$

where $|x\rangle$ are the eigenstates of the quadrature X , so that for any ρ_y

$$\text{tr}[\rho_y \Delta_X(x)] = \sqrt{\frac{2}{\pi}} \int_{-\infty}^{\infty} dx' e^{-2(x-x')^2} p_y(x'), \quad (\text{B.6})$$

where $p_y(x) = \langle x|\rho_y|x\rangle$ is the probability distribution of the X quadrature in the ρ_y state. The question is whether the equality $p(x|y) = \text{tr}[\rho_y \Delta_X(x)]$ can be fulfilled with a legitimate $p_y(x)$ distribution. The solution is simple since the equality can be inverted via Fourier analysis to determine $p_y(x)$ with the result:

$$p_y(x) = \sqrt{\frac{2}{\pi}} \frac{8x^2 + 2y^2 - 1}{1 + 2y^2} e^{-2x^2}. \quad (\text{B.7})$$

We can see that whenever $y^2 < 1/2$ the distribution $p_y(x) = \langle x|\rho_y|x\rangle$ takes negative values, for example at $x = 0$, and then $\rho_y \not\leq 0$.

References

- [1] Ballentine L E 1990 *Quantum Mechanics* (Englewood Cliffs: Prentice Hall) Ch 20
- [2] Bell J S 1964 On the Einstein Podolsky Rosen paradox *Physics* **1** 195–200
- [3] Fine A 1982 Hidden Variables, Joint Probability, and the Bell Inequalities *Phys. Rev. Lett.* **48** 291–5

- [4] Mandel L and Wolf E 1995 *Optical Coherence and Quantum Optics* (Cambridge: Cambridge University Press)
- [5] Feynman R P 1991 Negative Probabilities *Quantum Implications: Essays in Honour of David Bohm* (New York: Routledge) pp. 235–48
- [6] Feynman R P 1982 Simulating Physics with Computers *Int. J. Theor. Phys.* **21** 467–88
- [7] Ryu J, Hong S, Lee J-S, Seol K H, Jae J, Lim J, Lee J, Lee K-G and Lee J 2019 Optical experiment to test negative probability in context of quantum-measurement selection *Sci. Rep.* **9** 19021
- [8] Spekkens R W 2008 Negativity and Contextuality are Equivalent Notions of Nonclassicality *Phys. Rev. Lett.* **101** 020401
- [9] Rivas A 2019 On the role of joint probability distributions of incompatible observables in Bell and Kochen–Specker Theorems *Ann. Phys. (N. Y.)* **411** 167939
- [10] de Barros J A, Kujala J V and Oas G 2016 Negative probabilities and contextuality *J. Math. Psychol.* **74** 34–45
- [11] Suzuki Y, Iinuma M and Hofmann H F 2016 Observation of non-classical correlations in sequential measurements of photon polarization *New J. Phys.* **18** 103045
- [12] Hofmann H F 2012 Complex joint probabilities as expressions of reversible transformations in quantum mechanics *New J. Phys.* **14** 043031
- [13] Gleason A M 1957 Measures on the Closed Subspaces of a Hilbert Space *Journal of Mathematics and Mechanics* **6** 885–93
- [14] Masa E, Ares L and Luis A 2020 Nonclassical joint distributions and Bell measurements *Phys. Lett. A* **384** 126416
- [15] Galazo R, Bartolomé I, Ares L and Luis A 2020 Classical and quantum complementarity *Phys. Lett. A* **384** 126849
- [16] Luis A 2015 Nonclassical states from the joint statistics of simultaneous measurements arXiv:1506.07680 [quant-ph]
- [17] Luis A 2016 Nonclassical light revealed by the joint statistics of simultaneous measurements *Opt. Lett.* **41** 1789–92
- [18] Luis A and Monroy L 2017 Nonclassicality of coherent states: Entanglement of joint statistics *Phys. Rev. A* **96** 063802
- [19] Cónsul R and Luis A 2020 Detector self-tomography *Opt. Lett. A* **45** 6799
- [20] Masa E, Ares L and Luis A 2022 Inequalities for complementarity in observed statistics *Phys. Lett. A* **427** 127914
- [21] Guerra I 2013 On Quantum Conditional Probability *THEORIA* **76** 115–37
- [22] Busch P and Lahti P 2009 Lüders rule *Compendium of Quantum Physics* (Berlin: Springer)
- [23] Ozawa M 1985 Conditional probability and a posteriori states in quantum mechanics *Publ. RIMS Kyoto Univ.* **21** 279–95
- [24] Uffink J 1994 The joint measurement problem *Int. J. Theor. Phys.* **33** 199–212
- [25] Allahverdyan A E and Danageozian A 2018 Excluding joint probabilities from quantum theory *Phys. Rev. A* **97** 030102(R)
- [26] Fine A 1982 Hidden Variables, Joint Probability, and the Bell Inequalities *Phys. Rev. Lett.* **48** 291–295
- [27] Pitowsky I 1982 Resolution of the Einstein-Podolsk@-Rosen and Bell Paradoxes *Phys. Rev. Lett.* **48** 1299–1302
- [28] Gudder S P 1984 Reality, Locality, and Probability *Found. Phys.* **14** 997–1010
- [29] Malley J D 2004 All quantum observables in a hidden-variable model must commute simultaneously *Phys. Rev. A* **69** 022118
- [30] Miller D J 1997 Conditional probabilities in quantum mechanics from a time-symmetric formulation *Il Nuovo Cimento B* **112** 1577–92
- [31] Kirkwood J G 1933 Quantum Statistics of Almost Classical Assemblies *Phys. Rev.* **44**, 31–37
- [32] Dirac P A M 1945 On the Analogy Between Classical and Quantum Mechanics *Rev. Mod. Phys.* **17** 195–199

