

NEGATIVE PROBABILITY

ANDREAS BLASS AND YURI GUREVICH

ABSTRACT. The uncertainty principle asserts a limit to the precision with which position x and momentum p of a particle can be known simultaneously. You may know the probability distributions of x and p individually but the joint distribution makes no physical sense. Yet Wigner exhibited such a joint distribution $f(x, p)$. There was, however, a little trouble with it: some of its values were negative. Nevertheless Wigner's discovery attracted attention and found applications. There are other joint distributions, all with negative values, which produce the correct marginal distributions of x and p . But only Wigner's distribution produces the correct marginal distributions for all linear combinations of position and momentum. We offer a simple proof of the uniqueness and discuss related issues.

1. INTRODUCTION

“Trying to think of negative probabilities,” wrote Richard Feynman, “gave me a cultural shock at first” [6]. Yet quantum physicists tolerate negative probabilities. Feynman himself studied, in the cited paper, a probabilistic trial with four outcomes with probabilities 0.6, -0.1 , 0.3 and 0.2.

We were puzzled. The standard interpretation of probabilities defines the probability of an event as the limit of its relative frequency in a large number of trials. “The mathematical theory of probability gains practical value and an intuitive meaning in connection with real or conceptual experiments” [5, §I.1]. Negative probabilities are obviously inconsistent with the frequentist interpretation. Of course, that interpretation comes with a tacit assumption that every outcome is observable. In quantum physics some outcomes may be unobservable. This weakens the frequentist argument against negative probabilities but does not shed much light on the meaning of negative probabilities.

In the discrete case, a probabilistic trial can be given just by a set of outcomes and a probability function that assigns nonnegative reals to outcomes. One can generalize the notion of probabilistic trial by allowing negative values of the probabilistic function. Feynman draws an analogy between this generalization and the generalization from

positive numbers, say of apples, to integers. But a negative number of apples may be naturally interpreted as the number of apples owed to other parties. We don't know any remotely natural interpretation of negative probabilities.

We attempted to have a closer look on what goes on.

Heisenberg's uncertainty principle asserts a limit to the precision with which position x and momentum p of a particle can be known simultaneously: $\sigma_x \sigma_p \geq \hbar/2$ where σ_x, σ_p are the standard deviations and \hbar is the (reduced) Planck constant. You may know the probability distributions (or the density function or the probability function; we will use the three terms as synonyms) of x and p individually but the joint probability distribution with these marginal distributions of x and p makes no physical sense¹. Does it make mathematical sense? More exactly, does there exist a joint distribution with the given marginal distributions of x and p .

In 1932, Eugene Wigner exhibited such a joint distribution [20]. There was, however, a little trouble with Wigner's function. Some of its values were negative. The function, Wigner admits, "cannot be really interpreted as the simultaneous probability for coordinates and momenta." But this, he continues, "must not hinder the use of it in calculations as an auxiliary function which obeys many relations we would expect from such a probability" [20]. Probabilistic functions that can take negative values became known as *quasi-probability distributions*.

Richard Feynman described a specific quasi-probability distribution for discrete quantities, two components of a particle's spin [6]. The uncertainty principle implies that these two quantities can't have definite values simultaneously. So it seems plausible that an attempt to assign joint probabilities would again, as in Wigner's case, lead to something strange — like negative probabilities. "Trying to think of negative probabilities gave me a cultural shock at first . . . It is usual to suppose that, since the probabilities of events must be positive, a theory which gives negative numbers for such quantities must be absurd. I should show here how negative probabilities might be interpreted" [6]. His attitude toward negative probabilities echoes that of Wigner: a quasi-probability distribution may be used to simplify intermediate computations. The meaning of negative probabilities remains unclear. Those intermediate computations may not have any physical sense. But if the final results make physical sense and can be tested then the use of a quasi-probability is justified.

¹For general information about joint probability distributions and their marginal distributions see [5, §IX.1]

It bothered us that both Wigner and Feynman apparently pull their quasi-probability distributions from thin air. In particular, Wigner writes that his function “was found by L. Szilárd and the present author some years ago for another purpose” [20], but he doesn’t give a reference, and he doesn’t give even a hint about what that other purpose was. He also says that there are lots of other functions that would serve as well, but none without negative values. He adds that his function “seemed the simplest.”

We investigated the matter and made some progress. We found a characterization of Wigner’s function that might be considered objective.

Proposition 1 (Wigner Uniqueness). *Wigner’s function is the unique quasi-distribution on the phase space that yields the correct marginal distributions not only for position and momentum but for all their linear combinations.*

Quisani²: Wait, I don’t understand the proposition. Wigner’s function is not a true distribution, so the notion of marginal distribution of Wigner’s function isn’t defined. Also, what does it mean for a marginal to be correct?

Authors³: The standard definition of marginals works also for quasi-probability distributions. A marginal distribution is correct if it coincides with the prediction of quantum mechanics. We’ll return to these issues in §3 and §4 respectively.

