Information Entropy and Uncertainty Relations

Henry E. Montgomery, Jr.

Centre College, Danville, KY 40422-1394, mntgmry@centre.edu

Received July 17, 2002. Accepted September 25, 2002.

Abstract: Information entropy is introduced as a measure of quantum mechanical uncertainty. An uncertainty relation based on information entropy is obtained as an alternative to the Heisenberg inequality. In two typical examples, the entropic uncertainty relation is shown to be bounded in situations where the Heisenberg inequality diverges or grows too large to be useful.

The Heisenberg uncertainty *principle* [1, 2] introduces the idea that the probability distributions for the one-dimensional momentum, p_x, and position, x, of a quantum mechanical particle cannot be arbitrarily localized. This is usually expressed as the Heisenberg *inequality*

$$\Delta p_x \Delta x \geq \frac{\hbar}{2} \qquad (1)$$

Although Heisenberg did not give general definitions for the uncertainties Δp_x and Δx, they are usually associated with the standard deviation of a set of measurements of the position and momentum.

In this paper we discuss some limitations of the Heisenberg inequality that are a consequence of using the standard deviation as a measure of uncertainty. We show how a definition of uncertainty based on the definition of entropy from information theory leads to an alternative uncertainty relation. We then use the entropic uncertainty relation to discuss two model problems specifically chosen to illustrate the shortcomings of the Heisenberg inequality.

Uncertainty relations

For a quantum mechanical system described by a normalized wavefunction ψ, $\langle\Omega\rangle$, the average value of the operator $\hat{\Omega}$ is given by

$$\langle\Omega\rangle = \int \psi^* \hat{\Omega} \psi \, d\tau$$

where the integration is over the entire coordinate space. The standard approach to uncertainty relations is expressed by the Robertson inequality [3]

$$(\Delta A)(\Delta B) \geq \frac{1}{2}\left|\int \psi^* [\hat{A},\hat{B}] \psi \, d\tau\right| \qquad (2)$$

where $\hat{A}$ and $\hat{B}$ are Hermitian operators and their commutator $[\hat{A},\hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}$. $\Delta\Omega$ is the standard deviation of the operator $\hat{\Omega}$, given by

$$\Delta\Omega = \left[\int \psi^* \left(\hat{\Omega} - \langle\Omega\rangle\right)^2 \psi \, d\tau\right]^{1/2}$$

For one-dimensional momentum ($\hat{p}_x = -i\hbar\partial/\partial x$) and position ($\hat{x} = x$), the commutator $[\hat{p}_x, \hat{x}] = -i\hbar$ and the Robertson inequality reduces to the Heisenberg inequality.

If $[\hat{A},\hat{B}] = k$ where k is a constant, then the right-hand side of the Heisenberg inequality is *k/2* and a knowledge of ΔA gives an estimated lower bound for ΔB using

$$\Delta B \geq \frac{k}{2\Delta A}$$

However, if $[\hat{A},\hat{B}] = \hat{C}$ where $\hat{C}$ is another operator, then we must know ψ in order to evaluate the right-hand side of eq 2. But if we know ψ, ΔB can be calculated directly. If the right-hand side of eq 2 were always a constant, the resulting inequality would be stronger and thus more useful.

The derivation of the Robertson inequality is based on defining uncertainty as a measure of the distribution of values about their average value. This definition is consistent with the usual method of determining experimental uncertainty. It is most appropriate when the distribution of values is near-Gaussian. If the distribution has more than one peak, the standard deviation is not a good measure of uncertainty. As discussed by Hilgevoord [4], even if there is only one peak, the standard deviation may be a poor estimate of uncertainty if the distribution deviates significantly from Gaussian. It is reasonable to investigate measures of uncertainty other than the standard deviation and see if they lead to other uncertainty relations.

Intuitively, we equate uncertainty with a lack of information. It should come as no surprise that information theory provides a way to measure uncertainty. Information theory has its primary roots in two classic papers written by Claude Shannon in 1948 [5]. Shannon's purpose was to develop a mathematical theory to quantitatively analyze the passage of information from a source, through an information channel to a receiver. It has subsequently been applied to areas ranging from calculation of the ability of a material to be penetrated by charged particles [6] to analysis of binding sites on nucleotide sequences [7].

