Fisher information, disorder, and the equilibrium distributions of physics

B. Roy Frieden
Optical Sciences Center, University of Arizona, Tucson, Arizona 85721
(Received 26 June 1989; revised manuscript received 27 September 1989)

Consider an isolated statistical system specified by a coordinate x and its probability density $p(x)$. A functional of $p(x)$ called "Fisher information" can be used to measure the degree of disorder of the system due to the spread in $p(x)$. Fisher information may be minimized, subject to a physical constraint, to attain a temporal equilibrium solution $p(x)$. When the constraint is linear in the mean kinetic energy of the system, the equilibrium solution $p(x)$ often obeys the correct differential equation for the system. In this way, the Schrödinger (energy) wave equation, Klein-Gordon equation, Helmholtz wave equation, diffusion equation, Boltzmann law, and Maxwell-Boltzmann law may be derived from one *classical* principle of disorder. The convergence rate for Fisher information is about that for alternative use of maximum entropy (in problems where both have the same equilibrium solution). This suggests that Fisher information defines an arrow of time. The arrow points in the direction of decreasing accuracy for the determination of the mean, or ideal, value of a parameter.

INTRODUCTION

In a previous paper,[1] we showed that the Schrödinger wave equation can be derived from an uncertainty principle that is suggested by classical probability estimation theory. This is that the error in an optimal determination of mean particle position should be a maximum. No physical basis was given for this principle, except that it seemed reasonable on the basis of the perversity of nature, and because it led to the correct answer, the Schrödinger equation.

In this paper, we show that this uncertainty principle is actually the statement of a new principle of statistical disorder, that temporally $\delta I \leq 0$, where I is the Fisher information. We also formulate a consistent approach to deriving the major equilibrium distributions of physics, through the use of the Fisher disorder measure. This is that the Fisher I should be a minimum, subject to a linear constraint on mean kinetic energy.

GEDANKEN EXPERIMENT

Consider an isolated system, consisting either of many particles or the stochastic realizations of one particle. The system is specified by a physical parameter x. The particles may be material particles, or photons, etc., and x can specify position coordinate, velocity, or any other coordinate of interest. Let $p(x)$ describe the probability density for x. Assume that $p(x)$ is unknown, and is to be determined.

Imagine the following gedanken experiment to be performed. The observer makes one coordinate measurement y, where

$$y=\theta+x \ . \tag{1}$$

Quantity θ is the ensemble mean position (or velocity, etc.). See Fig. 1, which shows a two-dimensional version, for purposes of visualization. From this *single observation,* the observer tries to best infer θ. Call the resulting estimate $\hat{\theta}\equiv\hat{\theta}(y)$. How well can θ be determined, at any one time, and how does the expected error in $\hat{\theta}$ change with time?

The statistical increase in disorder that accompanies an increase in thermodynamic entropy suggests that the expected error should increase in time. Intuitively, the situation can be pictured as follows. Suppose that the particles are well localized initially, say, because they are in a small container. Then, a measurement of position y must be close to *mean* position θ. Hence an estimate $\hat{\theta}$ based on datum y must incur small error, on average. Next, the observer opens the walls of the container and repeats the experiment at successive times. As time progresses, disorder increases, and the particles become ever more randomly spread out, so that successive outputs y of the experiment tend to depart ever more widely from θ. Hence the expected error in $\hat{\theta}$ increases. Using a mean-square error measure e^2,

$$\delta(e^2)\geq 0 \ . \tag{2}$$

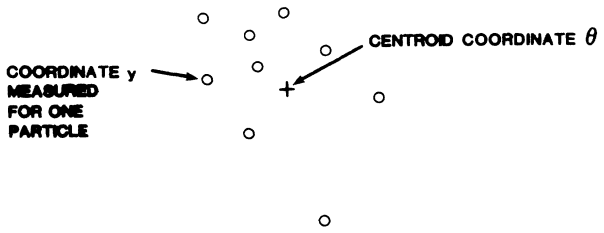

FIG. 1. The gedanken experiment. By measuring the position (or velocity, etc.) y of one particle, best infer the mean position θ (or velocity, etc.) of the ensemble. In optical and quantum-mechanical cases, y and θ are positions. In the Boltzmann ideal gas problem, y and θ are velocities. In the latter case, θ is the mean drift velocity, which could be due to the container's motion relative to the observer.

 ©1990 The American Physical Society

This is, of course, an average effect. Randomly, some new particle configurations might temporarily improve the estimate, but these should be relatively infrequent events.

FISHER INFORMATION AND NEGENTROPY

The situation can be quantified by a powerful result from estimation theory.[1] The best possible estimator $\hat{\theta}(y)$ suffers a mean-square error e^2 from θ that obeys

$$e^2 = 1/I \ , \tag{3a}$$

$$I = \int dx (dp/dx)^2/p(x) \ . \tag{3b}$$

(In this and all subsequent cases, integration limits are from $-\infty$ to $+\infty$ unless otherwise specified.) The particular $\hat{\theta}(y)$ that achieves error (3) is called the "efficient" estimator. Any other estimator must have a larger mean-square error. The only proviso to this result is that all estimators under consideration be unbiased, i.e., obey

$$\langle \hat{\theta}(y) \rangle = \theta \ .$$

A condition of unbiasedness is usually considered desirable for an estimator, just as it is for any physical apparatus whose aim would be to measure θ. The quantity I in Eq. (3b) is called the "Fisher information," after its discoverer.[2,3]

The effects (2) and (3a) together imply that the change in I should be negative,

$$\delta I \leq 0 \ . \tag{4}$$

That this represents a change toward increased disorder, can be seen as follows. As $p(x)$ becomes broader and smoother, its gradients decrease, so that by Eq. (3b) I decreases [as in (4)]. For example, if $p(x)$ is Gaussian with standard deviation σ, then Eq. (3b) gives $I = 1/\sigma^2$. A broad, smooth probability law $p(x)$ of course represents increased randomness in x. Hence condition (4) represents increased randomness in x.

Physically, according to the second law of thermodynamics,

$$\delta H \leq 0 \ , \tag{5}$$

where H is the "negentropy" [This is the negative of entropy. It is convenient to use negentropy, rather than entropy, in this paper, in order to facilitate comparisons between changes in I and H, which both tend to go in the same direction if H is the *negative* of entropy. See Eqs. (4) and (5).] This is the functional

$$H = \int dx \, p(x) \ln p(x) \ . \tag{6}$$

H is another measure of smoothness in $p(x)$. To continue the example above, if $p(x)$ is Gaussian then $H = c + \ln(1/\sigma)$, c an inconsequential number. Hence, once again as $p(x)$ becomes smoother the functional (now H) decreases, although logarithmically (versus directly for I).

EQUILIBRIUM STATES

With no physical constraints on x present, I will decrease, according to effect (4), until it reaches absolute zero [when $p(x) =$ constant[4]]. However, with a constraint on parameter x, such as a fixed $\langle x^2 \rangle$, I will decrease until it reaches a *finite* minimum,

$$I = \int dx (dp/dx)^2/p(x) = \min \ . \tag{7}$$

This defines a stationary (or equilibrium) state of $p(x)$, and is called the minimum-Fisher-information (MFI) principle. As we shall see, according to its constraints, various differential equations of physics are the solutions.

The time required to reach the equilibrium state $p(x)$ may be finite, as in thermodynamics and diffusion, or zero, as in the cases of quantum mechanics and diffraction optics considered. In the latter cases, the equality sign holds in disorder principle (4). The nature of the relaxation time will be seen to have a strong bearing upon which disorder principle, (4) or (5), will apply.

ENTROPY VERSUS FISHER INFORMATION

The second law of thermodynamics states that the equilibrium probability law $p(x)$ to a given statistical scenario must obey

$$H = \int dx \, p(x) \ln p(x) = \min \ . \tag{8}$$

This is compared with the MFI principle (7). The two principles give, in general, different equilibrium answers $p(x)$. When is each applicable?

