Math. Struct. in Comp. Science (2014), vol. 24, iss. 3, e240305, 18 pages. © Cambridge University Press 2014
doi:10.1017/S0960129512000709

Joint distributions, the uncertainty principle and positive distributions[†]

LEON COHEN

City University of New York, 695 Park Avenue, New York, NY 10065, U.S.A.
Email: Leon.Cohen@hunter.cuny.edu

Received 1 February 2011; revised 23 November 2011

We examine the construction of joint probabilities for non-commuting observables. We show that there are indications in standard quantum mechanics that imply the existence of conditional expectation values, which in turn implies the existence of a joint distribution. We also argue that the uncertainty principle has no bearing on the existence of joint distributions but only constrains the marginal distributions. In addition, we show that within classical probability theory there are mathematical quantities that are similar to quantum mechanical wave functions. This is shown by generalising a theorem of Khinchin on the necessary and sufficient conditions for a function to be a characteristic function.

1. Introduction

The fundamental basis of quantum mechanics is that it is a probability theory and that the results obtained are standard probabilistic quantities, such as expectation values and probability distributions. However, it calculates the probabilistic quantities in an unusual way that is radically different from standard probability theory. Quantum mechanical ideas and methods are seemingly foreign to standard probability theory as it deals with operators, wave functions, transformation theory, and yet, as just mentioned, it gives standard probabilistic results. One of the most widely accepted ideas of quantum mechanics is that there can be no joint distributions for physical quantities that are represented by non-commuting operators. We examine this issue and ask whether or not quantum mechanics allows the construction of proper joint distributions of non-commuting observables. Furthermore, we ask whether there are any quantities that are part of quantum mechanics that imply joint distributions. We also reverse the situation and argue that since the formalism of quantum mechanics is the most successful to date, perhaps it is standard probability theory that is the curiosity. Hence, we also discuss whether concepts like wave functions and operators exist within standard probability theory.

[†] This work was supported by the Office of Naval Research

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

2. Are there any results in standard quantum mechanics that suggest a joint distribution?

Quantum mechanical current is usually accepted as a legitimate quantum quantity, and we now examine how that concept can be generalised. We will argue that current is an average: specifically, the conditional average of momentum for a given position. If that is the case, we can ask for the standard deviation about that average. The standard deviation about a conditional average is called the conditional standard deviation. More generally, we will first discuss the general results in the following papers: Cohen (1996), Muga *et al.* (1998), Muga *et al.* (2005) and Sala *et al.* (1997). Suppose we have an observable represented by an operator $\mathbf{A}$ and a wave function $\psi(q)$. The global expectation value and standard deviation are then

$$\langle \mathbf{A} \rangle = \int \psi^*(q)\, \mathbf{A}\, \psi(q)\, dq \tag{1}$$

$$(\Delta \mathbf{A})^2 = \int \psi^*(q)\, (\mathbf{A} - \langle \mathbf{A} \rangle)^2\, \psi(q)\, dq, \tag{2}$$

respectively. It can be shown that this leads to

$$\langle \mathbf{A} \rangle_q = \int \left(\frac{\mathbf{A}\psi}{\psi} \right)_R |\psi(q)|^2\, dq \tag{3}$$

and

$$(\Delta \mathbf{A})^2 = \int \left(\frac{\mathbf{A}\psi}{\psi} \right)_I^2 |\psi(q)|^2\, dq + \int \left[\left(\frac{\mathbf{A}\psi}{\psi} \right)_R - \langle \mathbf{A} \rangle \right]^2 |\psi(q)|^2\, dq \tag{4}$$

where we have written

$$\frac{\mathbf{A}\psi}{\psi} = \left(\frac{\mathbf{A}\psi}{\psi} \right)_R + i \left(\frac{\mathbf{A}\psi}{\psi} \right)_I. \tag{5}$$

These are exact equations. What is the interpretation of Equations (3) and (4)? In particular, is there anything in standard probability theory like Equation (4)? Note that both terms are averages, the second term being the average of a quantity subtracted from the global mean.

We now argue that similar expressions exist in standard probability theory for a joint distribution. Suppose we have a probability distribution $P(a, b)$ of two variables a and b. For the marginals, we write

$$\int P(a, b)\, da = P(b) \tag{6}$$

$$\int P(a, b)\, db = P(a), \tag{7}$$

and the conditional probability distribution $P(a|b)$ is then given by

$$P(a|b) = \frac{P(a, b)}{P(b)}. \tag{8}$$

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

Therefore, the conditional average of a for a fixed b is

$$\begin{aligned}\langle a\rangle_b &= \int aP(a|b)\,da \\ &= \frac{1}{P(b)}\int aP(a,b)\,da\end{aligned} \tag{9}$$

and the conditional standard deviation is

$$\sigma^2_{a|b} = \frac{1}{P(b)}\int (a-\langle a\rangle)^2 P(a,b)\,da. \tag{10}$$

The global averages and global standard deviation are

$$\langle a\,\rangle = \iint P(a,b)\,da\,db \tag{11}$$

$$\sigma^2_a = \iint (a-\langle a\rangle)^2 P(a,b)\,da\,db. \tag{12}$$

We now ask, what is the relation between the conditional and global averages? It is straightforward to show that the standard deviation and conditional standard deviation are related by

$$\langle a\,\rangle = \int \langle a\rangle_b P(b)db \tag{13}$$

$$\sigma^2_a = \int \sigma^2_{a|b}\ P(b)\ db\ + \int \left(\langle a\rangle_b - \langle a\rangle\right)^2\ P(b)\ db \tag{14}$$

where $P(b)$ is the marginals distribution of b.