Q: To form a linear combination $ax + bp$ of position x and momentum p you add x , which has units of length like centimeters, and p , which has momentum units like gram centimeters per second. That makes no sense.

A: We are adding ax and bp . Take a to be a momentum and b to be a length; then both ax and bp have units of action (like gram centimeters squared per second), so they can be added. If you want to make $ax + bp$ a pure number, divide by \hbar .

²Readers of this column may remember Quisani, an inquisitive former student of the second author.

³speaking one at a time

Q: Finally, is it obvious that Wigner’s quasi-distribution is not determined already by the correct marginal distributions for just the position and momentum, without taking into account other linear combinations?

A: This is obvious. There are modifications of Wigner’s quasi-distribution that still give the correct marginal distributions for position and momentum. For an easy example, choose a rectangle R centered at the origin in the (x, p) phase plane and modify Wigner’s $f(x, p)$ by adding a constant c (resp. subtracting c) when $(x, p) \in R$ and the signs of x and p are the same (resp. different). For a smoother modification, you could add $cxp \exp(-ax^2 - bp^2)$ where a, b, c are positive constants (of appropriate dimensions).

The idea to consider linear combinations of position and momentum came from Wigner’s paper [20] where he mentions that projections to such linear combinations preserve the expectations. In fact, the projections give rise to the correct marginals. This led us to the proposition.

In the case of Feynman’s quasi-distribution mentioned above, one can’t use linear combinations of those two spin components to characterize the distribution. Nor is there a characterization using the spin component in yet another direction. Furthermore, if we only require the correct marginal distributions for the x and z spins, then there are genuine, nonnegative joint probability distributions with those marginals.

Our investigation was supplemented by digging into the literature and talking to our colleagues, especially Nathan Wiebe. That brought us to “Quantum Mechanics in Phase Spaces: An Overview with Selected Papers” [22]. It turned out that there was another, much earlier, approach to characterizing the Wigner quasi-probability distribution. The main ingredient for that earlier approach is a proposal by Hermann Weyl [19, §IV.14] for associating Hermitian operators on L^2 to well-behaved functions $g(x, p)$ of position and momentum. José Enrique Moyal used Weyl’s correspondence to characterize Wigner’s quasi-distribution in terms of only the expectation values but for a wider class of functions rather than the marginal distributions for just the linear functions of position and momentum [10]. There is a trade-off here. The class of functions is wider but the feature to match is narrower.

George Baker proved that any quasi-distribution on the position-momentum phase-space, satisfying his “quasi-probability distributional formulation of quantum mechanics,” is the Wigner function [2]. The

problem of an objective characterization of Wigner’s function also attracted the attention of Wigner himself [21, 12].

The volume [22] does not contain “our” characterization of Wigner’s function but it is known and due to Jacqueline and Pierre Bertrand [3]. They found an astute name for the approach: tomographic. The tomographic approach gives an additional confirmation of the fact that behavior of our quantum system is not classical. The approach can be used to establish that the behavior of some other quantum systems is not classical. It has indeed been used that way in quantum optics; see [15, 13] for example.

Still, in our judgment, our proof of the uniqueness of Wigner’s function is simpler and more direct than any other in the literature, and so we present it in §4. In section §5 we establish Moyal’s characterization of Wigner’s function.

§6 contains a cursory discussion related to Feynman’s four-outcome quasi-distribution. All our observations on that issue happened to be known as well, but we have yet to research the history of the negative probabilities in the discrete case. We intend to address the discrete case elsewhere.

Q: So what is the meaning of negative probability?

A: We don’t know.

Q: The use of negative probabilities to validate the quantum character of a quantum system reminds me proofs by contradiction. Assume that the behavior is classical, produce a unique joint distribution, prove the existence of negative values and establish a contradiction. If this is the only use of negative probabilities then there is no need to interpret them semantically.

A: There are some attempts to use negative probabilities as a measure of “quantumness” [17, 18]. We think that the jury is still out.

Q: I have yet another question. Recently, in this very column, Samson Abramsky wrote about contextuality which is another manifestation of non-classical behavior of quantum systems [1]. I wonder what is the relation, if any, between negative probabilities and contextuality.

A: The discussion of that relation is beyond the scope of this paper. But please have a look at Robert Spekkens’s article [16] with a rather telling title “Negativity and contextuality are equivalent notions of nonclassicality.”

Acknowledgment. Many thanks to Nathan Wiebe who was our guide to the literature and the state of art on quasi-probabilities.

2. PRELIMINARIES

We tried to make the paper as accessible as possible; hence this section. We still assume some familiarity with mathematical analysis. By default in this paper integrals are from $-\infty$ to $+\infty$. The baby quantum theory that we use is covered in §3 of book [8] titled “A first approach to quantum mechanics,”.