In thermodynamical situations, where classical particles form the statistics, of course entropy principle (8) gives the correct answer $p(x)$. Note that in problems of this type, there is a finite amount of time taken for the equilibrium state $p(x)$ to be attained. This is commonly called the "relaxation time." En route, at each time interval δt, the second law inequality (5) is obeyed by each physically acceptable perturbation to $p(x)$. This is the statement of Boltzmann's H theorem.

By contrast, consider quantum mechanics, and in particular where (as assumed in this paper) the potential energy function $V(x)$ is real and nontemporal. Here, the particles are decidedly nonclassical. Also, the equilibrium law $p(x)$ is the same as the initial law $p(x)$ at $t = 0$. Hence, in effect, the equilibrium distribution is attained instantaneously. In this scenario, there is no time interval available for physical enactment of the Boltzmann H theorem. Hence there is no reason to expect principle (8) of minimum negentropy to be valid. Indeed, there is no published claim to the effect that it is valid in quantum mechanics.[5] [That it gives wrong answers is shown at Eqs. (51)–(54) below.] The same argument holds for scalar diffraction theory, where the equilibrium intensity law $p(x)$ at any time is the law at $t = 0$ as well.

If, in these cases, entropy principle (8) is not the appropriate measure of disorder, is there an alternative? As we shall see, the MFI principle (7) is the appropriate measure of disorder for these cases, in the sense that it gives the correct equilibrium laws $p(x)$.

Moreover, even in those limited cases where entropy H

is appropriate to use, the alternative use of Fisher principle (7) will be seen to give *the same answers,* making it equally valid there [See Eqs. (26)–(36) and (D4).]

TEMPORAL EVOLUTION OF $p(x)$

Consider a thermodynamic problem, i.e., one where the equilibrium law $p(x)$ takes a finite amount of time to be converged upon. At what rate is it converged upon if Fisher principle (4) is used? Also, what is the rate if instead entropy principle (5) is used?

First, suppose that Fisher principle (4) guides the random evolution of a law $p(x)$: If a law $p(x)$ is perturbed to a new law, the new law is acceptable only if it causes a decrease in I. Does this define a convergent process, and if so, how fast does it converge?

To simulate the process, we wrote a computer program that generates a random initial function $p_0(x)$ by Fourier series, and then perturbs $p_0(x)$ by adding random changes to the series coefficients. The resulting law $p(x)$ is accepted if its I value is less than that of the initial law. If I instead increases, a new set of perturbations is generated, etc., until a new law *is* accepted. This law is now perturbed again, etc. In this way, the law $p(x)$ evolves in accordance with increasing disorder as measured by the Fisher I.

We found that regardless of starting function $p_0(x)$, the law evolves toward the correct stationary state[4] for this problem $p(x)=1.0$ for $0\le x\le 1$. For example, the evolved $p(x)$ law after 200 trial perturbations is shown in Fig. 2 (bold curve). For comparison, we also show the evolved law obtained similarly by use of the second law, principle (5) (thin curve). Both are close to the stationary answer.

By testing with many different initial laws $p_0(x)$ in this way, we found that principle (4) causes convergence toward the stationary solution at about the same rate as does the second law, principle (5).

This has an interesting ramification. Since the two rates are about equal, and since in thermodynamics the Fisher principle converges to *the same* (Boltzmann) law as does the entropy principle (see below), the Fisher principle (4) might be used to define an arrow of time. If so, the arrow points in the direction of decreasing ability to estimate mean particle position. Position determination must, in an overall sense, be getting worse with time. See further discussion of this point below.

The two curves in Fig. 2 have a curious difference in the character of their oscillations. The entropy solution (thin curve) has all frequencies present about equally, while the Fisher solution (bold) appears to lack high frequencies. This can be accounted for as follows. Consider a random perturbation $\epsilon f(x)$, where ϵ is small and $f(x)$ arbitrary, from the stationary solution $p(x)=1$. The changes δH and δI from the stationary values for H and I can be easily shown to obey

$$\delta H=\tfrac{1}{2}\epsilon^2\int_0^1 dx\, f^2(x),\quad \delta I=\epsilon^2\int_0^1 dx\, f'^2(x) \tag{9a}$$

to second order. In terms of the spectrum $F(\omega)$ of $f(x)$, these become

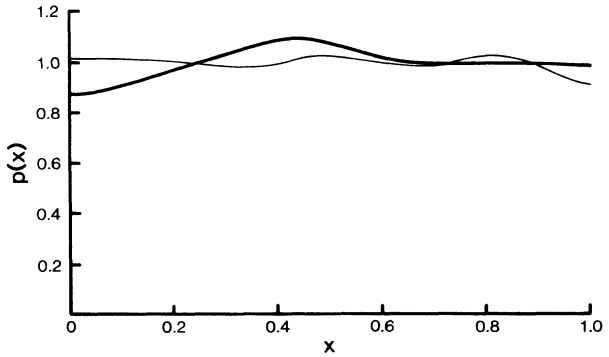

FIG. 2. Temporal evolution study. After 200 trial perturbations, an initial law $p_0(x)$ evolves, via Fisher principle (4), to the bold curve; or to the thin curve, via the entropy principle (5).

$$\delta H=\tfrac{1}{2}\epsilon^2\int d\omega|F(\omega)|^2\ ,$$
$$\delta I=\epsilon^2\int d\omega\,\omega^2|F(\omega)|^2\ . \tag{9b}$$

Since both these quantities are being minimized [by selection procedures (4) and (5)], because of the extra factor ω^2 in the last integral, the Fisher solution will tend to selectively lose more high-frequency content, as time progresses, than will the entropy solution. Hence, although the two solutions converge toward equilibrium at about the same rate, their oscillations about equilibrium differ structurally.

MULTIPARAMETER SYSTEMS—ADDITIVITY PROPERTY

It is well known[6] that the entropy of a system is an extensive, or additive, measure of disorder. The entropy from independent system parameters adds. Does Fisher information have this property as well? First we have to generalize Fisher information to the case of two or more parameters.

Suppose, then, that the unknown parameter mean θ has two components θ_1 and θ_2. The single datum y is replaced by a two-component vector $\mathbf{y}$ obeying

$$y_i=\theta_i+x_i,\quad i=1,2\ .$$

Let e_1^2 be the mean-square error in estimating θ_1, and e_2^2 be that in θ_2, by use of input $\mathbf{y}$. Use the precision[7]

$$1/e_1^2+1/e_2^2\equiv I(X_1,X_2) \tag{10}$$

as a measure of the quality of the estimates. The best possible estimates have a precision[8]

$$I(X_1,X_2)=\int\int dx_1dx_2\frac{(\partial p/\partial x_1)^2+(\partial p/\partial x_2)^2}{p}\ . \tag{11}$$

Equation (11) shows that $I(X_1,X_2)$ is a two-dimensional generalization of $I\equiv I(X_1)$ in Eq. (3b). Hence we adopt $I(X_1,X_2)$ as the *two*-dimensional Fisher information measure. The N-dimensional measure is analogous.

Next, assume that the system parameters x_1 and x_2 are independent, so that

$$p(x_1,x_2)=p_1(x_1)p_2(x_2) .$$

Then

$$\frac{\partial p}{\partial x_1}=p_1'p_2, \quad \frac{\partial p}{\partial x_2}=p_1p_2' ,$$

and substitution into Eq. (11) gives

$$I(x_1,x_2)=\int\int dx_1dx_2\frac{p_1'^2p_2^2+p_1^2p_2'^2}{p_1p_2} .$$

After integrating out x_2 in the first integral and x_1 in the second, we get

$$I(X_1,X_2)=\int dx_1p_1'^2/p_1+\int dx_2p_2'^2/p_2$$

or

$$I(X_1,X_2)=I(X_1)+I(X_2) \tag{12}$$

by definitions (3b) and (11).

