Fisher Information and Statistical Mechanics

Tech. Note 008v4
http://threeplusone.com/fisher

Gavin E. Crooks

January 4, 2012

Fisher information is an important concept in statistical estimation theory and information theory, but it has received relatively little consideration in statistical physics. In order to rectify this oversight, in this brief note I will review the correspondence between Fisher information and fluctuations at thermodynamic equilibrium, and discuss various applications of Fisher information to equilibrium and non-equilibrium statistical mechanics.

Given a family of probability distributions $p(x|\lambda)$ that vary smoothly with a parameter λ, then the Fisher information [1] is defined as

$$I(\lambda) = \sum_x p(x|\lambda) \left(\frac{\partial \ln p(x|\lambda)}{\partial \lambda} \right)^2 \qquad (1)$$
$$= \left\langle \left(\frac{\partial \ln p(x|\lambda)}{\partial \lambda} \right)^2 \right\rangle$$

Since the mean of the "score function" $\langle \partial \ln p / \partial \lambda \rangle$ is zero (See appendix), it follows that the Fisher information is the variance of the score.

The Fisher information is also equal to the negative mean second derivative of the score, under certain conditions (See appendix).

$$I(\lambda) = -\left\langle \frac{\partial^2}{\partial \lambda^2} \ln p(x|\lambda) \right\rangle \qquad (2)$$

The probability distribution of a system at thermal equilibrium is given by the canonical ensemble [2, 3]

$$\rho(x|\lambda) = \exp\Big(\beta F(\lambda) - \beta E(x,\lambda) \Big) , \qquad (3)$$

where $\beta = 1/k_B T$ is the inverse temperature T of the environment in natural units (k_B is the Boltzmann constant), $E(x,\lambda)$ is the energy of the system, which depends both on the internal state x and the external control λ, and F is the free energy

$$\beta F = -\ln \sum_x \exp\{-\beta E(x,\lambda)\} . \qquad (4)$$

Alternatively, instead of the free energy, we could write equivalent equations using the free entropy $\psi = -\beta F$ or the partition function $Z = \exp\{-\beta F\}$.

If we plug the canonical ensemble (3) into the Fisher information (1) we find that

$$I(\lambda) = \beta^2 \left\langle \left(\frac{dF}{d\lambda} - \frac{\partial E}{\partial \lambda} \right)^2 \right\rangle . \qquad (5)$$

From Eq. (4) it follows that the derivative of the free energy with respect to the control parameter, $dF/d\lambda$, is equal to the average derivative of the energy with respect to the same parameter.

$$\frac{dF}{d\lambda} = \left\langle \frac{\partial E}{\partial \lambda} \right\rangle \qquad (6)$$

Therefore, the Fisher information of a canonical ensemble with respect to a parameter λ that smoothly (but otherwise arbitrarily) controls the energy, is the variance of the infinitesimal change in energy with respect to a change in the control parameter.

$$I(\lambda) = \beta^2 \left\langle \left(\left\langle \frac{\partial E}{\partial \lambda} \right\rangle - \frac{\partial E}{\partial \lambda} \right)^2 \right\rangle \qquad (7)$$

If we consider the inverse temperature as the controllable parameter, then the Fisher information is equal to the energy fluctuations.

$$I(\beta) = \left\langle \big(\langle E \rangle - E \big)^2 \right\rangle \qquad (8)$$

In other words, the Fisher information of a thermodynamic system tells us the size of fluctuations about equilibrium.

Let us consider two examples. First, suppose the system under examination is a single polymer, and the parameter under control is the end-to-end distance L. The instantaneous tension $\mathcal{T} = \frac{\partial E}{\partial L}$ is the force exerted by the polymer on the apparatus constraining the distance between the polymer ends. The Fisher information for this system is equal to the variance of the tension at equilibrium.

$$I(L) = \beta^2 \left\langle \big(\langle \mathcal{T} \rangle - \mathcal{T} \big)^2 \right\rangle \qquad (9)$$

On the other hand, suppose control is exerted by applying constant tension to the ends of the polymer. The total energy is then a linear function of length and tension,

$E(x,\mathcal{T}) = U(x) - \mathcal{T}\, L(x)$, and the Fisher information is equal to the variation of the end-to-end polymer length at equilibrium.

$$I(\mathcal{T}) = \beta^2 \left\langle \left(\langle L \rangle - L \right)^2 \right\rangle \tag{10}$$

In each case the Fisher information has a simple physical interpretation in terms of the equilibrium fluctuations.

When the energy is a linear function of the control parameter, as in the last example, then the variance (and therefore the Fisher information) is equal to the negative second derivative of the free entropy [4, 5]. However, in general the relation between free energy and Fisher information contains an additional term.

$$\begin{aligned}
-\frac{\partial^2 \beta F}{\partial \lambda^2} &= -\frac{\partial}{\partial \lambda} \sum_x \exp\Big(\beta F(\lambda) - \beta E(x,\lambda)\Big) \frac{\partial \beta E}{\partial \lambda} \\
&= \beta^2 \left\langle \frac{\partial E}{\partial \lambda} \right\rangle^2 - \beta^2 \left\langle \left(\frac{\partial E}{\partial \lambda} \right)^2 \right\rangle - \beta \left\langle \frac{\partial^2 E}{\partial \lambda^2} \right\rangle \\
&= I(\lambda) - \beta \left\langle \frac{\partial^2 E}{\partial \lambda^2} \right\rangle
\end{aligned} \tag{11}$$

The last term will be zero if the control parameter is intensive (e.g. the pressure), and will be very small and entirely unimportant for an extensive parameter (e.g. volume) in the macroscopic thermodynamic limit. In these cases the Fisher information can be calculated given the free energy as a function of the control parameter.

