Negative Probabilities

Yuri Gurevich
Microsoft Research
Joint work with Andreas Blass of the U of Michigan

Heisenberg's uncertainty principle limits the precision with which position x and momentum p of a particle can be known simultaneously:

$$\sigma_x \sigma_p \geq \hbar/2$$

You may know the probability distributions of x and p individually (in a given state $|\psi\rangle$ of the particle) but the joint probability distribution with these marginal distributions of x and p makes no physical sense.

Does it make mathematical sense?
If yes, can it be used for anything?

In 1932, Eugene Wigner exhibited such a joint distribution. There was, however, a little trouble with it: some of its values were negative. (Here and below, the terms “probability distribution”, “probability function”, and “density function” are used as synonyms.)

“But of course this must not hinder the use of it in calculations as an auxiliary function which obeys many relations we would expect from such a probability,” wrote Wigner.

Such generalized probability distributions became known as *quasi-probability distributions*.

Wigner's discovery gave rise to a phase-space version of quantum mechanics; see

"Quantum Mechanics in Phase Spaces:
An Overview with Selected Papers"
C.K. Zachos et al. (eds.)
World Scientific, 2005.

In the 1980s, Richard Feynman came to negative probabilities from a different angle:

> *"if we combined the principles of quantum mechanics and those of relativity plus certain tacit assumptions, we seemed only able to produce theories . . . which gave infinity for the answer to certain questions. . . . In an attempt to . . . make a theory which would give only finite answers . . . , I looked into the tacit assumptions."*

One tacit assumptions was that probabilities are non-negative, and Feynman came with a natural quasi-probability distribution in a very different setting.

But, today, taking into account the time bound, I will not examine Feynman's approach.

What bothered us was that both Wigner and Feynman apparently pulled their quasi-probability distributions from thin air.

We wanted to understand how can one arrive at their distributions systematically. The question was not addressed in the World Scientific volume mentioned above. Hence our own little investigation.

Wigner writes that his function “was found by L. Szilárd and the present author some years ago for another purpose,” but he gives no reference and no hint about what that other purpose was.

He notes that there are lots of other quasi-probability distributions that would serve as well[1], but none without negative values. The particular Wigner function “was chosen from all possible expressions, because it seems to be the simplest.”

[1] Indeed, just add any distribution with zero x and p marginals. E.g. draw a square around the origin and make a chess-board pattern on the four quadrants, where values $+1$ and -1 alternate.

We found a characterization of Wigner's chosen quasi-probability distribution that might be considered objective:

Theorem

Wigner's distribution 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.

Eventually we discovered that we were scooped:

"A tomographic approach to Wigner's function,"
Jacqueline and Pierre Bertrand,
Foundations of Physics 17:4, 1987.

We can claim only a simpler proof. It is indeed much simpler than that of Bertrands.

The forward FT sends functions $f_1(x), f_2(x, y)$ to

$$\hat{f}_1(\xi) = \frac{1}{\sqrt{2\pi}} \int f_1(x)\, e^{-i\xi x}\, dx,$$
$$\hat{f}_2(\xi, \eta) = \frac{1}{2\pi} \iint f_2(x, y)\, e^{-i(\xi x + \eta y)}\, dx\, dy,$$

and the inverse FT sends $g_1(\xi), g_2(\xi, \eta)$ to

$$\check{g}_1(x) = \frac{1}{\sqrt{2\pi}} \int g_1(\xi) e^{i\xi x}\, dx,$$
$$\check{g}_2(x, y) = \frac{1}{2\pi} \iint g_2(\xi, \eta) e^{i(\xi x + \eta y)}\, d\xi\, d\eta.$$

- All integrals are from $-\infty$ to ∞.

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

Let $f(x,p)$ be an ordinary (nonnegative) probability distribution on $\mathbb{R}^2$, and $g(z)$ the marginal distribution of a linear combination $z = ax + bp$ where a, b are not both zero.