Comparing Equation (4) with Equation (14), we can see that they are identical in form. Now, of course, we cannot equate functions under the integral sign, but if we did, we would get interesting results. In particular,

$$\langle \mathbf{A}\rangle_q = \left(\frac{\mathbf{A}\psi}{\psi}\right)_R \tag{15}$$

and

$$\sigma^2_{a|q} = \left(\frac{\mathbf{A}\psi}{\psi}\right)^2_I. \tag{16}$$

Therefore, we can interpret these quantum results as a local average and conditional standard deviation. But for that to be the case, we must have an underlying joint probability.

As an example, consider the case of momentum. For the operator $\mathbf{A}$, we take the momentum operator $\mathbf{P}$ given by

$$\mathbf{P} = \frac{1}{i}\frac{d}{dq}. \tag{17}$$

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

We then have

$$\left(\frac{\mathbf{P}\psi}{\psi}\right)_R = \frac{1}{|\psi|^2}\frac{\hbar}{2i}\left(\psi^*\frac{d\psi}{dq} - \psi\frac{d\psi^*}{dq}\right) \tag{18}$$

$$\left(\frac{\mathbf{P}\psi}{\psi}\right)_I = \frac{1}{|\psi|^2}\frac{1}{2i}\left(\psi^*\frac{d\psi}{dq} + \psi\frac{d\psi^*}{dq}\right). \tag{19}$$

These expressions simplify considerably if we write the wave function in terms of its amplitude and phase

$$\psi(q) = R(q)e^{iS(q)}. \tag{20}$$

Then

$$\frac{\mathbf{P}\psi}{\psi} = \frac{\hbar}{i}\frac{R'(q)}{R(q)} + S'(q), \tag{21}$$

so

$$\begin{aligned}\langle \mathbf{P}\rangle_q &= \left(\frac{\mathbf{P}\psi}{\psi}\right)_R \\ &= \varphi'(q)\end{aligned} \tag{22}$$

and

$$\sigma^2_{p|q} = \left(\frac{R'(q)}{R(q)}\right)^2. \tag{23}$$

Hence, we see that quantum mechanics implies the existence of local quantities, and current has certainly been the historical example. In Section 5, we will give an alternative derivation based on windowed wave functions, which also implies the existence of conditional quantities.

3. Joint distributions, positivity, bilinearity and the Wigner impossibility theorem – general remarks

It is very often said that Wigner proved that positive joint distributions that satisfy the quantum mechanical marginals do not exist. This view has persisted, and it is our aim to examine it from the point of view of the actual construction of joint distributions. The concept of a joint distribution for position and momentum in quantum mechanics started in 1932 when Wigner (Wigner 1932) introduced a remarkable viewpoint, which has developed into what is now known as the phase-space formulation of quantum mechanics. At the same time, Kirkwood (Kirkwood 1933) also considered the issue and came up with a different distribution. Their aim was to produce a joint distribution of position and momentum that could be used as a practical calculational tool. In particular, they sought a $P(q,p)$ that satisfies the quantum mechanical marginal conditions

$$\int P(q,p)\,dp = |\psi(q)|^2 \tag{24}$$

$$\int P(q,p)\,dq = |\varphi(p)|^2 \tag{25}$$

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

where $\psi(q)$ is the wave function in position and $\varphi(p)$ is the momentum wave function related by

$$\psi(q) = \frac{1}{\sqrt{2\pi}} \int \varphi(p)\, e^{ipq} dp \tag{26}$$

$$\varphi(p) = \frac{1}{\sqrt{2\pi}} \int \psi(q)\, e^{-ipq}\, dq. \tag{27}$$

The marginals $|\psi(q)|^2$ and $|\varphi(p)|^2$ are the probability distributions of position and momentum. Wigner gave what is now called the Wigner distribution

$$P(q,p) = \frac{1}{2\pi} \int \psi^*(q - \frac{1}{2}\tau)\psi(q + \frac{1}{2}\tau)\, e^{-i\tau p}\, d\tau, \tag{28}$$

and almost at the same time Kirkwood gave

$$P(q,p) = \psi(q)\, \varphi^*(p) e^{-ipq}. \tag{29}$$

Other distributions have been given and studied, and we will discuss these in a later section. Of course, for a proper theory, we would also insist that

$$P(q,p) \geqslant 0. \tag{30}$$

In his original paper, Wigner pointed out that his choice of distribution may go negative, and then stated that

"It may be shown, however, that there does not exist an expression for P which is bilinear in ψ, satisfies ... [the marginals] and is everywhere positive."

Many subsequent authors generally assumed that Wigner had proved that positive joint distributions simply do not exist. Several years later, Wigner provided a proof of his assertion (Wigner 1971). While Wigner was clear in both papers that he is assuming bilinearity, many subsequent papers addressing the subject merely stated that Wigner had proved that no positive distributions can exist. However, as we will see, they do exist. The issue of bilinearity is crucial, and we will discuss this in detail in Section 7. A study of Wigner's proof was carried out in Mugur-Schächter (1977; 1979). Here we discuss in general terms the types of distributions that have been developed and the relation to the satisfaction of the marginals.

3.1. *Bilinear distributions satisfying the marginals*

All bilinear distributions satisfying the marginals can be generated by the following general form (Cohen 1966):

$$C(q,p) = \frac{1}{4\pi^2} \iiint \psi^*\left(u - \frac{1}{2}\tau\right)\, \psi\left(u + \frac{1}{2}\tau\right)\, \Phi(\theta,\tau)\, e^{-i\theta q - i\tau p + i\theta u}\, du\, d\tau\, d\theta \tag{31}$$

where $\Phi(\theta,\tau)$ is the kernel that characterises the distribution. To ensure the marginals are satisfied, we must have

$$\Phi(\theta,0) = \Phi(0,\tau) = 1.$$

If the kernel is taken to be independent of the wave function, the above form is said to be bilinear. However, we can take the kernel to be wave function dependent, in

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

which case manifestly positive distributions can be obtained. These distributions have been developed in both physics (Lee 1995) and signal processing (Loughlin *et al.* 1993; Choi and Williams 1989; Jeong and Williams 1992; Claasen and Mecklenbrauker 1980; Cohen 1995; Zhao *et al.* 1990).