Hence, with statistical independence, the net information is the sum of the component information values. Fisher information obeys an additivity property.

NET FISHER PRINCIPLE

Principle (7) is unconstrained, as it stands, by any prior knowledge describing the physical situation. The standard way to inject such information into a Lagrangian problem is to add in linear constraint terms, and we do this. It turns out that the key such constraint term is the expression for mean kinetic energy of the system. The expression for mean kinetic energy E_{kin} allows the user to inject the physics of the situation into MFI. The overall principle becomes, for one-dimensional problems,

$$\int dx\, p'^2(x)/p(x)+\lambda\int dx\, E_{\text{kin}}(x)p(x)=\min ,$$
$$p'(x)\equiv dp(x)/dx . \tag{13}$$

The second integral is the mean kinetic energy, $\langle E_{\text{kin}}\rangle$. Function $E_{\text{kin}}(x)$ is the kinetic energy expressed as a function of the coordinate x in use. As the simplest example, if x is the x-coordinate velocity of a molecule in an ideal gas, then $E_{\text{kin}}(x)$ is $\frac{1}{2}mx^2$.

It is found that the MFI principle (13) derives many of the basic equilibrium distributions of physics, when the following recipe is followed. When $\langle E_{\text{kin}}\rangle$ is known, then parameter λ is solved for, in the usual Lagrange equality-constraint way. Also, then the normalization constraint on $p(x)$ is explicitly added in as another equality-constraint term. If, however, $\langle E_{\text{kin}}\rangle$ is not known, then λ becomes instead a negative *weighting* factor that must be properly fixed by the user. The former case is valid in thermodynamics (see below), the latter in quantum mechanics and diffraction optics, as discussed in the next section.

In the special case where coordinate x is itself the kinetic energy, i.e., where the probability law on kinetic energy is sought, principle (13) has to be modified. See Appendix D.

The two-dimensional generalization of principle (13), for a law $p(x,y)$, is given in Appendix B. It is shown there that the solution $p(x,y)$ separates when the E_{kin} constraints are marginal in x and y. Hence, when the constraints are marginal, MFI predicts that the variables x and y are independent. This has important application to thermodynamics and diffusion, as seen below.

APPLICATION TO QUANTUM MECHANICS: THE SCHRÖDINGER WAVE EQUATION

Consider the basic scenario of a particle of mass m moving in a nontemporal, real potential field $V(x)$. The system state parameter x is here a position coordinate.

In the gedanken experiment of Fig. 1, now, one particle is present. One stochastic realization of position y is measured, in an attempt to measure the particle's classical (mean) position θ.

Calling the total energy $W(x)$,

$$E_{\text{kin}}(x)=W(x)-V(x) .$$

The mean E_{kin} is then

$$\langle E_{\text{kin}}\rangle=\int dx\, E_{\text{kin}}(x)p(x)$$
$$=\int dx[W(x)-V(x)]p(x) .$$

This simplifies to

$$\langle E_{\text{kin}}\rangle=\int dx[W-V(x)]p(x) , \tag{14}$$

where W is the mean total energy

$$W=\int dx\, W(x)p(x) , \tag{15}$$

and we used the normalization property of $p(x)$

$$\int dx\, p(x)=1 .$$

By Eqs. (13) and (14), the net Fisher principle is now

$$\int dx p'^2(x)/p(x)+\lambda\int dx[W-V(x)]p(x)=\min . \tag{16}$$

The solution to this variational problem is the ordinary Euler-Lagrange equation

$$\frac{d}{dx}\left[\frac{\partial L}{\partial p'}\right]=\frac{\partial L}{\partial p} , \tag{17}$$

with

$$L=p'^2/p+\lambda[W-V(x)]p .$$

This simplifies to[1]

$$q''(x)+\lambda q(x)[W-V(x)]=0, \quad p(x)=q(x)^2 . \tag{18}$$

Since $p(x)$ is *a priori* unknown, $\langle E_{\text{kin}}\rangle$ cannot be known from Eq. (14). By our recipe [see below Eq. (13)], since $\langle E_{\text{kin}}\rangle$ is unknown, parameter λ becomes a weight that must be fixed by the user. With the choice

$$\lambda=-2m/\hbar^2 , \tag{19}$$

stationary solution (18) is the time-independent Schrödinger wave equation. Function $q(x)$ then has the familiar role of a probability "amplitude," although here

it is purely real. In cases $V(x)$ where a complex probability amplitude is required, use of MFI and a supplemental binary variable leads to the complex Schrödinger equation. See Appendix A.

It is satisfying that the stationary MFI solution to this problem is the stationary Schrödinger equation. Of further interest is that the eigensolutions $q_n(x)$ to Eq. (18) define subsidiary minima in I over the space of $q(x)$ functions, and that the lowest eigenvalue solution $q_0(x)$ or $q_1(x)$ (as the case may be) defines the *absolute minimum* in I. Hence, of all possible eigensolutions to Eq. (18), the lowest eigenvalue solution defines the state of maximum disorder for the system. Mathematically, this follows from the fact that the lowest eigenvalue solution is also generally the smoothest.

Returning to our gedanken experiment, this also means that the average or (here) classical position θ of the particle would be most difficult (or inaccurate) to estimate, from one observation of position y, if the system were in the lowest eigenstate. The lowest eigenstate of a system is usually preferred because it is a lowest energy state. We see now that it is also preferred because it is a *state of maximum disorder.*

KLEIN-GORDON EQUATION

The MFI principle (13) also gives rise to this relativistic equation, in the particular case of zero fields. See Appendix C.

APPLICATION TO DIFFRACTION OPTICS: THE HELMHOLTZ WAVE EQUATION

Consider the situation of light entering a medium of known refractive index profile $n(x)$. The system state parameter x is again a position coordinate. We now use the semiclassical treatment of radiation to find the kinetic energy of a typical photon within the medium.

By the semiclassical approximation, a photon is treated as a "particle" in a field $V(x)=0$, where

$$E_{\text{kin}}(x)=P^2(x)/2m \ , \tag{20}$$

where $P(x)$ is the particle's momentum at a general position x, and m is its mass. Use DeBroglie's hypothesis, wavelength

$$\lambda_0=h/P \ . \tag{21}$$

Also, of course

$$\lambda_0\nu=v(x)=c/n(x) \ , \tag{22}$$

where $\lambda_0=\lambda_0(x)$ is the light wavelength at a general position x within the medium, ν is the frequency, $v(x)$ is the velocity within the medium, and c is the speed of light *in vacuo.* Combining Eqs. (20) through (22), we get

$$E_{\text{kin}}(x)=(h^2/2m)\nu^2n^2(x)/c^2 \ . \tag{23}$$

The net Fisher principle (13) is now

$$\int dx\, p'^2(x)/p(x) + \lambda(h^2/2m)(\nu^2/c^2)\int dx\, n^2(x)p(x)=\min \ .$$

Since the photon is being treated as a semiclassical particle, we make the same choice for λ, Eq. (19), as in the quantum-mechanical case. The result is

$$\int dx\, p'^2(x)/p(x)-(\omega/c)^2\int dx\, n^2(x)p(x)=\min \ , \tag{24}$$

$$\omega\equiv 2\pi\nu \ .$$

Using the Euler-Lagrange equation (17), the solution[8] to problem (24) is a differential equation

$$q''(x)+(\omega/c)^2n^2(x)q(x)=0, \quad p(x)=q(x)^2 \ . \tag{25}$$

This is the Helmholtz wave equation, governing scalar diffraction. Probability $p(x)$ is, here, the local intensity.