2.1. Fourier transform. The forward Fourier transform sends a function $f(x)$ to

$$\hat{f}(\xi) = \frac{1}{\sqrt{2\pi}} \int f(x) e^{-i\xi x} dx.$$

and the inverse Fourier transform sends a function $g(\xi)$ to

$$\check{g}(x) = \frac{1}{\sqrt{2\pi}} \int g(\xi) e^{i\xi x} d\xi,$$

Mathematically x and ξ are real variables. In applications, the dimension of ξ is the inverse of that of x so that ξx is a pure number.

The forward and inverse Fourier transforms are defined also for functions of several variables. In particular,

$$\begin{aligned} \hat{f}(\xi, \eta) &= \frac{1}{2\pi} \iint f(x, y) e^{-i(\xi x + \eta y)} dx dy, \\ \check{g}(x, y) &= \frac{1}{2\pi} \iint g(\xi, \eta) e^{i(\xi x + \eta y)} d\xi d\eta. \end{aligned}$$

Q: What about the convergence of the integrals? Are you going to ignore such details?

A: Yes, we are going to ignore such details. But Fourier transforms are used, with full mathematical rigor, even in some situations where the integrals don't converge.

Q: I do not understand this.

A: The idea is to first define the Fourier transform as an operator on nice functions in $L^2(\mathbb{R})$, for which the integrals clearly converge. Informally a function $f(x)$ is nice if it and its derivatives $f'(x), f''(x), f'''(x), \dots$ approach zero very rapidly as $x \rightarrow \infty$. The Fourier transform is an isometry on these nice functions, and the nice functions are dense in $L^2(\mathbb{R})$, so the isometry extends to all of L^2 . Details can be found in books on real analysis, like [9] and [14]; alternatively, see [8, Appendix A.3.2].

2.2. Dirac's delta function. Dirac's δ -function is a generalized function such that for any nice function f ,

$$\int f(x)\delta(x)dx = f(0).$$

It follows that

$$\int f(x)\delta(x-a)dx = \int f(x+a)\delta(x)dx = f(a).$$

Some divergent integrals, e.g. $\int e^{itx}dt$, can be seen as generalized functions in that sense. In fact, as generalized functions,

$$\int e^{itx}dt = 2\pi\delta(x).$$

Indeed,

$$\begin{aligned} \int dx f(x) \int e^{itx} dt &= \sqrt{2\pi} \int dt \frac{1}{\sqrt{2\pi}} \int f(x) e^{itx} dx \\ &= \sqrt{2\pi} \int \check{f}(t) dt \\ &= 2\pi \cdot \frac{1}{\sqrt{2\pi}} \int \check{f}(t) e^{-it0} dt = 2\pi f(0). \end{aligned}$$

Q: Are these nice functions the same as the nice functions mentioned earlier.

A: Yes, they are.

2.3. Exponential operators. The exponential e^O of an operator O over a topological vector space is the operator

$$e^O = \sum_{k=0}^{\infty} \frac{O^k}{k!} = I + O + \frac{1}{2}O^2 + \frac{1}{6}O^3 + \dots$$

If $(X\psi)(x) = x \cdot \psi(x)$ then $(e^X\psi)(x) = e^x\psi$, because

$$(e^X\psi)(x) = \sum_{k=0}^{\infty} \frac{1}{k!} (X^k\psi)(x) = \psi \cdot \sum_{k=0}^{\infty} \frac{1}{k!} x^k = \psi \cdot e^x.$$

If D is the derivative operator $\frac{d}{dx}$, then $e^{aD}\psi(x) = \psi(x+a)$. Indeed,

$$\begin{aligned} e^{aD}\psi(x) &= \sum_{k=0}^{\infty} \frac{(aD)^k\psi(x)}{k!} = \sum_{k=0}^{\infty} \frac{D^k f(x)}{k!} a^k \\ &= \psi(x) + \frac{\psi'(x)}{1!}a + \frac{\psi''(x)}{2!}a^2 + \frac{\psi'''(x)}{3!}a^3 + \dots \end{aligned}$$

which is the Taylor series of $\psi(x+a)$ around point x . (Think of a as Δx .)

Q: What functions f are you talking about? The Taylor series expansion of f suggests that f is analytic, that is real-analytic.

A: Our intention is that f ranges over L^2 . By the proof above, e^{aD} is a shift $f(x) \mapsto f(x+a)$ on analytic functions. In particular, e^{aD} is a shift on Gaussian functions

$$\exp\left(-\frac{(x-b)^2}{2c^2}\right).$$

But Gaussian functions span a dense subspace of $L^2(\mathbb{R})$, and there is a unique continuous extension of e^{aD} to L^2 , namely the shift $f(x) \mapsto f(x+a)$.

Q: The exponential e^O has got to be a partial operator in general.

A: Yes, $e^O(x)$ is defined whenever the operators O^k are defined at x , and the series $\sum_{k=0}^{\infty} \frac{O^k(x)}{k!}$ converges.

