Information theory and electron density

István Kolossváry
Department of Chemical Information Technology,
Technical University of Budapest, H-1521 Budapest, Hungary
E-mail:kolos@dumbo.inc.bme.hu

Received 3 October 1995

The intriguing concept of inherent uncertainty of probability schemes in information theory and statistical inference is applied to the molecular electron density. The electron density function is treated as a multimodal, three-dimensional probability density function describing the distribution of the electrons of a molecule in real space. A simple theory is proposed to introduce the amount of information associated with perturbations of the nuclear geometry such as molecular vibrations and reaction paths, in particular. It is shown by computations that the amount of information associated with the normal modes of vibration is related to the reduced mass. The proposed theory also suggests a novel Riemannian nuclear configuration space which is completely defined by the observable electron density of a molecular system.

1. Introduction

In this paper, the author proposes a theory furnishing a link between information theory and molecular electron density. The theory is based on the well-known concept of intrinsic uncertainty of probability schemes in information theory and statistical inference. The electron density function is treated as a multimodal, three-dimensional probability density function describing the distribution of the electrons of a molecule in real space. The modi of the electron density distribution appear at or near the atomic centers. The electron density function varies with the nuclear geometry, and translates and rotates together with the nuclei. The fact that the Heisenberg uncertainty principle does not allow the atoms to reside at an exact location is analogous to the common situation in statistics where the unknown center(s) of the distribution of a statistical population can only be estimated by some statistic applied to random samples. Consider a hypothetical experiment where points in three-dimensional space are picked randomly with respect to the electron density function (as a probability density function), and points near to the atomic centers of a molecule are picked more frequently than points far away from the atoms. This means the statistical population represented by the electron density function will be sampled and the location of the atomic centers can be estimated in an unbiased manner by, e.g., calculating the centroid of the clusters of the randomly drawn sample-points.

The Cramér–Rao theorem [1] imposes an intrinsic limitation on such statistical estimations based on the so-called Fisher information [2] – a measure of uncertainty – analogous to the Heisenberg uncertainty principle. The application of the Cramér–Rao theorem to the electron density function reveals the inherent uncertainty of the location of the atoms of a molecule in three-dimensional space. Indeed; although in a very different manner, the analogy of the Cramér–Rao theorem and the Heisenberg uncertainty principle has already been recognized in physics. Caianiello introduced a quantum-mechanical analogue of the Fisher information and used the corresponding quantum counterpart of the Cramér–Rao theorem to derive the Heisenberg uncertainty relationships [3,4].

The so-called Fisher information matrix [1] of the electron density is introduced. The Fisher information matrix gives rise to a unique interpretation of "information" associated with a small perturbation of a particular nuclear configuration. Perturbations of the nuclear configuration which represent *concerted* atomic displacements, such as vibrational modes and reaction coordinates, are of particular interest. The proposed theory suggests a classification of vibrational modes in terms of their information content. The Fisher information matrix also defines a unique (information) distance along reaction paths representing chemical reactions or conformational interconversions. The proposed theory is elucidated in a computational working example.

2. Theory

THE CRAMÉR–RAO THEOREM

A parametric family of probability density functions $p(\mathbf{r}, \mathbf{q})$ is defined in the random space of the $\mathbf{r} = (r^1, r^2, \ldots, r^n)$ random variables with some fixed set of the $\mathbf{q} = (q^1, q^2, \ldots, q^n)$ parameters. The Cramér–Rao theorem provides, subject to certain regularity conditions, a lower bound for the variance of any unbiased estimator $\tilde{q}^i$ of q^i [1]:

$$\operatorname{var}(\tilde{q}^i) \geqslant \left(\int p(\mathbf{r}, \mathbf{q}) \left(\frac{\partial \log p(\mathbf{r}, \mathbf{q})}{\partial q^i} \right)^2 d\mathbf{r} \right)^{-1} . \tag{1}$$

The expectation value in the outer parentheses on the right-hand side of eq. (1) is called the Fisher information about q^i in a single observation of a particular set of the $\mathbf{r}$ random variables [1,2]. The more information about q^i provided on average by a single observation, the smaller the lower bound for the variance of its estimator $\tilde{q}^i$. Of course, the central limit theorem assures a diminishing $\operatorname{var}(\tilde{q}^i)$ by the well-known factor of $1/\sqrt{N}$ in multiple observations so that $\operatorname{var}(\tilde{q}^i)$ approaches zero in the limit as the number N of observations goes to infinity.