An interesting aspect of solution (24) is its lack of dependence upon a phase function $\phi(x)$ supplemental to the amplitude $q(x)$. Evidently, diffraction intensity effects can be completely described without the need for the concept of phase. Phase is a useful, but artificial, concept. This is discussed further in Ref. 8.

APPLICATION TO THERMODYNAMICS: THE MAXWELL-BOLTZMANN LAW

Consider an ideal, monatomic gas not subjected to a field of potential. The container holding the gas may be in motion, with unknown velocity θ. See Fig. 1. Let x,y, and z be the components of *velocity* for a randomly selected atom. The gedanken experiment is now the measurement of one atom's velocity.

Note that a value θ represents *the special case* $n=1$ of the root-mean-square velocity among atoms taken n at a time. We will use this fact later.

Let the system be subjected to a constant temperature T. The principle of equipartition of energy states that when an equilibrium state $p(x)$ is attained the average kinetic energy associated with each degree of freedom (x,y, or z) has the same value,

$$\langle E_{\text{kin}}\rangle \equiv \int dx\,(\tfrac{1}{2}mx^2)p(x)=kT/2 \ , \tag{26}$$

where k is Boltzmann's constant. This is an *equality constraint,* so that the E_{kin} term in principle (13) is now a Lagrange equality-constraint term. Hence parameter λ will have to be solved for, not assigned as in previous applications.

This is a case of three *marginal* constraints (26). Hence, by Appendix C, $p(x,y,z)$ separates and it is sufficient to solve a one-dimensional MFI problem (C7). We arbitrarily choose the problem in x. Inserting, as well, the equality constraint of normalization [see discussion beneath Eq. (13)] yields an MFI principle

$$\int dx\, p'^2(x)/p(x)+\lambda\left[\int dx\tfrac{1}{2}mx^2p(x)-kT/2\right] +\mu\left[\int dx\, p(x)-1\right]=\min \ . \tag{27}$$

Using the Euler-Lagrange equation (17) for Lagrangian

$$L=p'^2/p+\tfrac{1}{2}\lambda mx^2p+\mu p \ ,$$

we have a solution

$$2\frac{d}{dx}\left[\frac{p'}{p}\right]+\left[\frac{p'}{p}\right]^2-\tfrac{1}{2}m\lambda x^2-\mu=0\ . \tag{28}$$

It is convenient to now define an auxiliary function

$$h(x)=p'(x)/p(x)\ . \tag{29}$$

Equation (28) becomes

$$2h'(x)+h^2(x)-\tfrac{1}{2}m\lambda x^2-\mu=0\ , \tag{30}$$

a Riccati equation. The plan is to solve this equation for $h(x)$, and then substitute the solution into Eq. (29) to yield $p(x)$, according to which

$$p(x)=A\ \exp\left[\int dx\ h(x)\right]\ . \tag{31}$$

The integral is here an indefinite one, and A is an arbitrary constant.

Solution 1 (lowest-order)

Riccati differential equations are commonly solved by trial power series, setting the coefficients of successive powers of x equal to zero. In this way, it is found that the trial solution

$$h(x)=a+bx\ , \tag{32}$$

a,b constant, works in Eq. (30). In order to satisfy the two constraints in Eq. (27), it turns out that $a=0$. Then by Eq. (31) the solution is of the form

$$p(x)=A\ \exp(Bx^2)\ , \tag{33}$$

where A and B are constants. Back substituting solution (33) into the two constraint equations allows A and B to be solved for. The result is

$$p(x)=\left[\frac{m}{2\pi kT}\right]^{1/2}\exp(-mx^2/2kT)\ , \tag{34}$$

the one-dimensional Boltzmann distribution law.

As we noted, the marginal laws $p(y)$ and $p(z)$ for the y and z components of velocity may be found in the same way, and give the same form of solution (34). We also found that x, y, and z are independent. Hence

$$p(x,y,z)=p(x)p(y)p(z)\ ,$$

so that

$$p(x,y,z)=\left[\frac{m}{2\pi kT}\right]^{3/2}\exp\left[-\frac{m}{2kT}(x^2+y^2+z^2)\right]\ . \tag{35}$$

Finally, we seek $p(v)$, where

$$v=(x^2+y^2+z^2)^{1/2}\ .$$

This is easily accomplished by Jacobian transforming the law (35) in terms of (x,y,z) to a law $p(v,\theta,\phi)$ where θ and ϕ are the usual polar and azimuthal angles in spherical polar coordinates. Then $p(v,\theta,\phi)$ is integrated out over θ and ϕ to yield

$$p(v)=\sqrt{2/\pi}\left[\frac{m}{kT}\right]^{3/2}v^2\exp(-mv^2/2kT)\ , \tag{36}$$

the Maxwell-Boltzmann law. This is the correct result.

Solution 2 (higher-order)

Because the MFI principle (13) has generally a differential equation as its solution, multiple solutions to any one problem may be expected. This is usually beneficial. For example, any principle that is to imply the Schrödinger equation must be capable of multiple solutions. In fact, the MFI solution to the present problem leads to a second solution, found next.

Instead of the trial solution (32), try[9]

$$h(x)=bx+c/x\ , \tag{37}$$

where b and c are constant, in Eq. (30). This is found to work, providing $c=2$. Hence, by Eq. (31), the solution $p(x)$ is of the form

$$p(x)=A\ \exp(bx^2/2+2\ln x)\ , \tag{38}$$

or more simply

$$p(x)=Ax^2\exp(Bx^2)\ . \tag{39}$$

Comparing this with the prior solution (33), we see that an extra factor x^2 now multiplies the Boltzmann solution.

Back-substituting solution (39) into the two constraint equations leads to solutions for A and B. The result is

$$p(x)=\frac{1}{\sqrt{2\pi}}\left[\frac{3m}{kT}\right]^{3/2}x^2\exp(-3mx^2/2kT)\ . \tag{40}$$

Again forming $p(x,y,z)$ as the product of three laws of form (40), transforming by Jacobian to spherical polar coordinates (v,θ,ϕ), and integrating out θ and ϕ, we find a law

$$p(v)=(\tfrac{54}{105})(27/2\pi)^{1/2}(m/kT)^{9/2}v^8\exp(-3mv^2/2kT)\ . \tag{41}$$

The two laws (36) and (41) are plotted in Fig. 3. Quantity v is in units of $(kT/m)^{1/2}$, and $p(v)$ has arbitrary units. It is also interesting to compare the v values at which each law has a maximum. Differentiating d/dv the law (36) and equating this to zero gives

$$v_1=(2kT/m)^{1/2}\cong 1.414(kT/m)^{1/2}\ . \tag{42a}$$

Doing the same operations on law (41) yields

$$v_2=(8kT/3m)^{1/2}\cong 1.633(kT/m)^{1/2}\ . \tag{42b}$$

The two values are surprisingly close. But, what can the distribution (41) physically represent?

The probability law for the root-mean-square (rms) velocity v over n atoms is

$$p(v)=\frac{2}{\Gamma(3n/2)}\frac{n^{3n/2}}{(\sqrt{2}\sigma)^{3n}}v^{3n-1}e^{-nv^2/2\sigma^2}\ ,$$

$$\sigma^2=kT/m\ . \tag{43}$$

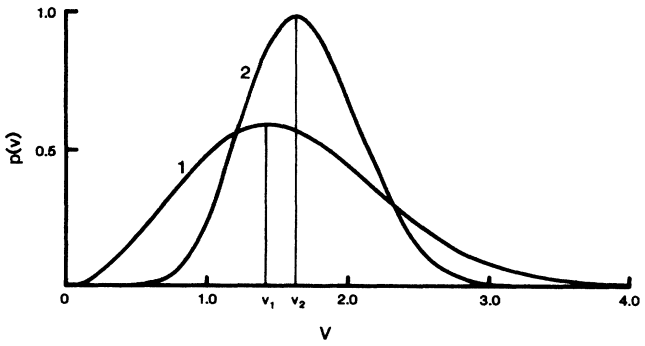

FIG. 3. Curve 1 is the Boltzmann solution (36) for the probability $p(v)$ of particle velocity v in a monatomic gas. Curve 2 is alternative solution (41), describing local rms velocity (and pressure) fluctuations due to three atoms. Velocity v is in units of $(kT/m)^{1/2}$; $p(v)$ is in arbitrary units.