If we have a message composed of n signals, each of which occurs with a probability p_i, Shannon defined the information associated with this discrete probability distribution as

$$H = -\sum_{i=1}^{n} p_i \ln(p_i) \quad (3)$$

By analogy to Boltzmann's formulation of entropy in statistical mechanics, Shannon called H the entropy of the signal distribution. The story is told that John von Neumann advised Shannon to use the term entropy because "no one knows what entropy really is, so in a debate you will always have the advantage" [8]. This probabilistic or information entropy measures the spread and sharpness of a probability distribution, independent of its actual values. For a continuous variable t and an associated probability density distribution $\rho(t)$ normalized so that

$$\int_{-\infty}^{\infty} \rho(t)dt = 1$$

the sum in eq 3 becomes an integral and the information entropy of ρ is given by

$$S_t = -\int_{-\infty}^{\infty} \rho(t)\ln\rho(t)\,dt \quad (4)$$

For the reader who desires to learn more about information theory, there are numerous resources. Chapter 9 of Goldstein and Goldstein [9] discusses the relationship between entropy and information at a level understandable by the nonscientist. Pierce [10] provides a more technical discussion and discusses the applicability to a variety of areas including art and psychology. Shannon's original papers [5] contain some difficult areas, but most of his work is straightforward, particularly in his development of the concept of entropy.

To better understand S_t, it will be helpful to consider a Gaussian distribution with mean $t = 0$ and standard deviation σ,

$$\rho(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-t^2}{2\sigma^2}\right) \quad (5)$$

Performing the integration in eq 4, we obtain

$$S_t = \frac{1}{2}\left[1+\ln(2\pi)\right] + \ln(\sigma) \quad (6)$$

As σ increases, $\rho(t)$ becomes less localized and S_t increases. This is consistent with the idea that decreased localization should result in increased entropy. In the case of a Gaussian distribution, σ and S_t are seen to provide similar information.

For a normalized, one-dimensional, position-space wavefunction $\psi(x)$, the conjugate momentum-space wavefunction $\phi(p)$ is given by the Fourier transform

$$\phi(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \exp\left(\frac{-ipx}{\hbar}\right)\psi(x)\,dx \quad (7)$$

With $\psi(x)$ we can associate a position-space density distribution $\rho(x) = \psi(x)^* \psi(x)$ and with $\phi(p)$ we can associate a momentum-space density distribution $\rho(p) = \phi(p)^* \phi(p)$. Using eq 4 we calculate the position-space and momentum-space information entropies as

$$S_x = -\int_{-\infty}^{\infty} \rho(x)\ln\rho(x)\,dx \quad (8)$$

and

$$S_p = -\int_{-\infty}^{\infty} \rho(p)\ln\rho(p)\,dp \quad (9)$$

Bialynicki-Birula and Mycielski (BBM) [11] used Fourier analysis to derive a relationship between S_x and S_p. Their derivation is synopsized in the appendix. They showed that S_x and S_p satisfy the relation

$$S_x + S_p \geq 1 + \ln\pi \cong 2.145 \quad (10)$$

We call this an entropic uncertainty relation.

Because the information entropy measures the localization of a distribution, eq 10 places a limit on the simultaneous localization of the position and momentum distributions. If one of the entropies becomes small, then the other must become large enough to preserve the inequality. This is philosophically consistent with the Heisenberg uncertainty principle. BBM [11] also showed that the Heisenberg inequality could be derived from their entropic uncertainty relation.

The BBM entropic uncertainty relation has a constant lower bound and thus overcomes one of the limitations of the Heisenberg inequality. These entropic uncertainty relations have recently received considerable interest in the literature. The interested reader is referred to Majerník and Richterek [12], Yáñez et al. [13], Majerník and Majerníková [14] and references therein.