3.2. *Bilinear (positive) distributions not satisfying the marginals*

Some time after Wigner, Bop (Bopp 1956) and Husimi (Husimi 1940) discussed quantum-like distributions that do not satisfy the marginals, but in some sense reflect quantum features. These distributions are standard in signal analysis (Cohen 1989; Cohen 1995). Instead of discussing the particular cases given by Bop and Husimi, we will formulate the issue in general with

$$P(q,p) = \left| \frac{1}{\sqrt{2\pi}} \int e^{-ipq'} \psi(q')\, w(q'-q) dq' \right|^2 \tag{32}$$

where $w(q)$ is called a window function (Janssen 1984; Janssen 1985; Cohen 1989; Cohen 1995). These distributions do not satisfy the marginals but, depending on the choice of window function, the marginals can be satisfied approximately. They can be obtained from the general form, Equation (31), by taking

$$\Phi(\theta,\tau) = \int w^* \left(u - \frac{1}{2}\tau\right)\, w\left(u + \frac{1}{2}\tau\right)\, e^{-i\theta u} du. \tag{33}$$

3.3. *Positive distributions satisfying the marginals*

We can use the following procedure (Cohen and Zaparovanny 1980) to construct manifestly positive distributions that satisfy the marginals, and that must therefore be non-bilinear:

$$P(q,p) = |\varphi(p)|^2\, |\psi(q)|^2\, \Omega(u(q), v(p)) \tag{34}$$

where for u, v we substitute

$$\begin{aligned} u(q) &= \int_{-\infty}^{q} |\psi(q')|^2\, dq' \\ v(p) &= \int_{-\infty}^{p} |\varphi(p')|^2\, dp'. \end{aligned} \tag{35}$$

The conditions required on Ω to ensure the marginals and positivity are satisfied will be discussed in Section 6.

4. The uncertainly principle, joint distributions and negative values of the Wigner distribution

The standard argument for the non-existence of joint distributions is that the uncertainty principle implies that we cannot have a manifestly positive joint distribution of non-commuting variables. Furthermore, it is often argued that the negative values of the

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

Wigner distribution are a consequence of the uncertainty principle. Whether people have a mathematical or philosophical argument in mind is usually not clear, and typically the arguments are made in a vague manner. The argument is often made by saying that we cannot define a joint distribution that would give a position–momentum value at a point because that would be a *clear* violation of the uncertainty principle. It is not clear what these statements mean, but whatever they mean, they cannot be correct because the uncertainty principle depends *only* on the marginals, so any joint density that satisfies the marginals also satisfies the uncertainty principle. We argue that the uncertainty principle by itself has no bearing on the question of joint distributions, but what it does do is constrain the marginals.

What does the uncertainty principle depend on?

The uncertainty principle *only* depends on the marginals, that is, on $|\psi(q)|^2$ and $|\varphi(p)|^2$, and on the fact that the underlying functions $\psi(q)$ and $\varphi(p)$ are Fourier transform pairs. To see this, recall the definitions of spread in position and momentum,

$$\sigma_q^2 = \int (q - \langle q \rangle)^2 \, |\psi(q)|^2 \, dq \tag{36}$$

$$\sigma_p^2 = \int (p - \langle p \rangle)^2 \, |\varphi(p)|^2 \, dp \,. \tag{37}$$

Taking *only* these definitions and the fact that $\psi(q)$ and $\varphi(p)$ are Fourier transform pairs, the uncertainty principle then follows:

$$\sigma_q \, \sigma_p \geqslant \frac{1}{2}. \tag{38}$$

Now suppose we did have a joint distribution that satisfies the marginals. We could then define the position standard deviation by

$$\sigma_q^2 = \iint (q - \langle q \rangle)^2 \, P(q,p) dq dp \tag{39}$$

$$= \int (q - \langle q \rangle)^2 \, |\psi(q)|^2 \, dq \tag{40}$$

and the momentum standard deviation by

$$\sigma_p^2 = \iint (p - \langle p \rangle)^2 \, P(q,p) dq dp \tag{41}$$

$$= \int (p - \langle p \rangle)^2 \, |\varphi(p)|^2 \, dp \tag{42}$$

where in going from Equation (39) to Equation (40) and from Equation (41) to Equation (42), we have used the fact that the joint distribution satisfies the marginals, Equations (24) and (25). Therefore, *any* joint distribution that satisfies the marginals also satisfies the uncertainty principle.

What does the uncertainty principle constrain?

The uncertainty principle does put constraints on the possible type of joint distributions. Since both marginals cannot be too concentrated, the joint distribution cannot be too

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

concentrated either. However, the fact that we cannot have joint distributions that are too concentrated does not mean that we cannot have joint distributions at all. The fact that the marginals are related imposes further constraints on the possible joint distributions. For example, since the marginals are the absolute square of Fourier transform pairs, and since we know that Fourier transform pairs cannot both be of finite extent, we cannot have a joint distribution that is confined to a finite region in the position–momentum plane.

4.1. *Uncertainty principle and negative areas of the Wigner distribution*

It was realised in the years following the introduction of the Wigner distribution that, generally speaking, the Wigner distribution can go negative, and it is often stated that this is a reflection of the uncertainty principle. The only wave function for which the Wigner distribution is manifestly positive is

$$\psi(q) = (\alpha/\pi)^{1/4} e^{-\alpha q^2/2 + i\beta q^2/2 + ip_0 q} \tag{43}$$

where the Wigner distribution is calculated to be

$$W(q,p) = \frac{1}{\pi} e^{-\alpha q^2 - (p-\beta q - p_0)^2/\alpha}, \tag{44}$$

which is indeed manifestly positive. In fact, this is the only wave function for which the Wigner distribution is positive. This distribution satisfies the uncertainty principle because the Wigner distribution satisfies the marginals for any wave function. Therefore, we see that even for the Wigner distribution, there is one case where the distribution is positive and also satisfies the uncertainty principle. This shows that the issue of whether the distribution goes negative or not is not related to the uncertainty principle, even for the Wigner distribution.