3. JOINT-TO-MARGINAL LEMMA

Let $f(x,p)$ be an ordinary probability distribution or a quasi-distribution on \mathbb{R}^2 . For any $z = ax + bp$ where a, b are not both zero, the marginal distribution $g(z)$ of z can be defined thus:

$$g(z) = \begin{cases} \frac{1}{b} \int f(x, \frac{1}{b}(z - ax)) dx & \text{if } b \neq 0 \\ \frac{1}{a} \int f(\frac{1}{a}(z - by), p) dp & \text{otherwise} \end{cases}$$

Here's a justification in the case $b \neq 0$. We have

$$p = \frac{1}{b}(z - ax),$$

$$dp = \frac{1}{b}(dz - a dx),$$

$$f(x,p) dx dp = f(x, \frac{1}{b}(z - ax)) \frac{1}{b} dx dz.$$

We are relying here on the formalism of differential 2-forms [7] for area elements, so that $dx dz$ really means $dx \wedge dz$ and we have used that $dx \wedge dx = 0$. The use of differential forms makes computations like this easier, and it fits well with physics, e.g., with Maxwell's equations and with general relativity. One could, however, avoid differential forms

here and get the same result by considering the Jacobian determinant of the change of variables.

For any real $u \leq v$, the probability that $u \leq z \leq v$ should be

$$\begin{aligned} \int_u^v g(z) dz &= \iint_{u \leq ax+bp \leq v} f(x, p) dx dp \\ &= \iint_{u \leq ax+bp \leq v} \frac{1}{b} f\left(x, \frac{1}{b}(z - ax)\right) dx dz \\ &= \int_u^v dz \int_{-\infty}^{\infty} \frac{1}{b} f\left(x, \frac{1}{b}(z - ax)\right) dx. \end{aligned}$$

Since the first and last expressions coincide for all $u \leq v$, we have

$$g(z) = \frac{1}{b} \int f\left(x, \frac{1}{b}(z - ax)\right) dx.$$

Lemma 2 (J2M). *For any a, b not both zero, the following statements are equivalent.*

- (1) $g(z)$ is the marginal distribution of $z = ax + bp$.
- (2) $\hat{g}(\zeta) = \sqrt{2\pi} \cdot \hat{f}(a\zeta, b\zeta)$.

Proof. To prove (1)→(2), suppose (1) and compare the forward Fourier transforms of g and f :

$$\begin{aligned} \hat{g}(\zeta) &= \frac{1}{\sqrt{2\pi}} \int g(z) e^{-i\zeta z} dz \\ &= \frac{1}{\sqrt{2\pi}} \iint f\left(x, \frac{1}{b}(z - ax)\right) e^{-i\zeta z} \frac{1}{b} dx dz \\ &= \frac{1}{\sqrt{2\pi}} \iint f(x, p) e^{-i\zeta(ax+bp)} dx dp. \end{aligned}$$

$$\hat{f}(\xi, \eta) = \frac{1}{2\pi} \iint f(x, p) e^{-i(\xi x + \eta p)} dx dp.$$

We have $\hat{g}(\zeta) = \sqrt{2\pi} \hat{f}(a\zeta, b\zeta)$.

To prove (2)→(1), suppose (2) and use the implication (1)→(2). If h is the marginal distribution of $z = ax + by$ then

$$\hat{h}(\zeta) = \sqrt{2\pi} \cdot \hat{f}(a\zeta, b\zeta) = \hat{g}(\zeta),$$

and therefore $g = h$. □

Corollary 3. *For any real α, β not both zero, $\hat{f}(\alpha, \beta) = \frac{1}{\sqrt{2\pi}} \hat{g}(\zeta)$ where $g(z)$ is the marginal distribution for the linear combination $z = ax + bp$ such that $\alpha = a\zeta$, $\beta = b\zeta$ for some ζ .*

4. WIGNER UNIQUENESS

The purpose of this section is to prove the Wigner Uniqueness proposition. For simplicity we work with one particle moving in one dimension, but everything we do in this section generalizes in a routine way to more particles in more dimensions.

In classical mechanics, the position x and momentum p of the particle determine its current state. The set of all possible states is the phase space of the particle. By Corollary 3, an ordinary distribution $f(x, p)$ on the phase space is uniquely determined by its marginal distributions for all linear combinations $ax + bp$ where a, b are not both zero.

In the quantum case, a state of the particle is given by a normalized (to norm 1) vector $|\psi\rangle$ in $L^2(\mathbb{R})$. The position and momentum are given by Hermitian operators X and P where

$$(X\psi)(x) = x \cdot \psi(x) \quad \text{and} \quad (P\psi)(x) = -i\hbar \frac{d\psi}{dx}(x).$$

For any a, b not both zero, the linear combination $z = ax + by$ is given by the Hermitian operator $Z = aX + bP$. In a state $|\psi\rangle$, there is a probability distribution $g(z)$ (for the measurement) of the values of z . For a function $h(z)$ of z , the expectation of $h(z)$ is $\langle\psi|h(Z)|\psi\rangle$.