- [33] Lee H-W 1995 Theory and application of the quantum phase-space distribution functions *Physics Reports* **259** 147–211
- [34] Asorey M, Kossakowski A, Marmo G and Sudarshan E C G 2006 Relations Between Quantum Maps and Quantum States *Open.Syst. Info. Dyn.* **12** 319–29
- [35] Leifer M S 2006 Quantum dynamics as an analog of conditional probability *Phys. Rev. A* **74** 042310
- [36] Cerf N J and Adami C 1999 Quantum extension of conditional probability *Phys. Rev. A* **60** 893–7
- [37] Barandes J A and Kagan D 2021 Quantum Conditional Probabilities and New Measures of Quantum Information arXiv:2109.07447 [quant-ph]
- [38] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
- [39] Ozawa M 1997 Quantum State Reduction and the Quantum Bayes Principle arXiv:quant-ph/9705030
- [40] Peres A 1993 *Quantum Theory: Concepts and Methods* (Dordrecht: Kluwer Academic)
- [41] Park J L and Margenau H 1968 Simultaneous measurability in quantum theory *Int. J. Theor. Phys.* **1** 211–83
- [42] Muynck W M 2002 *Foundations of Quantum Mechanics, an Empiricist Approach* (Dordrecht: Kluwer Academic Publishers)
- [43] Busch P 1986 Unsharp reality and joint measurements for spin observables *Phys. Rev. D* **33** 2253–61
- [44] Yu S, Liu N, Li L and Oh C H 2010 Joint measurement of two unsharp observables of a qubit *Phys. Rev. A* **81** 062116
- [45] Muynck W M and Abu-Zeid O 1984 On an alternative interpretation of the Bell inequalities *Phys. Lett. A* **100** 485–9
- [46] Khrennikov A 2000 Non-Kolmogorov probability models and modified Bell's inequality arXiv:0003.017 [quant-ph]
- [47] Hess K and Philipp W 2004 Bell's theorem: Critique of proofs with and without inequalities arXiv:0410.015 [quant-ph]
- [48] Matzkin A 2008 Is Bell's theorem relevant to quantum mechanics? On locality and non-commuting observables arXiv: 0802.0613 [quant-ph]
- [49] Nieuwenhuizen T M 2011 Is the Contextuality Loophole Fatal for the Derivation of Bell Inequalities? *Found. Phys.* **41** 580
- [50] Khrennikov A 2014 CHSH inequality: Quantum probabilities as classical conditional probabilities arXiv:1406.4886 [quant-ph]
- [51] Muynck W M 2001 Interpretations of quantum mechanics, and interpretations of violation of Bell's inequality arXiv:quant-ph/0102066
- [52] Luis A and Ares L 2020 Apparatus contribution to observed nonclassicality *Phys. Rev. A* **102**, 022222
- [53] Walker N G and Carroll J E 1986 Multiport homodyne detection near the quantum noise limit *Opt. Quantum Electron.* **18**, 355–363
- [54] Walker N G 1987 Quantum theory of multiport optical homodyning *J. Mod. Opt.* **34**, 15–60