Some classical examples of joint distributions with an uncertainty principle

To illustrate these ideas, we will now construct some classical type distributions that have marginals so that an uncertainty principle exists. Suppose, for example, that x and y have the following densities:

$$P(x) = \frac{1}{\sqrt{2\pi\eta/l}} \exp\left[-\frac{(x-a)^2}{2\eta/l}\right] \tag{45}$$

$$P(y) = \frac{1}{\sqrt{2\pi\eta(l^2+k^2)/l}} \exp\left[-\frac{(y-b)^2}{2\eta(l^2+k^2)/l}\right] \tag{46}$$

where l and k are any two positive numbers that are arbitrary and can be chosen at will. Suppose the 'universal constant' η is fixed and not at our disposal. These two marginals are perfectly good densities, but note that the parameter l appears in both marginals. Calculating the standard deviations, we get

$$\begin{aligned} \sigma_x &= \sqrt{\eta/l} \\ \sigma_y &= \sqrt{\eta(l^2+k^2)/l}. \end{aligned} \tag{47}$$

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

Hence we have an uncertainty principle

$$\sigma_x \sigma_y = \eta \sqrt{1 + l^2/k^2} \geqslant \eta, \tag{48}$$

and no matter how we choose l and k, we will always have

$$\sigma_x \sigma_y \geqslant \eta. \tag{49}$$

Does this mean that we cannot have a joint distribution? Of course not. The reason we have an uncertainty principle for this case is that the marginals are functionally related. In this case, it appears by way of the parameter l, which occurs in both distributions. Because the marginals are functionally related, if we change one marginal, we change the other.

Now consider the joint distribution

$$\begin{aligned} P(x, y) &= \frac{1}{2\pi\eta\sqrt{1 + l^2/k^2}} \exp\left[-\frac{(x-a)^2}{2\eta/l} - \frac{(y-b)^2}{2\eta(l^2+k^2)/l} \right] \\ &= P(x)\, P(y). \end{aligned} \tag{50}$$

This distribution is the independent joint distribution and is totally proper. Hence, even though the marginals are related, the x and y are not correlated. However, another joint density that satisfies the same marginals and has correlations is

$$\begin{aligned} P(x, y) = &\frac{1}{2\pi\eta\sqrt{1 + l^2/k^2}\sqrt{1 - r^2}} \\ &\exp\left[-\frac{1}{2(1-r^2)} \left\{ \frac{(x-a)^2}{\eta/l} - 2\frac{r(x-a)(y-b)}{\eta\sqrt{1 + l^2/k^2}} + \frac{(y-b)^2}{\eta(l^2+k^2)/l} \right\} \right] \end{aligned} \tag{51}$$

where r is any number between -1 and 1. For this case, there is correlation, and the correlation is positive or negative, depending on the sign of r. Hence, as we have emphasised, nothing can be concluded from the marginals about the existence or not of joint distributions or about any correlations between the variables.

5. Positive bilinear distributions not satisfying the marginals

In this section we will discuss windowed wave functions and their associated joint probabilities. Suppose we have a wave function $\psi(q)$, and suppose we want to study the momentum properties in a region centred about the point q. We can achieve this by modifying the wave function in such a way that it leaves the wave function more or less unaltered around the position q, but suppresses it for positions distant from q. This is achieved by multiplying the wave function by a window function $w(q)$ to produce a modified wave function, $\psi_q(q')$ given by

$$\psi_q(q') = \psi(q')\, w(q' - q). \tag{52}$$

We choose $w(q)$ such that

$$\psi_q(q') \sim \begin{cases} \psi(q') \text{ for } q' \text{ near } q \\ 0 \quad\quad \text{for } q' \text{ far from } q. \end{cases} \tag{53}$$

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

The windowed wave function is a function of two variables, the fixed position q that we are momentarily focused on and the variable position q'. Since the windowed wave function emphasises the wave function around the position q, the corresponding momentum wave function will reflect the distribution of momentum around that position:

$$\Phi_q(p) = \frac{1}{\sqrt{2\pi}} \int e^{-ipq'} \psi_q(q') dq' \tag{54}$$

$$= \frac{1}{\sqrt{2\pi}} \int e^{-ipq'} \psi(q')\, w(q'-q) dq'. \tag{55}$$

The momentum probability at position q is thus

$$\begin{aligned} P(q,p) &= |\,\Phi_q(p)\,|^2 \\ &= \left| \frac{1}{\sqrt{2\pi}} \int e^{-ipq'} \psi(q')\, w(q'-q) dq' \right|^2. \end{aligned} \tag{56}$$

Such distributions are standard in time-frequency analysis, and many of their mathematical properties have been developed over the last sixty years. They are called spectrograms.

Conversely, we may wish to study position properties at a particular momentum. We window the momentum wave function $\Phi(p)$ with a momentum window function $W(p)$, and define the short-momentum position wave function by

$$\psi_p(q) = \frac{1}{\sqrt{2\pi}} \int e^{ip'q}\, \Phi(p')\, W(p-p') dp'. \tag{57}$$

If we take $W(p)$ to be the momentum wave function corresponding to the window function $w(q)$ so that

$$W(p) = \frac{1}{\sqrt{2\pi}} \int w(q)\, e^{-ipq} dq, \tag{58}$$

it is then straightforward to show that

$$\Phi_q(p) = e^{-ipq}\, \psi_p(q). \tag{59}$$

Since the distribution is the absolute square, we have

$$P(q,p) = |\,\Phi_q(p)\,|^2 = |\,\psi_p(q)\,|^2. \tag{60}$$

This shows that the windowed density can be used to study the behaviour of position properties at a particular momentum. This is done by choosing a $W(p)$ that is narrow, or, equivalently, by taking a $w(q)$ that is broad.