The following technical lemma plays a key role in our proof of the uniqueness of Wigner's quasi-distribution.

Lemma 4.

$$\langle\psi|e^{-i(\alpha X + \beta P)}|\psi\rangle = e^{i\alpha\beta\hbar/2} \int \psi^*(y) e^{-i\alpha y} \psi(y - \beta\hbar) dy.$$

Proof. We want to split the exponential into a factor with X times a factor with P . This is not as easy as it might seem, because X and P don't commute. We have, however, two pieces of good luck. First, there is Zassenhaus's formula, which expresses the exponential of a sum of non-commuting quantities as a product of (infinitely) many exponentials, beginning with the two that one would expect from the commutative case, and continuing with exponentials of nested commutators:

$$e^{A+B} = e^A e^B e^{-\frac{1}{2}[A,B]} \dots,$$

where the "... " refers to factors involving double and higher commutators.

Q: You gave no reference to Zassenhaus's paper.

A: Apparently, Zassenhaus never published this result, but there's a paper [4] that shows how to compute the

next terms. It also has a pointer to early uses of the formula.

The second piece of good luck is that $[X, P] = i\hbar I$, where I is the identity operator. (In the future, we'll usually omit writing I explicitly, so we'll regard this commutator as the scalar $i\hbar$.) Since that commutes with everything, all the higher commutators in Zassenhaus's formula vanish, so we can omit the “ \dots ” from the formula. We have

$$\langle \psi | e^{-i\alpha X - i\beta P} | \psi \rangle = \langle \psi | e^{-i\alpha X} e^{-i\beta P} e^{\alpha\beta[X, P]/2} | \psi \rangle.$$

The last of the three exponential factors here arose from Zassenhaus's formula as

$$-\frac{1}{2}[-i\alpha X, -i\beta P] = \frac{1}{2}\alpha\beta[X, P] = i\alpha\beta\hbar/2.$$

That factor, being a scalar, can be pulled out of the bra-ket. Taking into account §2.3,

$$\langle \psi | e^{-i(\alpha X + \beta P)} | \psi \rangle = e^{i\alpha\beta\hbar/2} \int \psi^*(y) e^{-i\alpha y} \psi(y - \beta\hbar) dy.$$

□

Now we are ready to prove the Wigner Uniqueness proposition. Suppose that a quasi-distribution $f(x, p)$ yields correct marginal distributions for all linear combinations of position and momentum. For any real α, β not both zero, let a, b, g, ζ be as in Corollary 3. Then

$$\begin{aligned} \hat{f}(\alpha, \beta) &= \frac{1}{\sqrt{2\pi}} \hat{g}(\zeta) = \frac{1}{2\pi} \int g(z) e^{-i\zeta z} dz \\ (1) \qquad &= \frac{1}{2\pi} \langle e^{-i\zeta Z} \rangle = \frac{1}{2\pi} \langle \psi | e^{-i\zeta Z} | \psi \rangle \\ &= \frac{1}{2\pi} \langle \psi | e^{-i\zeta(aX + bP)} | \psi \rangle. \end{aligned}$$

By Lemma 4,

$$(2) \qquad \hat{f}(\alpha, \beta) = \frac{e^{i\alpha\beta\hbar/2}}{2\pi} \int \psi^*(y) e^{-i\alpha y} \psi(y - \beta\hbar) dy.$$

To get $f(x, p)$, apply the (two-dimensional) inverse Fourier transform.

$$f(x, p) = \frac{1}{(2\pi)^2} \iiint \psi^*(y) e^{-i\alpha y} e^{i\alpha\beta\hbar/2} \psi(y - \beta\hbar) e^{i\alpha x} e^{i\beta p} dy d\alpha d\beta.$$

Collecting the three exponentials that have α in the exponent, and noting that α appears nowhere else in the integrand, perform the integration over α and (recall §2.2) get a Dirac delta function:

$$\int e^{-i\alpha(y - \frac{\beta\hbar}{2} - x)} d\alpha = 2\pi\delta(y - x - \frac{\beta\hbar}{2}).$$

That makes the integration over y trivial, and what remains is

$$(3) \quad f(x, p) = \frac{1}{2\pi} \int \psi^*(x + \frac{\beta\hbar}{2})\psi(x - \frac{\beta\hbar}{2})e^{i\beta p} d\beta,$$

which is Wigner's quasi-distribution.

To check that Wigner's quasi-distribution yields correct marginal distribution note that the derivation of (3) from (1) is reversible. This completes the proof of the Wigner Uniqueness proposition.