Constant Wavefunction in Position Space

We can illustrate one of the shortcomings of the Heisenberg inequality by considering a quantum mechanical particle described by the normalized, position-space wavefunction $\psi(x) = 1/\sqrt{L}$ on $\hbar$ the interval $-L/2 \leq x \leq L/2$ and $\psi(x) = 0$ elsewhere. We will use units scaled such that $\hbar = m = 1$. This serves both to simplify the form of the equations and to emphasize that the entropy is a measure of the sharpness of the probability distribution, independent of dimension. Other ways of dealing with the issue of dimension are discussed in section 4 of [12]. The conjugate momentum-space wavefunction can be obtained from the Fourier transform, eq 7 as

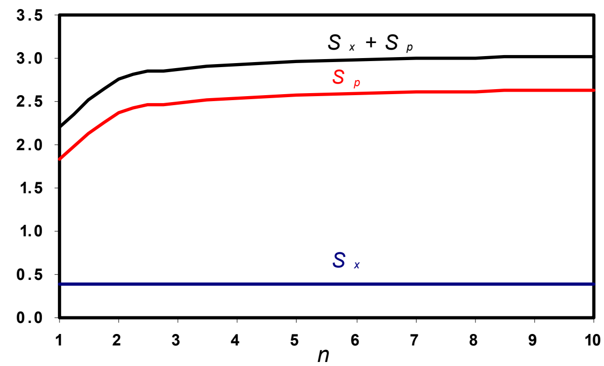

Figure 1. Information entropies for $L = 2$.

$$\phi(p) = \frac{1}{\sqrt{2\pi}}\int_{-L/2}^{L/2}\exp(-ipx)\frac{1}{\sqrt{L}}dx = \sqrt{\frac{2}{\pi L}}\frac{\sin(Lp/2)}{p} \quad (11)$$

The position-space information entropy is given by

$$S_x = -\int_{-L/2}^{L/2}\left(\frac{1}{\sqrt{L}}\right)^2\ln\left(\frac{1}{\sqrt{L}}\right)^2 dx = \ln(L) \quad (12)$$

consistent with the idea that the uncertainty in position increases as the interval length increases. The momentum-space information entropy results in the rather complicated integral

$$S_p = \int_{-\infty}^{\infty}\left\{\sqrt{\frac{2}{\pi L}}\frac{\sin\left(\frac{Lp}{2}\right)}{p}\right\}^2 \ln\left\{\sqrt{\frac{2}{\pi\ L}}\frac{\sin(Lp/2)}{p}\right\}^2 dp \quad (13)$$

Sánchez-Ruiz [15] showed that

$$\int_0^{\infty}\frac{\sin^2(x)}{x^2}\ln\left(\frac{\sin^2(x)}{x^2}\right)dx = \pi(\gamma-1) \quad (14)$$

where γ is Euler's constant $\gamma = 0.5772$. Thus $S_p = \ln(2\pi/L) + 2(1-\gamma)$. $S_p = 2.683$ at $L = 1$ and decreases to 0.381 at $L = 10$ while $S_x + S_p = \ln(2\pi) + 2(1-\gamma) = 2.683 \geq 1 + \ln\pi$.

Because the position- and momentum-space wavefunctions are symmetric about the origins of their respective spaces, $\langle p_x\rangle = \langle x\rangle = 0$. Calculation of the position-space standard deviation gives

$$\Delta x = \sqrt{\int_{-L/2}^{L/2}x^2\frac{1}{L}dx} = \frac{\sqrt{3}}{6}L \quad (15)$$

When we calculate the momentum-space standard deviation, we find

$$\Delta p_x = \sqrt{\int_{-\infty}^{\infty}p^2\left\{\sqrt{\frac{2}{\pi L}}\frac{\sin(Lp/2)}{p}\right\}^2 dp} = \infty \quad (16)$$

and $\Delta p_x\Delta x = \infty$. The Heisenberg inequality is certainly satisfied, but it is difficult to extract anything useful from the product $\Delta p_x\Delta x$.