For the univariate normal distribution $p(\mathbf{r}, \mathbf{q}) = 1/\sqrt{2\pi\sigma^2} \exp(-(x-\mu)^2/2\sigma^2)$, as an example, $\mathbf{r} = (x)$ and $\mathbf{q} = (\mu, \sigma)$. The estimators $(\tilde{\mu}, \tilde{\sigma})$ are equal to the sample mean and the sample standard deviation, respectively. The Fisher information with respect to μ is equal to $2/\sigma^2$, and the Fisher information with respect to σ is equal to $1/\sigma^2$. This example shows the meaning of Fisher information very intuitively; the inherent statistical uncertainty, represented by σ, imposes a limit on the variance of the mean and standard deviation of a random sample.

The Fisher information matrix is a straightforward multidimensional generalization of the Fisher information. The (i,j)th element of the symmetrical $n \times n$ Fisher information matrix is [1]

$$f_{ij} = \int p(\mathbf{r}, \mathbf{q}) \left(\frac{\partial \log p(\mathbf{r}, \mathbf{q})}{\partial q^i} \right) \left(\frac{\partial \log p(\mathbf{r}, \mathbf{q})}{\partial q^j} \right) d\mathbf{r} \,. \tag{2}$$

INFORMATION THEORY

The significance of the Fisher information matrix is its intimate relationship to cross-entropy. In the context of the proposed theory, cross-entropy (often termed relative entropy of Kullback–Leibler information) is the information provided, on average, in a single observation to discriminate between two closely related distributions $p(\mathbf{r}, \mathbf{q})$ and $p(\mathbf{r}, \mathbf{q} + \Delta\mathbf{q})$, respectively [5,6]:

$$I(\mathbf{q}, \mathbf{q} + \Delta\mathbf{q}) = \int p(\mathbf{r}, \mathbf{q} + \Delta\mathbf{q}) \log \frac{p(\mathbf{r}, \mathbf{q} + \Delta\mathbf{q})}{p(\mathbf{r}, \mathbf{q})} d\mathbf{r} \,. \tag{3}$$

It can be shown that subject to regularity conditions, the Taylor expansion of eq. (3), to within second order terms, reveals that cross-entropy can be expressed in terms of the Fisher information matrix [5]:

$$I(\mathbf{q}, \mathbf{q} + \Delta\mathbf{q}) = \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} \Delta q^i \Delta q^j + O\left((\Delta\mathbf{q})^3 \right) . \tag{4}$$

INFORMATION GEOMETRY

Since the Fisher information matrix is symmetric, positive definite, and invariant under coordinate (parameter) transformations, the quadratic form in eq. (4) defines a suitable distance measure, termed information metric [7–10]. Rao [7] introduced the information metric to establish a distance measure ds^2 used to quantify the similarity between two infinitesimally different distributions $p(\mathbf{r}, \mathbf{q})$ and $p(\mathbf{r}, \mathbf{q} + d\mathbf{q})$, respectively:

$$ds^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} dq^i dq^j \,. \tag{5}$$

The information metric allows for a fully geometrical interpretation and generalization of cross-entropy. Note that cross-entropy as defined in eq. (3), is not symmetrical with respect to swapping $p(\mathbf{r},\mathbf{q})$ with $p(\mathbf{r},\mathbf{q}+\Delta\mathbf{q})$. However, cross-entropy at the infinitesimal scale (eqs. (4) and (5)) is symmetrical. The information metric ds^2 in eq. (5) defines a Riemannian parameter space where the so-called information distance can be defined [7–10]. The information distance between two distributions is defined as the geodesic distance between their two representative points in the Riemannian parameter space. The information distance between two different univariate normal distributions, e.g., is equal to [10]

$$2\sqrt{2}\tanh^{-1}((\mu_1-\mu_2)^2+2(\sigma_1-\sigma_2)^2/(\mu_1-\mu_2)^2+2(\sigma_1+\sigma_2)^2)^{1/2}\,.$$

ELECTRON DENSITY

The electron density function $\rho(\mathbf{r},\mathbf{q})$ is treated as a parametric family of multimodal, three-dimensional probability density functions describing the distribution of the electrons. The modi of the electron density distribution appear at or near the atomic centers. The "random" variables are the $\mathbf{r}=(x,y,z)$ points of the three-dimensional space and the parameters are the $\mathbf{q}=(q^1,q^2,\ldots,q^n)$ nuclear coordinates. The shape of the electron density function varies with the nuclear geometry and, also, translates and rotates together with the nuclei. Different nuclear configurations (molecular geometries) and/or different molecular positions in the three-dimensional space are represented by different members of the parametric family of $\rho(\mathbf{r},\mathbf{q})$ electron density functions.