5. WEYL'S CORRESPONDENCE

There is another approach to characterizing the Wigner quasi-probability distribution, using the expectation values for a wide class of functions rather than the marginal distributions for just the linear functions of position and momentum. The main ingredient for this approach is a proposal by Hermann Weyl [19, §IV.14] for associating a Hermitian operator on $L^2(\mathbb{R})$ to any (well-behaved) function $g(x, p)$ of position and momentum. Weyl's proposal is to first form the Fourier transform $\hat{g}(\alpha, \beta)$ of $g(x, p)$, and then apply the inverse Fourier transform with the Hermitian operators X and P in place of the classical variables x and p . Thus, the Weyl correspondence associates to $g(x, p)$ the operator

$$g(X, P) = \frac{1}{2\pi} \iint \hat{g}(\alpha, \beta) e^{i(\alpha X + \beta P)} d\alpha d\beta.$$

If one grants that this is a reasonable way of converting phase-space functions $g(x, p)$ to operators $g(X, P)$, then a desirable property of a phase-space quasi-probability distribution $f(x, p)$ would be that the expectation of $g(X, P)$ in a quantum state $|\psi\rangle$ is the same as the expectation of $g(x, p)$ under $f(x, p)$. We shall show that the Wigner distribution is uniquely characterized by enjoying this desirable property for all well-behaved g .

Indeed, the expectation of $g(X, P)$ in state $|\psi\rangle$ is

$$\langle\psi|g(X, P)|\psi\rangle = \frac{1}{2\pi} \iint \hat{g}(\alpha, \beta) \langle\psi|e^{i(\alpha X + \beta P)}|\psi\rangle d\alpha d\beta,$$

and the expectation of $g(x, p)$ under the distribution $f(x, p)$ is

$$\iint g(x, p)f(x, p) dx dp = \iint \hat{g}(\alpha, \beta)\hat{f}(\alpha, \beta) d\alpha d\beta.$$

This last equation is a consequence of the fact, mentioned in §2.1, that the Fourier transform is a unitary operator and therefore preserves the inner product structure of $L^2(\mathbb{R})$. Since these two expectations agree for all (well-behaved) g ,

$$\hat{f}(\alpha, \beta) = \frac{1}{2\pi} \langle \psi | e^{i(\alpha X + \beta P)} | \psi \rangle.$$

But this is the part of equation(1) that was used to derive Wigner's formula (3).

Q: I wonder how Weyl arrived at his proposal.

A: Weyl presents his proposal in [19] without any motivation, so we don't know how he came up with it, but we can speculate. The title of [19] indicates that Weyl was working in a group-theoretic context. As a result, the Fourier transform, expressing functions on \mathbb{R} as combinations of the characters $e^{i\alpha x}$ of the group $(\mathbb{R}, +)$ would be in the forefront of his considerations. Now consider his goal — to somehow convert a classical function $g(x, p)$ into an operator. Roughly speaking, he would want to substitute the operators X and P for the classical variables x and p . An obvious difficulty is that the same functions $g(x, p)$ might have two different expressions, for example $xp = px$, which are no longer equivalent when operators are substituted, $XP \neq PX$. So it is reasonable to try to choose, from the many expressions for a function $g(x, p)$, one particular, reasonably canonical expression, into which one can substitute X and P . The Fourier expansion, $\int \hat{g}(\alpha, \beta) \exp(i(\alpha x + \beta p)) dx dp$ has those properties. It depends only on the function g , not on how one chooses to express it, and there is no problem substituting X and P for x and p .

6. FEYNMAN AND SPINS

Richard Feynman studied “an analogue of the Wigner function for a spin $\frac{1}{2}$ system or other two state system” [6]. He chose the z and x components of the spin to serve as the analogs of the position and momentum in Wigner's formula.

Q: His case should be much simpler than Wigner's case.

A: Not necessarily. While the commutator $[X, P]$ is a scalar, the commutator of the z and x components of a spin is the y -component times $i\hbar$. This is but one of several complications.

Q: Is Feynman's quasi-distribution determined by the correct marginals for all linear combinations of the x and z spins?

A: That question sounds reasonable until you look at it a little more closely. To fix notation, let's describe spin by means of the standard Pauli matrices

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The usual matrix representation of the spins for a spin $\frac{1}{2}$ particle is given by these matrices divided by 2, but it's convenient to skip those extra factors $\frac{1}{2}$; if you like, imagine that we are measuring angular momentum in units of $\hbar/2$ instead of \hbar .

So each of our matrices has eigenvalues ± 1 , with $+1$ meaning spin along the corresponding positive axis and -1 along the corresponding negative axis. For example, the two basis states of spin up and spin down along the z axis are the eigenvectors for eigenvalues 1 and -1 of Z ; equivalently, they correspond to eigenvalues 1 and 0 for $(I + Z)/2$, where I is the identity operator. This point of view is useful because it implies that, in any state $|\psi\rangle$, $(1 + \langle Z \rangle)/2$ is the probability that the z spin is up. Here, as before, angle brackets denote expectations. Similarly, $(1 - \langle Z \rangle)/2$ is the probability that the z spin is down. Of course, analogous formulas apply to the x and y components of the spin.