Often, the results obtained are particularly revealing if expressed in terms of amplitudes and phases of the wave function and window function,

$$\begin{aligned} \psi(q) &= R_\psi(q)\, e^{iS_\psi(q)} \\ w(q) &= R_w(q)\, e^{iS_w(q)}. \end{aligned} \tag{61}$$

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

For the normalisation, we have

$$\begin{aligned} \varrho(q) &= \int P_{WD}(q,p)dp \\ &= \int R_\psi^2(q')\ R_w^2(q'-q)dq'. \end{aligned} \tag{62}$$

Also, direct calculation yields

$$\langle p \rangle_q = \frac{1}{\varrho(q)} \int R_\psi^2(q')\ R_w^2(q'-q)\ [S'_\psi(q') - S'_w(q'-q)]dq' \tag{63}$$

and

$$\begin{aligned} \sigma_{p|q}^2 =& \frac{1}{\varrho(q)} \int \left(\frac{\partial}{\partial q'} R_\psi(q')\ R_w(q'-q) \right)^2 dq' \\ &+ \frac{1}{2\varrho^2(q)} \iint R_\psi^2(q_1') R_\psi^2(q_2') R_w^2(q_1'-q) R_w^2(q_2'-q) \\ &\qquad \left[S'_\psi(q_1') - S'_\psi(q_2') + S'_w(q_1'-q) - S'_w(q_2'-q) \right]^2 dq_1' dq_2'. \end{aligned} \tag{64}$$

Suppose we use a real window by taking $S_w(q) = 0$ and narrow it so that $R_w^2(q)$ approaches a delta function:

$$R_w^2(q) \rightarrow \delta(q). \tag{65}$$

From Equation (63), we have

$$\langle p \rangle_q \rightarrow \frac{\int R_\psi^2(q')\,\delta(q'-q)\ S'(q')dq'}{\int R_\psi^2(q')\,\delta(q'-q)dq'} = \frac{R_\psi^2(q)\ S'(q)}{R_\psi^2(q)} \tag{66}$$

or

$$\langle p \rangle_q \rightarrow S'(q), \tag{67}$$

which is the quantum mechanical current. That is, if the window is narrowed to get increasing position localisation, the limiting value of the estimated current $\langle p \rangle_q$ approaches the quantum mechanical current. However, while this is a desirable result, the estimate becomes harder to ascertain as we narrow the window. This can be seen by direct mathematical substitution of the delta function in Equation (64), which results in

$$\sigma_{p|q} \rightarrow \infty \qquad [\text{for} \quad R_w^2(q) \rightarrow \delta(q)]. \tag{68}$$

The reason for this is that we have taken a wave function and made it narrow; in fact, we have made it infinitely narrow by letting the magnitude square of the window go to a delta function. The standard deviation of the modified wave function must therefore go to infinity. However, an interesting result is obtained in that it can be shown that

$$\sigma_{p|q}^2 \rightarrow \left(\frac{R'(q)}{R(q)} \right)^2 + \text{window dependent terms.} \tag{69}$$

The first term, which depends only on the wave function, is identical to the conditional spread in momentum derived by totally different methods in Section 2.

The spectrogram has the desirable property that it is manifestly positive, but it has limitations. Among them are that it introduces a window into the analysis, and hence

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

the results are window dependent. It is often impossible to disentangle the effects of the window to obtain the inherent information and characteristics of the wave function. It also suffers from a position–momentum resolution trade-off, meaning that if we want fine position resolution, we have to make the window narrow. However, doing this artificially modifies the wave function into a narrow wave function with a momentum distribution that may have little or no relation to the original wave function.

6. Positive distributions and their construction

There are positive distributions that satisfy the marginals and are easily constructed. Wigner's theorem says that no bilinear distribution exists that satisfies the marginals, so we must search for non-bilinear distributions. Positive distributions can be obtained as follows (Cohen and Zaparovanny 1980):

$$P(q,p) = |\varphi(p)|^2 |\psi(q)|^2 \Omega(u(q), v(p)) \tag{70}$$

where Ω is a positive function that characterises the particular distribution. It can be obtained as follows. Choose any positive function $\Omega(u,v)$ of two variables such that it satisfies

$$\begin{aligned} \int_0^1 \Omega(u,v)\, dv &= 1 \\ \int_0^1 \Omega(u,v)\, du &= 1. \end{aligned} \tag{71}$$

We then substitute for u and v the functions of position and momentum $u(q)$ and $v(p)$ given by

$$\begin{aligned} u(q) &= \int_{-\infty}^{q} |\psi(q')|^2 \, dq' \\ v(p) &= \int_{-\infty}^{p} |\varphi(p')|^2 \, dp'. \end{aligned} \tag{72}$$

Note that u and v are the cumulative distributions of the marginals. To show that the marginals are satisfied, we note that

$$dv = |\varphi(p)|^2 \, dp, \tag{73}$$

and integrate with respect to p to give

$$\begin{aligned} \int P(q,p)dp &= |\psi(q)|^2 \int |\varphi(p)|^2 \Omega(u,v)\, dp \\ &= |\psi(q)|^2 \int_0^1 \Omega(u,v)\, dv \\ &= |\psi(q)|^2, \end{aligned} \tag{74}$$

and similarly for the momentum marginals. The distributions given by Equation (70) are positive provided Ω is positive for the range $0 \leqslant u,v \leqslant 1$. Note that Ω may or may not be a functional of the marginals or wave function. Examples of Ωs satisfying the above

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

conditions are easily generated. For example,

$$\Omega(u,v) = 1 - (nu^{n-1} - 1)(mv^{m-1} - 1) \tag{75}$$

where n, m are any integers. A particular choice is

$$\Omega(u,v) = 1, \tag{76}$$

in which case

$$P(q,p) = |\varphi(p)|^2 |\psi(q)|^2. \tag{77}$$

This is the correlationless distribution. We do not argue that it is correct in any sense, but it does show that even for the simplest case the wave function comes in four times rather than bilinearly. We note and emphasise that all these positive distributions satisfy the uncertainty principle because they satisfy the marginals.