The Fisher information matrix of the electron density associated with a particular nuclear configuration is defined as follows:

$$f_{ij}=\frac{1}{n_e}\int\rho(\mathbf{r},\mathbf{q})\left(\frac{\partial\log\rho(\mathbf{r},\mathbf{q})}{\partial q^i}\right)\left(\frac{\partial\log\rho(\mathbf{r},\mathbf{q})}{\partial q^j}\right)d\mathbf{r}\,. \tag{6}$$

The logarithmic derivatives of the electron density function with respect to the q^i $(i=1,\ldots,n)$ nuclear coordinates are calculated at a particular nuclear configuration and the region of integration is the entire three-dimensional space. The integral is normalized by the number of electrons n_e. The $n\times n$ Fisher information matrix is symmetric and positive definite; moreover, the choice of nuclear coordinates is immaterial, because the Fisher information matrix is invariant under coordinate transformations (see above). Nonetheless, internal and external coordinates should be distinguished, because internal coordinates do not take into account that electron density translates and rotates together with the nuclei.

It is straightforward to apply eqs. (4) and (5) using the electronic Fisher information matrix of eq. (6). The parameter space of the electronic Fisher information matrix is nothing but the nuclear configuration space. Different (electron) distributions correspond to different nuclear configurations, that is, to different molecular geometries. The Riemannian nuclear configuration space induced by the information metric (eq. (5)), based on the electronic Fisher information matrix (eq. (6)), is uniquely defined by the observable electron density of a molecular system.

An important difference, nevertheless, between information theory and the proposed theory should be noted. In information theory, the path connecting two distributions in parameter space is usually irrelevant. The geodesic distance is usually associated with the shortest possible path. In the proposed theory, however, a path in nuclear configuration space has physical significance: it corresponds to a perturbation of the nuclear configuration. Perturbations of the nuclear configuration which represent *concerted* atomic displacements such as vibrational modes and reaction coordinates, are of particular interest. The author therefore suggests the following applications of the electronic Fisher information matrix:

$$I_{vib} = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} \Delta q^i \Delta q^j \,. \tag{7}$$

I_{vib} is the Fisher information associated with a particular vibrational mode represented by the corresponding unit-length eigenvector $\Delta \mathbf{q}$ of the mass-weighted Hessian matrix. The components of $\Delta \mathbf{q}$ represent the individual atomic displacements in that particular vibrational mode. The Fisher information matrix f_{ij} is defined in eq. (6), and the number of nuclear coordinates n depends on the coordinate system. Note that I_{vib} is invariant under coordinate transformations.

Fisher information associated with a particular reaction path can also be defined by integrating eq. (5) along the reaction path:

$$I_{rc} = \int_C \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} \frac{dq^i}{ds} \frac{dq^j}{ds}} ds \,. \tag{8}$$

I_{rc} is defined in eq. (8) as the curve-integral of the square root of the information metric. The curve describes a particular reaction path in the nuclear configuration space. Note that this particular form of eq. (8) assumes that the curve is parameterized with respect to arc length. I_{rc} is also invariant under coordinate transformations.

3. Computations

The proposed theory is elucidated in a computational working example involving ammonia, formaldehyde, and hydrogen peroxide. The calculations were completed on a Hewlett-Packard 9000/705 workstation running the Gaussian 92/DFT program package [11]. The wavefunctions were calculated by density functional theory using the 6-311G** basis set at fully optimized geometries. Density functional theory was employed as defined by the "B3P86" Gaussian keyword. The Fisher information matrix was calculated utilizing a modified version of the Extreme program of the AIMPACK package [12]. Extreme was used for extracting the electron density from the Gaussian wavefunction and additional programs were written by the author to calculate the electron density derivatives with respect to the Cartesian nuclear coordinates and, to render the integration. The electron density derivatives were calculated numerically employing central differences. The three-dimensional integrals were also calculated numerically employing a spherical quadrature comprised of concentric shells (radial part: Gauss–Laguerre quadrature [13], angular part: Lebedev–Skorokhodov quadrature [14]).

The results of the calculations with respect to I_{vib} are summarized in Table 1. It is noteworthy that Fisher information is characteristic of the vibrations in terms of the involvement of heavy atoms in a particular vibrational mode. This is clearly shown by the correlation between I_{vib} and the reduced mass. For ammonia, it is the *s*-deform vibration possessing the largest amount of Fisher information. Note that the *s*-deform vibration is primarily responsible for the ammonia inversion. Formaldehyde and hydrogen peroxide are both displaying vibrations involving mainly heavy atoms; CO stretch and OO stretch, respectively. Note that the associated Fisher information I_{vib} is in both cases larger than those of the remaining vibrations, by an order of magnitude. The corresponding reduced masses are also

Table 1
Fisher information along the vibrational modes of ammonia, formaldehyde and hydrogen peroxide.