Feynman, in analogy to Wigner, introduces a quasi-probability distribution f for the pair of non-commuting observables Z and X . So f has four components, f_{++}, f_{+-}, f_{-+} , and f_{--} , as the quasi-probability of Z and X having the values ± 1 given by the subscripts of f .

Now let's look at a linear combination of Z and X , for example the simplest nontrivial one, $Z + X$. From the point of view of quasi-probabilities,

- with probability f_{++} , Z has value 1 and X has value 1 , so $Z + X$ has value 2 ,
- with probability f_{+-} , Z has value 1 and X has value -1 , so $Z + X$ has value 0 ,
- with probability f_{-+} , Z has value -1 and X has value 1 , so $Z + X$ has value 0 , and
- with probability f_{--} , Z has value -1 and X has value -1 , so $Z + X$ has value -2 .

Altogether, the possible values of $Z + X$ are 2, 0, and -2 , with quasi-probabilities f_{++} , $f_{+-} + f_{-+}$, and f_{--} , respectively.

So your proposed analog of the result for Wigner's distribution would assume that f is chosen so that these probabilities agree, in a given state, with the probabilities computed by quantum mechanics. But such agreement is impossible, because, according to quantum mechanics, the possible values of $Z + X$ are the eigenvalues of this operator, namely $\pm\sqrt{2}$, which are completely different from the 2, 0, -2 arising from the quasi-probabilities. It is easy to check that the possible values of any nontrivial linear combination of Z and X are completely different from the values arising from the quasi-probabilities.

Q: OK, let's require the minimum that Feynman obviously intended, namely that f should produce the correct marginal distributions of Z and X . Does this determine f uniquely?

A: At first sight, this looks promising. Requiring the correct marginals for the two variables, each having two possible values, gives us four equations for the four unknown components $f_{\pm\pm}$ of f :

$$\begin{aligned} f_{++} + f_{+-} &= \frac{1}{2}(1 + \langle Z \rangle) \\ f_{-+} + f_{--} &= \frac{1}{2}(1 - \langle Z \rangle) \\ f_{++} + f_{-+} &= \frac{1}{2}(1 + \langle X \rangle) \\ f_{+-} + f_{--} &= \frac{1}{2}(1 - \langle X \rangle). \end{aligned}$$

But there's redundancy in the equations; only three of them are independent, so there's one free parameter in the general solution. In fact, it's easy to write down the general solution:

$$\begin{aligned} f_{++} &= \frac{1}{4}(1 + \langle Z \rangle + \langle X \rangle + t) \\ f_{+-} &= \frac{1}{4}(1 + \langle Z \rangle - \langle X \rangle - t) \\ f_{-+} &= \frac{1}{4}(1 - \langle Z \rangle + \langle X \rangle - t) \\ f_{--} &= \frac{1}{4}(1 - \langle Z \rangle - \langle X \rangle + t), \end{aligned}$$

where t is arbitrary. Feynman's formulas correspond to $t = \langle Y \rangle$ but we see no reason to prefer $\langle Y \rangle$ over, for example, $-\langle Y \rangle$.

Q: Put the freedom in choosing t to some use. How about minimizing the negativity in f ? In other words, adjust t to bring f as close as possible to being a genuine probability distribution.

A: That idea works better than we originally expected. One can get rid of the negativity altogether. For each state $|\psi\rangle$, there is a choice of t that makes all four components of f nonnegative.

Indeed, write down the four inequalities $f_{\pm\pm} \geq 0$ using the formulas above for these $f_{\pm\pm}$'s. Solve each one for t . You find two lower bounds on t , namely

$$\begin{aligned} -1 - \langle Z \rangle - \langle X \rangle & \text{ (from } f_{++} \geq 0) \\ -1 + \langle Z \rangle + \langle X \rangle & \text{ (from } f_{--} \geq 0), \end{aligned}$$

and two upper bounds, namely

$$\begin{aligned} 1 + \langle Z \rangle - \langle X \rangle & \text{ (from } f_{+-} \geq 0) \\ 1 - \langle Z \rangle + \langle X \rangle & \text{ (from } f_{-+} \geq 0). \end{aligned}$$

An appropriate t exists if and only if both of the lower bounds are less than or equal to both of the upper bounds. That gives four inequalities, which simplify to $-1 \leq \langle Z \rangle \leq 1$ and $-1 \leq \langle X \rangle \leq 1$. But these are always satisfied, because the eigenvalues of Z and X are ± 1 .

Q: I am confused. The uncertainty principle asserts that you cannot measure Z and X at once. Accordingly one would expect that the joint probability distribution $f_{\pm\pm}$ should not exist or, as in Wigner's case, should have at least one negative value.

A: The relevant difference between Wigner's and Feynman's cases seems to be this. In Wigner's case, there is naturally a rich set of marginals that the joint probability distribution is supposed to produce, namely the probability distributions of all linear combinations of the position x and the momentum p of the particle. In Feynman's case, the natural set of marginals is too poor, just the probability distributions of Z and X .