[This may readily be derived by noting that v^2 is the sum over $3n$ squared, Cartesian component velocities, each of which is an independent, Boltzmann degree of freedom. Then v^2 is a χ^2-random variable.[10] Jacobian transforming from the χ^2 law for v^2 to that for v gives the result (43).] With $n=1$, law (43) becomes the MFI solution (36), while with $n=3$, law (43) becomes the higher-order solution (41).

This essentially solves the problem of interpretation. Solution (36) represents the scenario where unknown parameter θ is the rms velocity over particles taken one at a time; while solution (41) corresponds to the problem θ for particles taken three at a time. In the latter case, data value y in the gedanken experiment of Fig. 1 is now the empirical rms mean velocity over the *three* observed velocities. Hence the solutions (36) and (41) correspond to different problems. Denote these as problems $\theta(1)$ and $\theta(3)$, respectively.

It is interesting to find, by Eqs. (3b) and (12), that the information I for problem $\theta(1)$ is $3m/kT$ while that for $\theta(3)$ is $27m/kT$, considerably larger. In the next section, we discuss this effect.

The average pressure on the walls containing the given gas, and the temperature T, are proportional to the first moment of the law (43), with n large.[11] Therefore law (41) has the added significance of defining local pressure fluctuations due to three atoms.

General solution

Solutions to differential equation (27) are generally the squares of parabolic cylinder functions,[12] of which the two lowest order are (34) and (40). These solutions have the general form of a Gaussian function times the square of a polynomial of finite order. The order of the polynomial defines the order of the solution.

As before, each term of the polynomial gives rise to a factor v^m multiplying a Boltzmann exponential. Hence, by Eq. (43), it represents the MFI solution to the problem of estimating the rms velocity θ among $n=(m+1)/3$ particles.

Hence each MFI solution represents the solution to a different problem $\theta=\theta(n)$. It is important to note that each is consistent with the Maxwell-Boltzmann law, since (a) Eq. (43) derived from the law, and (b) because the solution (36) for $\theta(1)$ *is* the Maxwell-Boltzmann law.

We noted, for the case $\theta(3)$, an I value that much exceeded that for $\theta(1)$. In general, I should increase with n. First of all, by direct use of Eq. (43) in definition (3b), I increases with n. Functionally, this is because $p(v)$ becomes more sharply concentrated about its mean as n increases (see the curves in Fig. 3). Hence the gradients increase and (3b) gives a larger output.

But perhaps the more interesting reason for the effect is from the standpoint of estimation. Parameter θ is here the *theoretical* rms mean velocity over n particles, while data value y is the *empirical* rms mean over n randomly selected particles. As n increases, the two means must approach one another, since (as we noted) $p(v)$ becomes more sharply concentrated about its mean θ. Hence an estimate of θ based on data y must suffer less error e^2, so that by Eqs. (3a) and (3b) I increases.

It is interesting to compare the MFI and minimum negentropy approaches to this problem. MFI gives a number of solutions, representing various estimation scenarios $\theta(n)$. Minimum negentropy, by comparison, gives just one solution, the most probable one, for the problem $\theta(1)$. The two approaches agree in the principal solution $n=1$, but approach the problem from different vantage points and hence make different *kinds* of predictions about the phenomenon. MFI outputs relate to the ability to estimate, while entropy outputs relate to the probability of the distribution.

DIFFUSION EQUATION

Let a typical particle in the ideal gas previously discussed be described by position (now) x at time t. We seek the joint law $p(x,t)$ defining maximum disorder for the gas, in the Fisher sense. For simplicity, assume that the particles move in a frictionless, nonviscous medium.

Denote a particle's position at time $t=0$ as x_0. Then after time t, its position x is

$$x=x_0+\Delta x \ , \tag{44}$$

$$\Delta x=\int_0^t dt'\Delta v(t') \ , \tag{45}$$

where Δv is the particle's random velocity fluctuation at time t'. We show next that, by the MFI principle, $\Delta v(t')$ does not correlate with $\Delta v(t'')$, $t'\neq t''$, so that by (45) Δx obeys the central limit theorem and must be Gaussian.

First estimate a law $p(x,y)$, by the MFI principle, where x is velocity component Δv_x at time t' and y is velocity component Δv_x at a different time t''. Assume, as usual, that the velocity distribution law in Δv_x has reached its equilibrium state at both times t' and t''. Then, the *spatial average* of $\frac{1}{2}m\Delta v_x^2$ at t' must equal that at t''. Assuming also that ergodicity holds, these averages $\langle \frac{1}{2}m\Delta v_x^2 \rangle$ must equal $kT/2$ as well. In summary, we seek the law $p(x,y)$, $x=\Delta v_x(t')$, $y=\Delta v_x(t'')$ obeying the *marginal* E_{kin} constraints

$$\int\int dx\,dy\tfrac{1}{2}mx^2p(x,y)=kT/2\ ,$$
$$\int\int dx\,dy\tfrac{1}{2}my^2p(x,y)=kT/2\ . \tag{46}$$

Using these as constraints in the two-dimensional MFI principle (B1), by result (B5) the law $p(x,y)$ separates. Hence $\Delta v_x(t')$ is independent of $\Delta v_x(t'')$, as was to be proven.

Equivalently,

$$\langle \Delta v(t)\Delta v(t')\rangle=\frac{kT}{m}\delta(t-t'),\quad \langle \Delta v(t)\rangle=0\ . \tag{47a}$$

Then by Eqs. (45) and (47a)

$$\langle \Delta x\rangle=0,\quad \langle \Delta x^2\rangle=at,\quad a=kT/m\ . \tag{47b}$$

Also by Eqs. (45) and (47a) Δx is effectively the sum of many independent random variables $\Delta v(t')$, so that by the central limit theorem[10] Δx is Gaussian in its statistics.[13] This holds true regardless of the probability law obeyed by Δv, so that it is true for any of the MFI solutions to (27). Combining this with result (47b), the probability density $p_{\Delta x}(x)$ for Δx obeys

$$p_{\Delta x}(x)=\frac{1}{\sqrt{2\pi at}}e^{-x^2/2at}\ . \tag{48}$$

Position value x_0, in Eq. (44), is a random number that depends on Δv values at times prior to interval $(0,t)$. Therefore, by independence effect (47a), x_0 must be independent of Δx. Then the probability law for their sum x obeys a convolution[10]

$$p(x)\equiv p(x,t)=\int dy\,p_0(y)\frac{1}{(2\pi at)^{1/2}}e^{-(x-y)^2/2at}\ . \tag{49}$$

Direct partial differentiation of this equation shows that[14]

$$\frac{\partial p(x,t)}{\partial t}=\frac{a}{2}\frac{\partial^2 p(x,t)}{\partial x^2},\quad a=kT/m\ . \tag{50}$$

The position of the particle obeys the diffusion equation.

We observe that the key to this derivation was establishing the independence effect (47a). This followed because of the propensity of MFI to deduce that variables are independent when their only constraints are marginal (nonjoint). Another application of this useful effect follows.

JOINT PROBABILITY LAW FOR POSITION AND VELOCITY IN IDEAL GAS

Let x denote the x-component position of a particle, and y denote its x-component velocity. We are interested in establishing $p(x,y)$ for particles in an ideal gas which, additionally, is subjected to central potential energy function $V(x)$.