Infinite Square Well

We consider a well of length L, centered at the origin, with a quantum mechanical particle confined to the interval $-L/2 \leq x \leq L/2$. We will again use units scaled such that $\hbar = m = 1$. The system is symmetric about $x = 0$ and the solutions to the Schrödinger equation divide into symmetric cosine functions and asymmetric sine functions. Because the Fourier transform of a sine function results in a complex-valued function, we will only consider the symmetric states

$$\psi_n(x) = \sqrt{\frac{2}{L}}\cos\left(\frac{(2n-1)\pi x}{L}\right) \quad (17)$$

where n = 1,2,3,…. The conjugate momentum-space wavefunctions are given by

$$\phi_n(p) = 2(-1)^{n+1}\sqrt{\pi\ L}\frac{(2n-1)\ \cos(Lp/2)}{\pi^2(2n-1)^2 - p^2L^2} \quad (18)$$

Calculation of the entropic uncertainties requires evaluation of integrals involving logarithms of trigonometric functions

$$S_x = \int_{-L/2}^{L/2}\left[\frac{2}{L}\cos^2\left(\frac{(2n-1)\pi x}{L}\right)\right]\ln\left[\frac{2}{L}\cos^2\left(\frac{(2n-1)\pi x}{L}\right)\right] = \ln(2L) - 1 \quad (19)$$

independent of the value of n. S_p can be evaluated using numerical integration. Figure 1 shows the information entropies and their sums for $L = 2$ and increasing n. S_p appears to be approaching an asymptotic limit which is slightly greater than 2.6. Majerník et al. [16] have shown that the limiting value is ≈2.6564. We see that $S_x + S_p$ ranges from 2.2120 to 3.0175 so the BBM inequality is satisfied.

Because the position- and momentum-space wavefunctions are again symmetric about the origins of their respective spaces, $\langle p_x\rangle = \langle x\rangle = 0$. The standard deviations are readily evaluated [17] as

$$\Delta x = \int_{-L/2}^{L/2}\left[\sqrt{\frac{2}{L}}\cos\left(\frac{(2n-1)\pi x}{L}\right)\right]x^2\left[\sqrt{\frac{2}{L}}\cos\left(\frac{(2n-1)\pi x}{L}\right)\right]dx = \frac{\sqrt{3}L}{6}\sqrt{1-\frac{6}{\pi^2(2n-1)^2}} \quad (20)$$

and

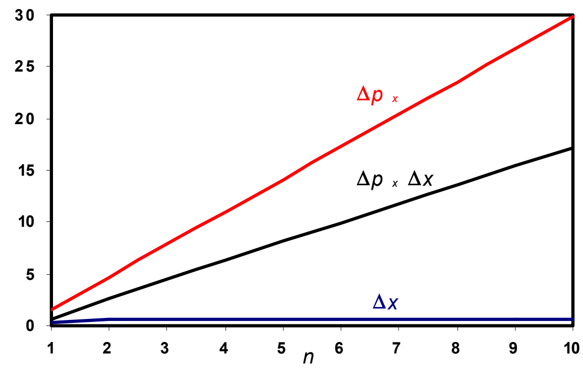

Figure 2. Standard deviations for $L = 2$.

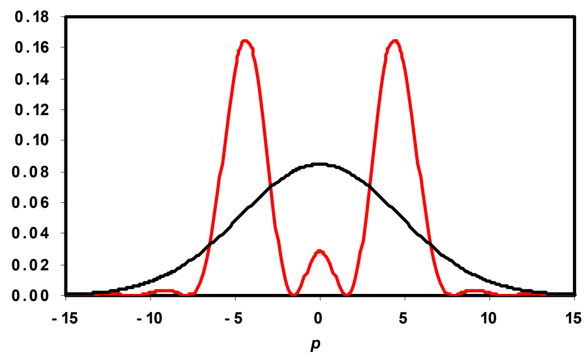

Figure 3. $\rho_2(p)$, the momentum-space distribution for $n = 2$, $L = 2$, and a Gaussian distribution having the same standard deviation.

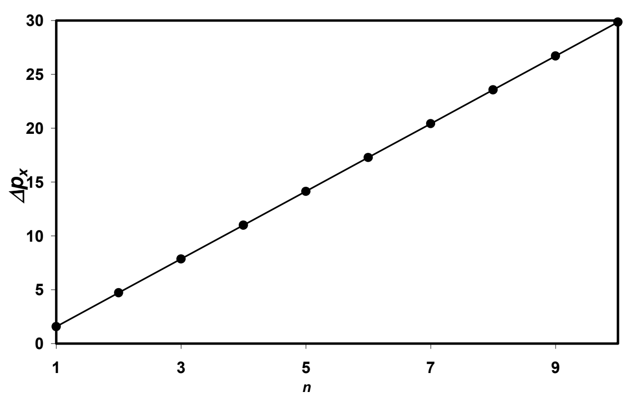

Figure 4. Momentum standard deviation dependence on n.