Remarkably, $\hat{g}$ is essentially the restriction of $\hat{f}$.

Lemma (J2M)

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

We prove the lemma in the next two slides. By symmetry, we may assume that $b \neq 0$.

- ax and bp should have the same dimension. For example, if x is a length and p a momentum then a is a momentum and b a length so that ax and bp have units of action (like gram centimeters squared per second). To make z a pure number, divide by $\hbar$.

We have $p = \frac{1}{b}(z - ax)$, $dp = \frac{1}{b}(dz - a\,dx)$ and
$f(x, p)\,dx\,dp = f\left(x, \frac{1}{b}(z - ax)\right)\frac{1}{b}\,dx\,dz$.

- Here and below we use differential 2-forms for area elements, so that $dx\,dz$ really means $dx \wedge dz$ here. One gets the same results using appropriate Jacobians.

To compute g, we integrate $f(x, p)$ along the lines on which z is constant:

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

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

By definition, δ-function is not a function but a linear functional (or "distribution"): for any nice "test function" f,

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

So

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

Usual functions also can be seen as distributions. As distribution,

$$\delta(x) = \frac{1}{2\pi}\int e^{itx}dt.$$

Indeed,

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

$e^{aD}f(x) = f(x+a)$

Here $D = \frac{d}{dx}$ and

$$e^{aD} = \sum_{k=0}^{\infty} \frac{(aD)^k}{k!} = I + aD + \frac{1}{2}a^2D^2 + \frac{1}{6}a^3D^3 + \dots$$

We have

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

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

Q: Taylor series expansions are for real-analytic functions, whereas ψ is merely in L^2.

A: On analytic functions, e^{aD} is simply a shift $f(x) \mapsto f(x+a)$. Gaussian densities $a \exp\left(-\frac{(x-b)^2}{2c^2}\right)$ are analytic and span a dense subspace of $L^2(\mathbb{R})$. There is a unique continuous extension of e^{aD} to L^2, namely the shift $f(x) \mapsto f(x+a)$.

Q: How do you know that Gaussian densities span a dense subspace?

A: Every smooth function with a compact support is approximated by a sum of Gaussian densities, and such functions are known to be dense on L^2.

Consider a particle (with no electric charge or spin or any other quantum number) moving in (for the simplicity of exposition) one dimension. Its state $|\psi\rangle$ is given by a normalized (to norm 1) wave function $\psi(x)$ in $L^2(\mathbb{R})$. Compute the probability densities $g(z)$ of linear combinations $z = ax + bp$ of the position and momentum of the particle.

Because of the uncertainty principle, we have no guarantee that these distributions are the marginals of any probability distribution $f(x, p)$.

Nevertheless, let's use the functions $g(z)$, for all choices of a and b, to reconstruct what f would have to be if it existed. The simple connection between $\hat{f}$ and any $\hat{g}$, suggest reconstructing $\hat{f}$ and then applying the inverse FT.

The Hermitian operators corresponding to position and momentum are X and P, given by

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

As before, we fix a and b, not both zero, and consider the linear combination $z = ax + bp$ and its associated Hermitian operator $Z = aX + bP$. We compute the forward Fourier transform $\hat{g}$ of the probability distribution $g(z)$ of $z = ax + bp$ in the state $|\psi\rangle$.

An aside. Eigenfunctions of X are δ-functions. Eigenfunctions of P have the form $e^{\frac{1}{\hbar}ix}$. Neither belong to $L^2(\mathbb{R})$. Enriching $L^2(\mathbb{R})$ with these functions leads to an example of the so-called rigged Hilbert spaces, a.k.a. Gelfand triples.