Obviously the mean kinetic energy obeys

$$\langle \tfrac{1}{2}my^2\rangle=W-\langle V(x)\rangle\ ,$$

where $W=kT/2$ is the mean total energy (15). Hence the MFI principle now has a constraint term

$$\lambda\int\int dx\,dy[W-V(x)-\tfrac{1}{2}my^2]p(x,y)$$
$$=-\lambda\int dx\,V(x)p_1(x)-\lambda\int dy\tfrac{1}{2}my^2p_2(y)\ ,$$

where $p_1(x)$ and $p_2(y)$ are the marginal probabilities. A term in W is dropped because it is independent of p and hence does not affect the minimization of I.

Hence the one constraint becomes a sum of marginal constraints. Therefore, by Appendix C, the solution separates,

$$p(x,y)=p_1(x)p_2(y)\ .$$

This is the correct result.[15]

DISCUSSION

Minimum negentropy is often used nowadays as a method for estimating an unknown probability law $p(x)$ in the presence of insufficient information. Specifically, the principle is[16]

$$\int dx\,p(x)\ln[p(x)/h(x)]$$
$$+\sum_{n=1}^{N}\lambda_n\left[\int dx\,p(x)k_n(x)-K_n\right]=\min\ , \tag{51}$$

where data $\{K_n\}$ and constraint kernels $\{k_n(x)\}$ are known. Also, the "prior probability law" $h(x)$ must be known. With these provisos, the solution $p(x)$ to principle (51) is actually maximum probable.[16] Let us use this principle *under the same constraint information* as for the MFI principle, i.e., when $E_{\text{kin}}(x)$ is a known constraint kernel $k(x)$. Accordingly, in analogy with the MFI principle (13), form a principle

$$\int dx\,p(x)\ln[p(x)/h(x)]+\lambda\int dx\,E_{\text{kin}}(x)p(x)=\min\ . \tag{52}$$

The Lagrangian is here

$$L=p\ln(p/h)+\lambda E_{\text{kin}}(x)p\ .$$

The Euler-Lagrange equation (17) here simplifies to $\partial L/\partial p=0$, or

$$1+\ln p-\ln h+\lambda E_{\text{kin}}=0\ ,$$

with solution

$$\hat{p}(x)=h(x)e^{-1-\lambda E_{\text{kin}}(x)}\ . \tag{53}$$

How often is this the correct *physical* solution?

In the E_{kin} cases (14) and (23) of quantum mechanics and diffraction, respectively, the solutions (53) are directly

$$\hat{p}(x)=h(x)e^{-1-\lambda W+\lambda V(x)} \tag{54a}$$

and

$$\hat{p}(x)=h(x)e^{-1-Kn^2(x)}\ , \tag{54b}$$

K a constant. Surprisingly, the estimated solutions do not have to obey differential equations. This is clearly incorrect for general $V(x)$ and $n(x)$ functions. Nature

does not give simple, exponential answers to the *general* problem in quantum mechanics or diffraction.

In thermodynamic applications, it was found that MFI solutions are Boltzmann solutions to a class of problems denoted as $\theta(1),\ldots,\theta(n)$. We showed this for velocity distributions, but similar results hold for energy distributions as well. For example, when mean kinetic energy E is constrained by (26) once again, the MFI solution for $p(E)$ is the *ordinary exponential law* $\langle E\rangle^{-1}\exp(-E/\langle E\rangle)$. The latter follows because the unconstrained MFI solution on the interval $0\leq x\leq\infty$ is an exponential law with a free mean. See Appendix D. The mean is then fixed by the $\langle E\rangle$ constraint.

SUMMARY

Fisher information, an outgrowth of classical estimation theory, is a physically meaningful measure of disorder for many physical problems. This is in the sense that its use implies the correct equilibrium distributions for such problems.

The equilibrium solution $p(x)$ is defined as that which minimizes the Fisher information subject to a constraint on mean kinetic energy. In this way, the Schrödinger wave equation,[17] Klein-Gordon equation, Helmholtz wave equation, and diffusion equation may be derived. In the case of thermodynamic equilibrium under constant temperature, the equilibrium Fisher I solution for a monatomic gas is either the usual Boltzmann law, or alternative laws consistent with the Boltzmann law (the squares of parabolic cylinder functions). The alternative laws are the laws of mean-square velocity fluctuation due to finite numbers of atoms.

Attempts at estimating distributions $p(x)$ using minimum negentropy in place of minimum I, with the same prior knowledge of $\langle E_{\rm kin}\rangle$, do not lead to correct laws $p(x)$, except in thermodynamic cases. Hence the Fisher measure of disorder applies to a much broader range of phenomena than does the entropy measure.

As we noted, minimum negentropy applies as a principle of disorder only in cases where classical particles are present, and where equilibrium is attained after a finite amount of time. By contrast, in cases of real, nontemporal potential $V(x)$ in quantum mechanics, and in diffraction theory, the solution after finite t is the same as for $t=0$. Hence there is no time span available for relaxation phenomena such as the Boltzmann H theorem and minimum negentropy. They do not apply, and this is why they give incorrect results $p(x)$ in these cases.

In thermodynamic cases, where minimum negentropy does apply, as we saw the use of the Fisher I gives the correct results anyhow. This is in the form of the Boltzmann energy distribution and the Maxwell-Boltzmann law. In addition, higher-order Fisher outputs $p(x)$ describe the correct laws for rms velocity, or pressure, fluctuations due to finite numbers of gas particles. The Fisher approach has the added benefit of giving the answers $p(x)$ to many problems simultaneously. The problems are ranked according to level of disorder.

The sense in which the Fisher equilibrium solution to (13) is attained should be addressed. In thermodynamic problems, the equilibrium solution has the usual sense—as time progresses the empirical $p(x)$ law over the particles evolves in shape, *through many intermediary shapes* $p(x)$, toward an equilibrium shape. By contrast, in quantum mechanics (for real, nontemporal potentials) and diffraction theory, we saw that the solution at any time is the same as that for $t=0$. Here there is no temporal evolution of $p(x)$ through intermediary shapes. In what sense, then, is the solution $p(x)$ to (13) an equilibrium solution? In this case, perhaps the word "equilibrium" should be replaced by the more accurate term "stationary," since the solution to (13) is always a stationary one, by definition. Only virtual (infinitesimal) changes in $p(x)$ are considered, during an infinitesimal time interval dt from time $t=0$. In this sense (13) resembles d'Alembert's variational principle of virtual work.

An important distinction between the two measures of disorder is their physical manifestations. Whereas H connects the concept of disorder with that of heat, I connects disorder with the ability to estimate, i.e., to know in a quantitative sense.

It was found that both the Fisher I and negentropy cause convergence to thermodynamic solutions at about the same rate. This suggests that δI might provide a useful alternative measure of the direction (arrow) of time. If so, the inescapable conclusion seems to be that, due to the inexorable increase in disorder, *the ability to estimate must be decreasing with time.*

As an example, consider an apparatus for measuring the speed of light c. Suppose that the apparatus-estimation rule combination is unbiased (as we required). Then, as time progresses, the accuracy in estimating the mean (true) value c must be decreasing. Basically, because of the second law of thermodynamics, the component parts of this apparatus wear out and contribute to ever-more faulty estimates.

This is a universal effect which can be reversed locally, e.g., if the apparatus is physically improved by replacing worn out parts. However, as with entropy, any local gain in the ability to estimate must be offset somewhere else by an even greater loss: the radiant heat caused by making the replacements will perturb other apparatuses, and hence decrease their accuracies.