$$\Delta p_x = \left[2\sqrt{\pi L}\frac{(2n-1)\cos(Lp/2)}{\pi^2(2n-1)^2 - p^2L^2}(-1)^{n+1}\right]p^2 \times\left[2\sqrt{\pi L}\frac{(2n-1)\cos(Lp/2)}{\pi^2(2n-1)^2 - p^2L^2}(-1)^{n+1}\right]dx = \frac{\pi}{L}(2n-1) \quad (21)$$

As n increases, Δx asymptotically approaches $\left(\sqrt{3}/6\right)L$. The increase in Δp_x is linear with n and the product of the standard deviations is

$$\Delta p_x \Delta x = \frac{1}{2\sqrt{3}}\sqrt{\pi^2(2n-1)^2 - 6} \quad (22)$$

independent of L. Figure 2 shows the standard deviations and their product.

Although S_p and Δp are both measures of momentum uncertainty, as n increases they behave in strikingly different ways. For $n > 1$ the momentum density distribution, $\rho_n(p)$, has two distinct peaks, symmetrically located above and below the mean. $\rho_2(p)$ is shown in Figure 3. Superimposed on the graph is a Gaussian distribution having the same standard deviation as $\rho_2(p)$. Given the difference between the momentum density distribution and the Gaussian "fit" to it by the standard deviation, it is not surprising that the standard deviation is not a good measure of the momentum uncertainty. As n increases, the separation between the peaks increases. Figure 4 shows that the increase in standard deviation is linear with increasing n. The uncertainty measured by the standard deviation is primarily a measure of the separation between the peaks of the momentum density distribution.

The momentum-space information entropy measures the area under the density distribution and primarily depends on the area under the two main peaks. Thus, it is bounded and well-behaved. Because $S_x = \ln(2L) - 1$, we have a quantitative relation between the position-space and momentum-space uncertainties, independent of the value of n.

Summary

We have presented a formulation of uncertainty based on information entropy and shown how entropic uncertainty relations are bounded in situations where the Heisenberg inequality is not bounded. If we consider $\Delta p_x \Delta_x$ and $S_x + S_p$ to measure orbits in their respective phase spaces, the former are sometimes not bounded whereas the latter are bounded. Further, the entropic uncertainty relations always have a finite lower bound.

The only case of which the author is aware where the Heisenberg inequality is bounded and the entropic uncertainty relation is unbounded is a distribution composed of separated Dirac delta functions. Everett [18] pointed out that such a distribution satisfies the Heisenberg inequality but results in infinite information entropy. For typical quantum chemical distributions, such as atomic wavefunctions, the position-space and momentum-space wavefunctions are well behaved and the entropic uncertainty relations lead to finite bounds [19].

This material should be useful in a classroom discussion of uncertainty and uncertainty principles. It also provides examples that give a student the opportunity for quantitative calculations involving one of quantum theory's fundamental concepts.

Acknowledgment. The author is grateful to P. M. Wilt and K. Dutch for their comments on this manuscript and to K. D. Sen for introducing him to entropic uncertainty relations. He would also like to acknowledge the helpful comments of the anonymous reviewers.

Appendix

Consider a one-dimensional function $f(x)$ and its Fourier transform $g(k)$ where $f(x)$ and $g(k)$ are normalized so that

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} g(k)\,dk = 1 \qquad \text{(A1)}$$

In 1957, Everett [18] and Hirshman [20] independently conjectured, but did not prove, that

$$-\int_{-\infty}^{\infty} f(x)\ln\left[f(x)\right]dx - \int_{-\infty}^{\infty} g(k)\ln\left[f(k)\right]dk \geq 1+\ln(\pi) \qquad \text{(A2)}$$

Their conjecture was based on the observation that if $f(x)$ and $g(k)$ are Gaussian functions, the inequality reduces to an equality whereas any variation from a Gaussian function increases the left side of the inequality. Analogous behavior for Gaussian functions is also noted for the Heisenberg inequality.