There are an infinite number of positive distributions for any wave function, and they are each characterised by a particular choice for the kernel $\Omega(u,v)$. This situation is analogous to the bilinear distributions, which are characterised by the kernel $\Phi(\theta,\tau)$. As with the bilinear distributions, the issue with the positive distributions is how to choose and construct them in some reasonable or systematic way. The first practical procedure for accomplishing this objective was developed by Loughlin *et al.* (Loughlin 1992; Loughlin 1997; Loughlin and Davidson 1998; Loughlin *et al.* 1994) and others based on maximum entropy (Groutage 1997). Subsequently, other methods have been developed for constructing positive distributions.

7. The significance of bilinearity

We now address the issue as to why there seems to be a preference for bilinearity. One argument is that since the marginals are bilinear in the wave function, the joint density should also be bilinear, but this argument is false. Another reason for the emphasis on bilinear distributions is that the bilinear distributions give a simple formulation for the calculation of averages by phase space methods. In particular, if we want to find the average value of the operator $\mathbf{G}(\mathbf{Q},\mathbf{P})$, we calculate

$$\langle \mathbf{G}(\mathbf{Q},\mathbf{P}) \rangle = \int \psi^*(q)\, \mathbf{G}(\mathbf{Q},\mathbf{P})\, \psi(q)\, dq, \tag{78}$$

but we can also calculate it by phase-space integration with respect to a distribution $C(q,p)$:

$$\langle \mathbf{G}(\mathbf{Q},\mathbf{P}) \rangle = \int\int g(q,p)\, C(q,p)\, dq\, dp \tag{79}$$

where

$$C(q,p) = \frac{1}{4\pi^2} \iiint \psi^*\left(u - \frac{1}{2}\tau\right) \psi\left(u + \frac{1}{2}\tau\right) \Phi(\theta,\tau)\, e^{-i\theta q - i\tau p + i\theta u}\, du\, d\tau\, d\theta. \tag{80}$$

In order for the two procedures to give the same answer, the relationship between the function g and the operator $\mathbf{G}$ is given by (Cohen 1966; Moyal 1949)

$$\mathbf{G}(\mathbf{Q},\mathbf{P}) = \int\int \gamma(\theta,\tau)\Phi(\theta,\tau)\, e^{i\theta\mathbf{Q} + i\tau\mathbf{P}}\, d\theta\, d\tau \tag{81}$$

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

where

$$\gamma(\theta,\tau) = \frac{1}{4\pi^2}\int\int g(q,p)\,e^{-i\theta q - i\tau p}\,dq\,dp \tag{82}$$

and $\Phi(\theta,\tau)$ is the kernel of the distribution as discussed in Section 3. Note that Equation (78) is inherently bilinear if we assume that the operator is wave function independent, which is usually the case. Also, in the proof of Equation (79), the distribution $C(q,p)$, as given by Equation (80) with the assumption that the kernel is wave function-independent, is bilinear. Obtaining averages by Equation (78) so that they equal the averages given by Equation (79) has been well-developed for the case where $C(q,p)$ is bilinear, and indeed this is called the phase space formulation of quantum mechanics. This formalism has not been developed for the positive distributions. It is important to emphasise that if the operator is of the form

$$\mathbf{G}(\mathbf{Q},\mathbf{P}) = f(\mathbf{Q}) + h(\mathbf{P}), \tag{83}$$

then *any* distribution $C(q,p)$ that satisfies the marginals will give the correct averages where

$$g(q,p) = f(q) + h(p). \tag{84}$$

That is, if and only if the distribution satisfies the marginals are we assured that for any arbitrary functions f and h do we obtain the correct average from the distribution:

$$\begin{aligned}\langle f(\mathbf{Q}) + h(\mathbf{P})\rangle &= \int \psi^*(q)\,[f(\mathbf{Q}) + h(\mathbf{P})]\psi(q)\,dq \\ &= \iint [f(q) + h(p)]\,C(q,p)\,dq\,dp.\end{aligned} \tag{85}$$

In fact, this equation is the condition that Wigner imposed in his original paper. Now, although the first integral is indeed bilinear, the second integral need not be bilinear. In particular, the equation above is satisfied by the positive distributions because all that is required is that the distribution give the correct marginals. Restriction to bilinear distributions is not necessary. In fact, such a restriction forces the distribution to be negative in order for it to yield the correct marginals.

8. Are there 'wave functions' in standard probability theory?

In most fields one deals with distributions directly, but in some fields, such as electromagnetism and acoustics, we have wave functions and the density is obtained by taking the absolute square of the wave function. Similarly, in quantum mechanics, we have wave functions and obtain probability densities by taking the absolute square of the wave function. What we gain by using wave functions is that a single function carries information about an infinite number of possible physical variables. How is this done? In quantum mechanics we associate a physical variable with an operator $\mathbf{A}$ and a transform $F(a)$ associated with it. Let us call the physical variable a. The transform, $F(a)$ for that variable is related to the wave function by

$$F(a) = \int \psi(q)\,u^*(a,q)\,dq \tag{86}$$

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

where $u(a,q)$ is the transformation matrix. The density of a is

$$\begin{aligned} P(a) &= |F(a)|^2 \\ &= \left| \int \psi(q)\, u^*(a,q)\, dq \right|^2 . \end{aligned} \tag{87}$$