ACKNOWLEDGMENTS

This research was supported, in part, by a grant from the Office of Naval Research. It is a pleasure to acknowledge important and helpful conversations with B. H. Soffer of Hughes Research Laboratories and D. Donoho of the University of California at Berkeley. Finally, the papers of E. T. Jaynes have strongly influenced this author to regard the concept of disorder as fundamental to understanding the laws of physics.

APPENDIX A: COMPLEX AMPLITUDE CASES

In order to allow complex solutions to the quantum MFI Eq. (18), the system must require now two parameters for its description. One is x, as before. The other is an internal state variable i, which can only take on two values $i=1,2$. Let

$$p(x,i)\equiv p_i(x), \quad i=1,2 \tag{A1}$$

denote the joint probability of x and i. Since states 1 and 2 are disjoint, the overall $p(x)$ obeys simply[10]

$$p(x)=p_1(x)+p_2(x) . \tag{A2}$$

The total Fisher information over both states obeys

$$I=\sum_{i=1}^{2}\int dx\, p_i'^2(x)/p_i(x) \tag{A3}$$

$$\equiv I_1+I_2 . \tag{A4}$$

We can now combine results. Using, again, a principle of minimum Fisher information constrained by $\langle E_{\rm kin}\rangle$, by Eqs. (A2) and (A3) we get

$$\int dx\, p_1'^2(x)/p_1(x)+\int dx\, p_2'^2(x)/p_2(x) + \int dx[p_1(x)+p_2(x)][W-V(x)]=\min . \tag{A5}$$

Now, $p_1(x)$ and $p_2(x)$ are two analytically distinct functions, so that they separately minimize (A5). Hence use the two Euler-Lagrange equations

$$\frac{d}{dx}\left[\frac{\partial L}{\partial p_i'}\right]=\frac{\partial L}{\partial p_i}, \quad i=1,2 \tag{A6}$$

with L the Lagrangian for Eq. (A5). These give solutions (18),

$$q_i''(x)+\lambda q_i(x)[W-V(x)]=0 , \tag{A7}$$

$$p_i(x)=q_i^2(x), \quad i=1,2 . \tag{A8}$$

Since the two Schrödinger equations (A7) are linear in their $q(x)$ functions, they may be added (after multiplying the $i=2$ equation by $\sqrt{-1}$) to give one Schrödinger equation in a combined quantity

$$\psi(x)=q_1(x)+jq_2(x), \quad j=\sqrt{-1} . \tag{A9}$$

This is the usual complex Schrödinger equation. Finally, by Eqs. (A2) and (A8)

$$p(x)=q_1^2(x)+q_2^2(x) , \tag{A10}$$

and by (A9) this results in

$$p(x)=|\psi(x)|^2 , \tag{A11}$$

the standard result for the complex case.

In summary, in order to force complexity into the MFI quantum solution, a supplementary binary variable $i=1,2$ must be assumed present. State i is a correlant with the ordinary system state parameter x. It directly gives rise to the real and imaginary contributors to $p(x)$, via Eq. (A10), as might have been expected.

The addendum of binary variable i to the *general* MFI principle (13) will result, in the same way that Eqs. (A7) and (A8) were derived, in a problem

$$q_i''(x)+\lambda q_i(x)E_{\rm kin}(x)=0 , \tag{A12}$$

$$p_i(x)=q_i^2(x), \quad i=1,2 \tag{A13}$$

$$p(x)=q_1^2(x)+q_2^2(x) . \tag{A14}$$

As before, this is the prescription for a generally complex solution. Hence, for general problems, MFI supplemented by binary complexity allows for complex solutions.

APPENDIX B: MARGINAL CONSTRAINTS

Consider the case of two system parameters x and y, and given marginal constraints $\langle E_{\rm kin1}(x)\rangle$ and $\langle E_{\rm kin2}(y)\rangle$. By Eq. (11) the MFI principle then becomes

$$\int\int dx\,dy\frac{p_x^2+p_y^2}{p}+\lambda_1\int\int dx\,dy\,E_{\rm kin1}(x)p(x,y) +\lambda_2\int\int dx\,dy\,E_{\rm kin2}(y)p(x,y)=\min , \tag{B1}$$

$$p_x=\frac{\partial p}{\partial x}, \quad p_y=\frac{\partial p}{\partial y} .$$

Under these constraint conditions, principle (51) of minimum negentropy is well known to give a separable solution $p(x,y)=p_1(x)p_2(y)$. We show next that MFI likewise separates under these conditions.

The Euler-Lagrange equation for this two-parameter problem is[18]

$$\frac{\partial}{\partial x}\left[\frac{\partial L}{\partial p_x}\right]+\frac{\partial}{\partial y}\left[\frac{\partial L}{\partial p_y}\right]=\frac{\partial L}{\partial p} . \tag{B2}$$

Using

$$L=(p_x^2+p_y^2)/p+\lambda_1 E_{\rm kin1}(x)p+\lambda_2 E_{\rm kin2}(y)p , \tag{B3}$$

from Eq. (B1), gives the differential equation

$$2p_{xx}/p+2p_{yy}/p-p_x^2/p^2-p_y^2/p^2 -\lambda_1 E_{\rm kin1}(x)-\lambda_2 E_{\rm kin2}(y)=0 . \tag{B4}$$

By substitution, the trial solution

$$p(x,y)=p_1(x)p_2(y) \tag{B5}$$

gives two distinct equations in x and y,

$$2p_i''/p_i-p_i'^2-\lambda_i E_{{\rm kin}i}=0, \quad i=1,2 . \tag{B6}$$

Hence the solution separates. When the constraints are marginal, MFI predicts that the variables are independent. Finally, note that the solution (B6) is equivalent to two distinct *one-dimensional* MFI problems

$$\int dx_i p_i'^2/p_i+\lambda_i\int dx_i E_{{\rm kin}i}(x_i)p_i(x_i)=\min, \quad i=1,2 . \tag{B7}$$

APPENDIX C: THE KLEIN-GORDON EQUATION

For simplicity, consider a field-free, one-dimensional case. Then the Klein-Gordon (KG) equation of relativistic quantum mechanics is[19]

$$\left[-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}+\frac{d^2}{dx^2}\right]\psi-\frac{m^2c^2}{\hbar^2}\psi=0 . \tag{C1}$$

Quantity c is the speed of light, t is the time, x is position, m is the particle rest mass, and $\hbar$ is Planck's constant$/2\pi$. In this field-free case, only positive energy eigenvalues E,

$$E \geq 0 , \tag{C2}$$

need be considered, as contrasted with the possibility for negative E when fields are present.[20] We use this fact below.

Seeking a stationary solution, as usual, let[21]

$$\psi(x,t)=u(x)e^{-jEt/\hbar} . \tag{C3}$$

Substitution into Eq. (C1) gives

$$u''+\left[\frac{E^2-m^2c^4}{c^2\hbar^2}\right]u=0 \tag{C4}$$

as the relativistic stationary solution.[22]

We next derive the stationary KG solution (C4) from the Fisher standpoint. Start with the general MFI principle (13). As in the Schrödinger wave equation derivation, let

$$p=q^2 . \tag{C5}$$

We thereby get a principle

$$\int dx\, q'^2(x)+\lambda\int dx\, E_{\rm kin}(x)q^2(x)=\min . \tag{C6}$$

Since there are no fields, we anticipate that $E_{\rm kin}(x)=E_{\rm kin}$, a constant. Using the Euler-Lagrange equation

$$\frac{d}{dx}\left[\frac{\partial L}{\partial q'}\right]=\frac{\partial L}{\partial q} , \tag{C7}$$

with Lagrangian

$$L=q'^2+\lambda E_{\rm kin}q^2(x) , \tag{C8}$$

the solution is

$$q''-\lambda E_{\rm kin}q=0 . \tag{C9}$$

Compare this with the stationary KG Eq. (C4). They will be identical for the particular choice of λ

$$\lambda=-\frac{(E^2-m^2c^4)}{E_{\rm kin}c^2\hbar^2} . \tag{C10}$$

This choice of λ does not outwardly resemble the nonrelativistic choice Eq. (19). However, we shall see next that it does, in fact, closely resemble (19).