Almost 20 years later BBM [11] proved eq A2 using Fourier analysis. Their proof is synopsized below. Define

$$\|f\|_p = \left[\int_{-\infty}^{\infty} |f(x)|^p\,dx\right]^{\frac{1}{p}} \text{ and } \|g\|_q = \left[\int_{-\infty}^{\infty} |g(x)|^q\,dx\right]^{\frac{1}{q}} \qquad \text{(A3)}$$

The (p,q) norm of this Fourier transform pair is defined as the smallest number $k(p,q)$ for which the inequality

$$\|g\|_q = k(p,q)\|f\|_p \qquad \text{(A4)}$$

holds, where

$$\frac{1}{p}+\frac{1}{q} = 1 \text{ and } q \geq 2 \qquad \text{(A5)}$$

Beckner [21] showed that

$$k(p,q) = \left(\frac{2\ \pi}{q}\right)^{\frac{1}{2\ q}}\left(\frac{2\ \pi}{p}\right)^{\frac{-1}{2\ p}} \qquad \text{(A6)}$$

Writing the difference from eq A4 as

$$W(q) = k(p,q)\|f\|_p - \|g\|_q \qquad \text{(A7)}$$

and expressing p as a function of q, BBM showed that the derivative of $W(q)$ evaluated at $q = 2$ reduces to the inequality of eq A2.

References and Notes

1. Heisenberg, W. *Z. für Phys.* **1927,** *43,* 172–198.
2. Heisenberg, W. *The Physical Principles of the Quantum Theory;* Translated by Eckart, C.; Hoyt, F. C.; Dover: New York, 1949; pp 13–46.
3. Robertson, H. P. *Phys. Rev.* **1929,** *34,* 163–164.
4. Hilgevoord, J. *Am J. Phys.* **2002,** *70,* 983.
5. Shannon C. E. *Bell Syst. Tech. J.* **1948,** *27,* 379–473; 623–656. Reprinted in Shannon, C. E.; Weaver, W. *The Mathematical Theory of Communication;* Illinois Press: Urbana, 1949; pp 29–125 and http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html (accessed Oct 2002).
6. Hồ, M.; Weaver, D. F.; Smith, V. H.; Sagar, R. P.; Esquivel, R. O. *Phys. Rev. A* **1998,** 57, 1412–4517.
7. Schneider, T. D.; Stormo, G. D.; Gold, L.; Ehrenfeucht, A. *J. Mol. Biol*. **1986,** *188,* 415–431.
8. Tribus, M.; McIrvine, E. C. *Sci. Am.* **1971,** *225,* 179–188.
9. Goldstein, M.; Goldstein, I. F. *The Refrigerator and the Universe;* Harvard University Press: Cambridge, 1993; pp 212–228.
10. Pierce, J. R. *An Introduction to Information Theory: Symbols, Signals and Noise;* Dover: New York, 1980.
11. Bialynicki-Birula, I.; Mycielski, J. *Comm. Math. Phys.* **1975,** *44,* 129–132.
12. Majerník, V.; Richterek, L. *Eur. J. Phys.* **1997,** *18*, 79–89.
13. Yáñez, R. J.; Van Assche, W.; Dehesa, J. S. *Phys. Rev. A* **1994,** *50,* 3065–3079.
14. Majerník, V.; Majerníková, E. *J. Phys. A: Math. Gen.* **2002,** *35,* 5721–5761.
15. Sánchez-Ruiz, J. *Phys. Rev. A* **1998,** 57*,* 1519–1525.
16. Majerník, V.; Charvot, R; Majerníková, E. *J. Phys. A: Math. Gen*. **1999,** *32,* 2207–2216.
17. Mulder, J. C. *Chem. Educator* [Online], **2002,** *7*(2): DOI 10.1007/s00897020545a.
18. Everett, H., In *The Many-Worlds Interpretation of Quantum Mechanics;* DeWitt, B. S. and Graham, N., Eds.; Princeton University Press: Princeton, 1973; pp 51–52.
19. Gadre, S. H.; Sears, S.B. *Phys. Rev. A* **1985,** *32,* 2602–2606.
20. Hirschman, I. I. *Am. J. Math.* **1957,** *79,* 152–156.
21. Beckner, W. *Proc. Nat Acad. Sci. USA* **1975,** *72,* 638–641.