We have

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

By the J2M lemma, if f existed its Fourier transform would satisfy

$$\sqrt{2\pi}\,\hat{f}(a\zeta, b\zeta) = \hat{g}(\zeta) = \frac{1}{\sqrt{2\pi}}\langle\psi|e^{-i\zeta(aX+bP)}|\psi\rangle.$$

Setting $\alpha = a\zeta$ (of dimension reciprocal length) and $\beta = b\zeta$ (of dimension reciprocal momentum), we would have

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

It would be helpful to split the exponential $e^{-i\alpha X - i\beta P}$ into a product: a factor with X times a factor with P. But X and P don't commute. We have, however, two pieces of good luck. First there is Zassenhaus's formula

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

where the dots refer to factors involving higher commutators. Second $[X, P]$ is equal to $i\hbar I$ and thus commutes with everything, so we can omit the dots from the formula. We have

$$\begin{aligned}\hat{f}(\alpha, \beta) &= \frac{1}{2\pi} \langle \psi | e^{-i\alpha X - i\beta P} | \psi \rangle \\ &= \frac{1}{2\pi} \langle \psi | e^{-i\alpha X} e^{-i\beta P} e^{\alpha\beta [X,P]/2} | \psi \rangle .\end{aligned}$$

The exponential $\frac{1}{2}\alpha\beta[X, P] = i\alpha\beta\hbar/2$ can be pulled out of the bra-ket:

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

Here we used $(P\psi)(y) = -i\hbar\frac{d\psi}{dy}(y)$. Since
$e^{-\beta\hbar\frac{d}{dy}}\psi(y) = \psi(y - \beta\hbar)$,

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

Now apply the 2D inverse FT.

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

Since

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

$$\int \psi^*(y)\psi(y-\beta\hbar)\delta(y-x-\frac{\beta\hbar}{2})dy = \psi^*(x+\frac{\beta\hbar}{2})\psi(x-\frac{\beta\hbar}{2}),$$

we have

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

which is Wigner's quasi-distribution.

There is another, much earlier, approach to characterizing the Wigner quasi-probability distribution, using 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.

There is a trade-off. The class of functions is wider but the feature to match is narrower.

The earlier approach builds on a 1927 proposal by Hermann Weyl.

Eine physikalische Größe ist durch ihren Funktionsausdruck f(p,q) in den kanonischen Variablen p, q mathematisch definiert. Es blieb ein Problem, wie ein derartiger Ausdruck auf die Matrizen zu übertragen war. Ohne weiteres klar war das nur für die Potenzen p^k, q^l und damit für Polynome. Freilich trat schon hier die Schwierigkeit auf, daß man nicht wußte, ob man einen Term wie $p^2 q$ als $P^2 Q$ oder QP^2 oder PQP usw. zu interpretieren hatte. Der Ansatz ist offenbar viel zu formal. Unsere gruppentheoretische Auffassung zeigt sogleich den rechten Weg: die Hermitesche Form

$$F = \iint \exp(P\sigma + Q\tau)\, \xi(\sigma, \tau)\, d\sigma d\tau$$

repräsentiert die Größe

$$f(p, q) = \iint \exp(p\sigma + q\tau)\, \xi(\sigma, \tau)\, d\sigma d\tau.$$

Note that every (well-behaved) function $f(p, q)$ of position and momentum can be written in the form

$$f(p, q) = \iint \exp(p\sigma + q\tau)\,\xi(\sigma, \tau)\,d\sigma d\tau.$$

The desired $\xi(\sigma, \tau) = 2\pi\,\hat{f}(p, q)$.

Returning to our notations, let us restated the Weyl correspondence as follows. For any (well-behaved) function of $g(x, p)$ of position and momentum, the associated linear operator is

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

If you buy 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 $f(x,p)$.

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

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

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 that the Fourier transform is a unitary operator and therefore preserves the inner product structure of L^2.

In order that these two expectations agree, for all (well-behaved) g, we need that

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

This is the same equation that we derived above from the requirement that the marginal distributions of linear functions $ax + bp$ be correct. And we saw there that this equation leads, via Zassenhaus's Lemma and the inverse Fourier transform, to the formula for Wigner's quasi-distribution.