The transformation matrix is obtained by solving the eigenvalue equation for the operator:

$$\mathbf{A}\, u(a,q) \;=\; a\, u(a,q). \tag{88}$$

If the quantum mechanical approach is so powerful, so economical, so simple, why is it that 'wave functions' never appear in standard probability theory? Well, perhaps they do, as we will now discuss. In probability theory we define the characteristic function $M(\theta)$ of a density $P(x)$ by

$$M(\theta) = \int e^{i\theta x} P(x)\; dx \tag{89}$$

$$P(x) = \frac{1}{2\pi} \int M(\theta) e^{-i\theta x} d\theta, \tag{90}$$

and note that the moments of the distribution can be obtained using

$$\langle x^n \rangle = \frac{1}{i^n} \frac{\partial^n M(\theta)}{\partial x^n} |_{\theta=0}\,. \tag{91}$$

However, while the characteristic function is in general a complex function, not every complex function is a characteristic function since it must be the Fourier transform of some *positive* density. Khinchin (Khinchin 1937) proved that $M(\theta)$ is a characteristic function if and only if there exists a function $g(x)$ such that

$$M(\theta) = \int g^*(x) g(x+\theta)\, dx. \tag{92}$$

If there is such a function, it must be normalised to one:

$$\int\; |g(x)|^2\, dx = 1. \tag{93}$$

We have generalised this theorem and have shown that for a Hermitian operator $\mathbf{A}$, we have that $M(\theta)$ is a characteristic function if and only if there exists the representation

$$M(\theta) = \int g^*(x)\, e^{i\theta \mathbf{A}} g(x)\, dx \tag{94}$$

for some function $g(x)$.

We will now show that the Khinchin theorem is a special case. If for the operator we take the momentum operator

$$\mathbf{A} = \mathbf{P} = \frac{1}{i} \frac{d}{dx}, \tag{95}$$

then

$$\begin{aligned} M(\theta) &= \int g^*(x)\, e^{i\theta \frac{1}{i} \frac{d}{dx}} g(x)\, dx \\ &= \int g^*(x) g(x+\theta)\, dx, \end{aligned} \tag{96}$$

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

which is Khinchin's expression.

Now consider the distribution $P(a)$ connected with the characteristic function given by Equation (94). We have

$$\begin{aligned} P(a) &= \frac{1}{2\pi}\int M(\theta)e^{-i\theta\alpha}d\theta \\ &= \left|\int u^*(a,x)g(x)\,dx\right|^2 \\ &= |F(a)|^2, \end{aligned} \tag{97}$$

which is the identical result to that given by quantum mechanics. Furthermore, consider the expectation value

$$\begin{aligned} \langle A\rangle &= \frac{1}{i}\frac{\partial M(\theta)}{\partial x}|_{\theta=0} \\ &= \frac{1}{i}\frac{\partial}{\partial x}\int g^*(x)\,e^{i\theta\mathbf{A}}g(x)\,dx|_{\theta=0} \end{aligned} \tag{98}$$

or

$$\langle A\rangle = \int g^*(x)\,\mathbf{A}g(x)\,dx, \tag{99}$$

which is the quantum mechanical way for getting expectation values.

We speculate, therefore, that the gs of classical probability theory are the wave functions of quantum mechanics. While there is a one-to-one relationship between the probability distribution and the characteristic function, this is not the case with the gs. There are many gs that yield the same density, and it is this non-uniqueness that allows the freedom to carry information about many physical variables with one function.

9. Conclusion

We have discussed the current state for constructing joint distributions of non-commuting observables. We have shown that quantum mechanics does lead to the concept of conditional quantities, which in turn implies the existence of joint distributions. No complete theory currently exists wherein quantum mechanics can be formulated in the phase-space of position and momentum, but there are certainly indications that such a phase-space must exist. We have also indicated that some of the methods that appear in quantum mechanics also appear in classical probability theory.

We have also pointed out that an interesting extension of these ideas is to consider the joint probability for the same variable at different times. Suppose we have a wave function $\psi(q,t)$ at time t. The probability distribution is then given by

$$P(q,t) = |\psi(q,t)|^2\,. \tag{100}$$

Now suppose we evolve the wave function to a time t' through

$$\psi(q,t') = e^{-i\mathbf{H}(t'-t)}\psi(q,t), \tag{101}$$

or, equivalently, we evolve the operator in the Heisenberg representation

$$\mathbf{q}(t') = e^{i\mathbf{H}(t'-t)/\hbar}\mathbf{q}(t)e^{-i\mathbf{H}(t'-t)/\hbar}, \tag{102}$$

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

where **H** is the Hamiltonian. But $\mathbf{q}(t')$ and $\mathbf{q}(t)$ do not commute, and hence, according to standard quantum mechanics, we cannot have a joint distribution that satisfies the marginals. In this case we seek a joint distribution $P(q'q;t',t)$ with the marginal conditions

$$\int P(q',q;t',t)dq = |\psi(q,t')|^2 \tag{103}$$

$$\int P(q',q;t',t)dq' = |\psi(q,t)|^2 . \tag{104}$$

We are then led to ask whether we can develop a theory for $P(q'q;t',t)$? All these issues are interesting and challenging and provide insight into the nature of quantum mechanics.

References

Baraniuk, R. G. and Jones, D. L. (1995) Unitary Equivalence: A New Twist on Signal Processing. *IEEE Transactions on Signal Processing* **43** (10) 2269–2282.

Bopp, F. (1956) La Mechanique Quantique Est-Elle Une Mechanique Statistique Classique Particuliere? *Annales de l'Institut Henri Poincaré* **15** 81–112.

Choi, H. I. and Williams, W. J. (1989) Improved time-frequency representation of multicomponent signals using exponential kernels. *IEEE Transactions on Acoustics, Speech and Signal Processing* **37** 862–871.

Claasen, T. and Mecklenbrauker, W. (1980) The Wigner distribution – A tool for time-frequency analysis, Parts I-III. *Philips Journal of Research* **35** (3-6) 217–250, 276–300, 372–389.

Cohen, L. (1966) Generalized phase–space distribution functions. *Journal of Mathematical Physics* **7** 781–786.