Q: Did Feynman find a good use for quasi-probabilities?

A: He introduced negative probabilities in connection to a problem of infinities in quantum field theory. "Unfortunately I never did find out how to use the freedom of allowing probabilities to be negative to solve the original problem of infinities in quantum field theory!" [6].

Q: Still, the idea to use quasi-probability distributions to simplify intermediate computations looks attractive to me.

A: You are in good company.

REFERENCES

- [1] Samson Abramsky, “Contextual semantics: From quantum mechanics to logic, databases, constraints, and complexity,” in *Logic in Computer Science Column*, Bulletin of EATCS 113, 26 pages (2014).
- [2] George A. Baker, Jr., “Formulation of quantum mechanics based on the quasi-probability distribution induced on phase space,” *Physical Review* 109:6 (1958) 2198–2206. Reprinted in [22], 235–243.
- [3] Jacqueline Bertrand and Pierre Bertrand, “A tomographic approach to Wigner’s function,” *Foundations of Physics* 17:4 (1987) 397–405.
- [4] Fernando Casas, Ander Murua, and Mladen Nadinic, “Efficient computation of the Zassenhaus formula,” *Computer Physics Communications* 183 (2012) 2386–2391.
- [5] William Feller, *An Introduction to Probability Theory and Its Applications, Volume 1*, John Wiley (1950), 3rd edition (1968).
- [6] Richard P. Feynman, “Negative probabilities,” in *Quantum Implications: Essays in Honor of David Bohm*, ed. F. D. Peat and B. Hiley, Routledge & Kegan Paul (1987) 235–248. Reprinted in [22], 426–439.
- [7] Harley Flanders, *Differential Forms with Applications to the Physical Sciences*, Academic Press (1963). 2nd edition, Dover (1989).
- [8] Brian C. Hall, *Quantum Theory for Mathematicians*, Springer-Verlag, Graduate Texts in Mathematics 267 (2013).
- [9] John F. C. Kingman and S. James Taylor, *Introduction to Measure and Probability*, Cambridge Univ. Press (1966).
- [10] José E. Moyal, “Quantum mechanics as a statistical theory,” *Proc. Cambridge Phil. Soc.* 45 (1949) 99–124. Reprinted in [22], 167–192.
- [11] G. Nogues, A. Rauschenbeutel, S. Osnaghi, P. Bertet, M. Brune, J. M. Raimond, S. Haroche, L. G. Lutterbach, and L. Davidovich, “Measurement of a negative value for the Wigner function of radiation,” *Physical Review A* 62, 054101 (2000).
- [12] Robert F. O’Connell and Eugene P. Wigner, “Quantum-mechanical distribution functions: conditions for uniqueness,” *Physics Letters* 83A:4 (1981), 145–148.
- [13] Alexei Ourjountsev, Rosa Tualle-Broui, and Philippe Grangier, “Quantum homodyne tomography of a two-photon Fock state,” arXiv:quant-ph/0603284 (2006).
- [14] Walter Rudin, *Real and Complex Analysis*, McGraw-Hill (1966). 3rd edition (1987).
- [15] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, “Measurement of Wigner distribution and the density matrix of a light mode using optical homodyne tomography,” *Physical Review Letters* 70:9 1244–1247 (1993).
- [16] Robert W. Spekkens, “Negativity and contextuality are equivalent notions of nonclassicality,” arXiv:0710.5549v2 (2008).
- [17] Victor Veitch, Christopher Ferrie, David Gross, and Joseph Emerson, “Negative quasi-probability as a resource for quantum computation,” arXiv:1201.1256 (2012).
- [18] Victor Veitch, S. A. Hamed Mousavian, Daniel Gottesman, and Joseph Emerson, “The resource theory of stabilizer quantum computation,” *New Journal of Physics* 16 013009 (Jan. 9, 2014).

- [19] Hermann Weyl, *Quantenmechanik und Gruppentheorie*, Zeitschrift für Physik 46 1–46. Reprinted in [22], 45–90.
- [20] Eugene P. Wigner, “On the quantum correction for thermodynamic equilibrium,” *Physical Review* 40 (1932) 749–759. Reprinted in [22], 100–110.
- [21] Eugene P. Wigner, “Quantum mechanical distribution functions revisited,” in W. Yourgrau and A. van der Merwe (eds), *Perspective in Quantum Theory*, MIT Press, MA, 1971) 25–36 (1971).
- [22] Cosmas K. Zachos, David B. Fairlie, and Thomas L. Curtright (eds.), *Quantum Mechanics in Phase Space: An Overview with Selected Papers*, World Scientific Series in 20th Century Physics, vol. 34 (2005).

MATHEMATICS DEPARTMENT, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109–1043, U.S.A.

E-mail address: `ablass@umich.edu`

MICROSOFT RESEARCH, ONE MICROSOFT WAY, REDMOND, WA 98052, U.S.A.

E-mail address: `gurevich@microsoft.com`