	Vib. mode [a]	Freq. [b] (cm^{-1})	Red. mass (amu)	I_{vib} [c]
Ammonia	*s*-deform	1064	1.18	.211
	d-deform	1676	1.07	.112
	s-stretch	3485	1.03	.138
	d-stretch	3612	1.09	.193
Formaldehyde	CH_2 wag	1201	1.37	.193
	CH_2 rock	1269	1.35	.199
	CH_2 scis.	1537	1.10	.127
	CO stretch	1845	7.78	4.10
	CH_2 *s*-stretch	2890	1.04	.104
	CH_2 *a*-stretch	2944	1.12	.143
Hydrogen peroxide	torsion	368	1.08	.0986
	OO stretch	976	14.4	3.81
	OH bend	1310	1.10	.0658
	OH bend	1461	1.09	.112
	OH stretch	3815	1.07	.0913
	OH stretch	3816	1.07	.116

[a] The vibrational modes are depicted with the standard spectroscopical terms.

[b] The frequencies were calculated analytically with the B3P86/6-311G** wavefunction [11].

[c] The step size for numerical differentiation was set to 0.001 Å. The radial part of the numerical integration employed the 250 point Gauss-Laguerre quadrature ($\alpha = 2$) [13] with a 5 Å cutoff, resulting in a reduced set of the innermost 30 shells. The angular part of the integration employed the largest Lebedev–Skorokhodov grid containing 986 integration points [14] per shell. The unit of Fisher information is $1/distance^2$ in atomic units.

much larger than those associated with vibrations primarily involving the hydrogen atoms.

4. Concluding remarks

The correlation between I_{vib} and the reduced mass can be attributed to the electron density derivatives in eq. (6). It is intuitively clear that displacement of a heavy atom results in a more significant variation of the electron density than displacement of a light atom, i.e., hydrogen. Therefore, the variation of the electron density along a particular direction in the nuclear configuration space according to eq. (7) depends on the relative amount of heavy atom displacements (reduced mass) in that particular vibrational mode. It remains to be seen, however, whether a model can be found to establish the link between Fisher information and the vibrational modes in analytical form.

Acknowledgements

It is my privilege to thank P.G. Mezey for encouragement and Zs. Révay for carefully reading the manuscript. I am also indebted to M. Thoma, D. Petz and W.C. Guida for inspiring discussions. Financial support from the Hungarian Research Foundation (OTKA W15587 and F16263) is acknowledged.

References

[1] S.D. Silvey, *Statistical Inference* (Chapman and Hall, 1975) pp. 21–45.

[2] R.A. Fisher, Proc. Cambridge Philos. Soc. 122 (1925) 700.

[3] E.R. Caianiello, in: *Frontiers of Nonequilibrium Statistical Physics*, eds. G.T. Moore and M.O. Scully (Plenum Press, 1986) pp. 453–464.

[4] E.R. Caianiello and W. Guz, Phys. Lett. A126 (1988) 223.

[5] S. Kullback, *Information Theory and Statistics* (Wiley, 1959) pp. 1–45.

[6] E.S. Soofi, J. Am. Stat. Assoc. 89 (1994) 1243.

[7] C.R. Rao, Bull. Calcutta Math. Soc. 37 (1945) 81.

[8] E.R. Caianiello, Lett. Nuovo Cimento 38 (1983) 539.

[9] S.-I. Amari, *Differential-Geometrical Methods in Statistics* (Springer, 1985).

[10] S.-I. Amari, O.E. Barndorff-Nielsen, R.E. Kass, S.L. Lauritzen and C.R. Rao, *Differential Geometry in Statistical Inference* (Institute of Mathematical Statistics, 1987).

[11] *Gaussian 92/DFT Revision G.2* (1993), available from: Gaussian Inc., Carnegie Office Park, Building 6, Pittsburgh, Pennsylvania 15106, USA.

[12] *AIMPAC*, available from: Prof. R.F.W. Bader, McMaster University, Department of Chemistry, Hamilton, Ont., Canada, L8S 4L8.

[13] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes in C Second Edition* (Cambridge University Press, 1992).

[14] V.I. Lebedev and A.L. Skorokhodov, Russian Acad. Sci. Dokl. Math. 45 (1992) 587.