Cohen, L. (1989) Time-Frequency Distributions – A Review. *Proceedings of the IEEE* **77** 941–981.

Cohen, L. (1995) *Time-Frequency Analysis*, Prentice-Hall.

Cohen, L. (1996) Local Values in Quantum Mechanics. *Physics Letters A* **212** 315–319.

Cohen, L. (2000) The Uncertainty principle for Windowed Wave Functions. *Optics Communications* **179** 221–229.

Cohen, L. and Zaparovanny, Y. (1980) Positive Quantum Joint Distributions. *Journal of Mathematical Physics* **21** 794–796.

Davidson, K. and Loughlin, P. (2000) Compensating for window effects in the calculation of spectrographic instantaneous bandwidth. *IEEE Transactions on Biomedical Engineering* **47** (4) 556–558.

Entralgo, E. E., Kuryshkin, V. V. and Zaparovanny, Yu. I. (1988) *Microphysical Reality and Quantum Formalism*, Kluwer Academic Publishers.

Groutage, D. (1997) A fast algorithm for computing minimum cross-entropy positive time-frequency distributions. *IEEE Transactions on Signal Processing* **45** (8) 1954–1970.

Husimi, K. (1940) Some Formal Properties of the Density Matrix. *Proceedings of the Physico-Mathematical Society of Japan* **22** 264–314.

Janssen, A. J. E. M. (1984) Positivity properties of phase-plane distribution functions. *Journal of Mathematical Physics* **25** 2240–2252.

Janssen, A. J. E. M. (1985) Bilinear phase-plane distribution functions and positivity. *Journal of Mathematical Physics* **26** 1986–1994.

Jeong, J. and Williams, W. (1992) Kernel design for reduced interference distributions. *IEEE Transactions on Signal Processing* **40** (2) 402–412.

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press

Khinchin, A. (1937) A new derivation of a formula of P. Levy. *Bulletin of the Moscow State University* **1** 1–5.
Kirkwood, J. G. (1933) Quantum statistics of almost classical ensembles. *Physical Review* **44** 31–37.
Kuryshkin, V. V. (1973) Some problems of quantum mechanics possessing a non-negative phase-space distribution function. *International Journal of Theoretical Physics* **7** 451–466.
Kuryshkin, V. V. and Zaparovanny, Yu. I. (1984) *Comptes Rendus de l'Académie des Sciences* **2** 17.
Kuryshkin, V. V., Lybas, I. A. and Zaparovanny, Yu. (1980) *Annales de la Fondation Louis de Broglie* **5** 105.
Lee, H. W. (1995) Theory and application of the quantum phase-space distribution-functions. *Physics Reports* **259** 147–211.
Loughlin, P. (1992) *Time-Frequency Energy Density Functions: Theory and Synthesis*, Ph.D. dissertation, University of Washington.
Loughlin, P. (1997) Cohen–Posch (positive) time-frequency distributions: development and applications. *Applied Signal Processing* **4** 122–130.
Loughlin, P. and Davidson, K. (1998) Positive local variances of time-frequency distributions and local uncertainty. *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis* 541–544.
Loughlin, P., Pitton, J. and Atlas, L. (1993) Bilinear time-frequency representations: new insights and properties. *IEEE Transactions on Signal Processing* **41** (2) 750–767.
Loughlin, P., Pitton, J. and Atlas, L. (1994) Construction of positive time-frequency distributions. *IEEE Transactions on Signal Processing* **42** (10) 2697–2705.
Mourgues, G., Feix, M. R., Andrieux, J. C. and Bertrand, P. (1985) Not necessary but sufficient conditions for the positivity of generalized Wigner functions. *Journal of Mathematical Physics* **26** 2554–2555.
Moyal, J. E. (1949) Quantum mechanics as a statistical theory. *Proceedings of the Cambridge Philosophical Society* **45** 99–124.
Muga, J. G., Palao, J. P. and Sala, R. (1998) Average local values and local variances in quantum mechanics *Physics Letters A* **238** 90–94.
Muga, J. G. Seidel, D. and Hegerfeldt. G. C. (2005) Quantum kinetic energy densities: an operational approach. *Journal of Chemical Physics* **122** (1) 154106.
Mugur-Schächter, M. (1979) A study of Wigner's theorem on joint probabilities. *Foundations of Physics* **9** 389–404.
Mugur-Schächter, M. (1977) The Quantum Mechanical One-System Formalism, Joint Probabilities and Locality. In: Lopes, J. and Patty, M. (eds.) *Quantum Mechanics, a Half Century Later*, Springer-Verlag 107–146.
O'Connell, R. F. and Wigner, E. P. (1981) Quantum–mechanical distribution functions: Conditions for uniqueness. *Physics Letters* **83A** 145–148.
Sala, R., Palao, J. P. and Muga, J. G. (1997) Phase space formalisms of quantum mechanics with singular kernel. *Physics Letters A* **231** 304–310.
Scully, M. O. and Cohen, L. (1987) Quasi-Probability Distributions for Arbitrary Operators. In: Kim, Y. S. and Zachary, W. W. (eds.) The Physics of Phase Space. *Springer-Verlag Lecture Notes in Physics* **278** 253–260.
Wigner, E. P. (1932) On the quantum correction for thermodynamic equilibrium. *Physical Review*, **40** 749–759.
Wigner, E. P. (1971) Quantum–mechanical distribution functions revisited. In: Yourgrau, W. and van der Merwe, A. (eds.) *Perspectives in Quantum Theory*, MIT Press 25–36.
Zhao, Y., Atlas, L. E. and Marks, R. J. (1990) The use of cone–shaped kernels for generalized time–frequency representations of nonstationary signals. *IEEE Transactions on Acoustics, Speech and Signal Processing* **38** 1084–1091.

https://doi.org/10.1017/S0960129512000709 Published online by Cambridge University Press