Use

$$E_{\rm kin}=E-mc^2 \tag{C11}$$

and

$$E^2=c^2P^2+m^2c^4 \tag{C12}$$

in Eq. (C10), where P is the particle momentum. Equation (C10) becomes

$$\lambda=-\frac{P^2/\hbar^2}{(c^2P^2+m^2c^4)^{1/2}-mc^2} . \tag{C13}$$

Now eliminate P through

$$P=\gamma mv , \tag{C14}$$

where γ is the Lorentz contraction factor and v is the particle speed relative to a reference frame. This gives

$$\frac{1}{\lambda}=-\hbar^2\frac{(c^2\gamma^2m^2v^2+m^2c^4)^{1/2}-mc^2}{\gamma^2m^2v^2} . \tag{C15}$$

After cancellation of a factor m, and use of

$$\gamma=(1-v^2/c^2)^{-1/2} , \tag{C16}$$

we get simply

$$\lambda=-\frac{m(\gamma+1)}{\hbar^2} . \tag{C17}$$

A positive square root was taken in (C15), since $E_{\rm kin}$ must be positive by Eq. (C2).

Compare the relativistic choice (C17) of λ with the nonrelativistic choice Eq. (19). We see that factor 2 in Eq. (19) has been generalized to factor $(\gamma+1)$ in Eq. (C17). As a check, by Eq. (C16), $(\gamma+1)\rightarrow 2$ as $v\rightarrow 0$, so that the two results are consistent.

In conclusion, the MFI solution (C9) with the parameter choice (C17) is the stationary KG free-space equation.

Note that we have restricted attention to the free-space case. This is for reasons of consistency with MFI. In free-space cases, $|\psi|^2$ retains the interpretation of a probability law, since it is then positive definite.[20] However, in cases where finite electromagnetic (or other) fields exist, $|\psi|^2$ in general goes negative[20] and hence cannot be interpreted as a probability law. On the other hand, the MFI principle (13) is restricted in scope to the derivation of probability laws. Therefore it would not make sense to attempt to derive by MFI the KG equation when finite fields can exist.

It is serendipitous that, in the field-free case, the *Dirac* relativistic equation is equivalent to the KG equation.[23] Hence, in this sense, the Dirac equation has also been derived by MFI principle (13).

In the usual derivation of the KG equation, the eigenvalues of a *squared* energy operator E^2 are sought.[24] This is usually regarded as unsatisfactory, on the grounds that linear operators are preferred. It is interesting that the MFI route to KG used, in fact, a linear expression in energy, Eq. (C6).

APPENDIX D: UNCONSTRAINED MFI SOLUTION OVER $(0,\infty)$

Consider the problem

$$\int_0^\infty dx\, p'^2(x)/p(x)+\lambda\left[\int_0^\infty dx\, p(x)-1\right]=\min . \tag{D1}$$

By the Euler-Lagrange equation (17), the solution obeys

$$q''(x)=\lambda q(x), \quad p(x)=q(x)^2 . \tag{D2}$$

Of course, the general solution is

$$q(x)=Ae^{\sqrt{\lambda}x}+Be^{-\sqrt{\lambda}x} . \tag{D3}$$

All parameters A, B, and $\sqrt{\lambda}$ must be real so that $p(x)\equiv q^2(x)$ remains real. In order for $p(x)$ to obey normalization over the infinite interval, necessarily $A=0$. Then enforcing normalization gives the unique solution

$$p(x)=\langle x\rangle^{-1}e^{-x/\langle x\rangle}\ . \tag{D4}$$

When $x=E_{\text{kin}}$, the kinetic energy, the solution is the ordinary Boltzmann law.

Notice that the variational principle (D1) is not of the general form (13). It lacks a constraint term on $\langle E_{\text{kin}}\rangle$. This is because the constraint term is now inconsistent with the Fisher (first) term in (13). In the gedanken experiment of Fig. 1, the mean θ is to be *unknown,* and this gives rise to the first term in principle (13). But here θ is $\langle E_{\text{kin}}\rangle$, which is *known,* by equipartition law (26). It would be inconsistent, then, to include both terms in (13). Hence the constraint term is dropped, giving principle (D1).

[1]B. R. Frieden, Am. J. Phys. **57**, 1004 (1989).

[2]R. A. Fisher, *Statistical Methods and Scientific Inference,* 2nd ed. (Oliver and Boyd, London, 1959).

[3]H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part I* (Wiley, New York, 1968), p. 80.

[4]B. R. Frieden, J. Mod. Opt. **35**, 1297 (1988).

[5]Note that the discrete probability law P_m describing the *empirical occurrence rate* of m samples from the underlying law $p(x)$ does relate, albeit approximately, to a discrete version of entropy principle (8). See R. Kikuchi and B. H. Soffer, J. Opt. Soc. Am. **67**, 1656 (1977). However, our aim here is to estimate $p(x)$, not P_m.

[6]I. Prigogine, *Thermodynamics of Irreversible Processes* (Thomas, Springfield, IL, 1955), p. 16.

[7]G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York, 1968), p. 633.

[8]B. R. Frieden, Opt. Lett. **14**, 199 (1989).

[9]Michael H. Rose (private communication).

[10]B. R. Frieden, *Probability, Statistical Optics and Data Testing* (Springer-Verlag, New York, 1983).

[11]G. Joos, *Theoretical Physics* (Stechert, New York, 1934), p. 532.

[12]*Handbook of Mathematical Functions.* Natl. Bur. Stand. Appl. Math. Ser. No. 55, edited by M. Abramowitz and I. Stegun (U.S. GPO, Washington, D.C., 1964), p. 686.

[13]A. Papoulis, *Probability, Random Variables and Stochastic Processes* (McGraw-Hill, New York, 1965), pp. 292, 293, 436, and 437.

[14]N. M. Blachman, IEEE Trans. Inf. Theory **IT-11**, 267 (1965).

[15]L. Landau and E. Lifshitz, *Statistical Physics* (Pergamon, New York, 1958), pp. 112–113.

[16]B. R. Frieden, J. Opt. Soc. Am. **73**, 927 (1983); also J. E. Shore and R. W. Johnson, IEEE Trans. Inf. Theory **IT-27**, 472 (1981).

[17]Past attempts at deriving the SWE from stochastic theory are summarized in M. Baublitz, Prog. Theor. Phys. **80**, 232 (1988). Of further interest is Schrödinger's first paper on quantum mechanics, Ann. Phys. **79**, 361 (1926). In it, he wrote down Eq. (C6) and then (C9), with little motivation. He could not attach any further significance to principle (C6), and in his second paper referred to the derivation of (C9) from (C6) as "incomprehensible." On the other hand, R. A. Fisher had already published his theory of information four years before, in Phil. R. Soc. London **222**, 309 (1922). Did Schrödinger not know of Fisher's work? For further details on Schrödinger in this regard, see C. G. Gray, Am. J. Phys. (to be published).

[18]G. A. Korn and T. M. Korn, Ref. 7, p. 355.

[19]H. A. Bethe and R. W. Jackiw, *Intermediate Quantum Mechanics* (Benjamin, New York, 1968).

[20]H. A. Bethe and R. W. Jackiw, Ref. 18, p. 343.

[21]H. A. Bethe and R. W. Jackiw, Ref. 18, p. 345.

[22]L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1955), p. 321.

[23]H. A. Bethe and R. W. Jackiw, Ref. 18, p. 373.

[24]H. A. Bethe and R. W. Jackiw, Ref. 18, p. 341.