When we have many control parameters $\lambda \equiv \{\lambda^1, \lambda^2, \dots, \lambda^N\}$ we can construct a Fisher information matrix $\mathcal{I}_{ij}$ (Also commonly g_{ij})

$$I_{ij} = \sum_x p(x|\lambda) \left(\frac{\partial \ln p(x|\lambda)}{\partial \lambda^i} \right) \left(\frac{\partial \ln p(x|\lambda)}{\partial \lambda^j} \right) \tag{12}$$

If we plug the canonical ensemble (3) into the definition of the Fisher information matrix (12), then we find that the Fisher matrix is a covariation matrix of fluctuations around equilibrium.

$$I_{ij} = \beta^2 \left\langle \left(\left\langle \frac{\partial E}{\partial \lambda^i} \right\rangle - \frac{\partial E}{\partial \lambda^i} \right) \left(\left\langle \frac{\partial E}{\partial \lambda^j} \right\rangle - \frac{\partial E}{\partial \lambda^j} \right) \right\rangle \tag{13}$$

Fisher information is central to the field of 'information geometry'. Since the Fisher information matrix I_{ij} is essentially a covariance matrix, it follows that the matrix is symmetric and positive semi-definite (all the eigenvalues are positive), and can therefore can be used a metric tensor to define a notion of distance between points in the space of parameters. This equips this manifold of parameters with the structure of a Riemannian metric [6, 7]. The length of a curve through parameter space, measured using this Fisher metric (also known as the Rao metric or entropy differential metric) is

$$\mathcal{L} = \int_0^1 \sqrt{\sum_{ij} \frac{d\lambda^i}{ds} \mathcal{I}_{ij} \frac{d\lambda^j}{ds}}\, ds \tag{14}$$

Recall that a metric provides a measure of 'distance' between points. It is a real function $f(a,b)$ such that (1) distances are non-negative, $f(a,b) \geqslant 0$ with equality if, and only if, $a = b$, (2) symmetric, $f(a,b) = f(b,a)$ and (3) it is generally shorter to go directly from point a to c than to go by way of b, $f(a,b) + f(b,c) \geqslant f(a,c)$ (The triangle inequality). Moreover, a Riemannian metric is a length space; there is always a point b 'between' a and c such that the equality $f(a,b) + f(b,c) = f(a,c)$ holds.

We can also define a corresponding divergence, or entropic action,

$$\mathcal{J} = \int_0^1 \sum_{ij} \frac{d\lambda^i}{ds} \mathcal{I}_{ij} \frac{d\lambda^j}{ds}\, ds \tag{15}$$

which is related to the length (14) by the inequality,

$$\mathcal{J} \geqslant \mathcal{L}^2 \tag{16}$$

which can be derived as a consequence of the Cauchy-Schwarz inequality $\int_0^\tau f^2 dt \int_0^\tau g^2 dt \geqslant \left[\int_0^\tau fg\, dt\right]^2$ with $g(t) = 1$. The value of the divergence depends on the parametrization. The minimum value $\mathcal{L}^2$ is attained only when the integrand is a constant along the path.

This Fisher information metric has been extensively applied to thermodynamic systems under the term "thermodynamic length" [8–11], although the connection to Fisher information has not been widely appreciated [4, 5]. The divergence measures the number of natural fluctuations along a path, and thermodynamic length is the cumulative root-mean-square deviations measured along the path.

Another application of Fisher information is the Cramér-Rao inequality [1]. Suppose that we have a function of state $\hat{\lambda}(x)$ that is an unbiased estimator for the parameter λ, i.e.

$$\int p(x|\lambda) \hat{\lambda}(x) dx = \lambda\,, \tag{17}$$

then the Cramér-Rao inequality states that the variance of the estimate is greater or equal to the inverse Fisher information,

$$\langle (\hat{\lambda} - \lambda)^2 \rangle \geqslant \frac{1}{I(\lambda)}\,. \tag{18}$$

For a system in thermodynamic equilibrium, the Cramér-Rao inequality can be interpreted as a thermodynamic uncertainty relation [12] (in rough analogy to the quantum uncertainty relations). For instance, a single measurement of a system determines the instantaneous energy, but this is not sufficient to infer the temperature with certainty. The Cramér-Rao inequality then relates the standard deviation of the energy fluctuations ΔE to the minimum uncertainty of the temperature measurement $\Delta\beta = \langle (\hat{\lambda} - \lambda)^2 \rangle^{1/2}$,

$$\Delta E\, \Delta\beta \geqslant 1\,. \tag{19}$$

Appendix: Miscellaneous Mathematics

It is useful to recall the behavior of logarithms under differentiation,

$$\frac{\partial \ln p(x|\lambda)}{\partial \lambda} = \frac{1}{p(x|\lambda)} \frac{\partial p(x|\lambda)}{\partial \lambda}$$

and that integration and differentiation can be interchanged, provided that certain mild technical conditions are met. See "Leibniz integral rule".

$$\frac{\partial}{\partial \lambda} \int p(x|\lambda)\, dx = \int \frac{\partial}{\partial \lambda} p(x|\lambda)\, dx$$

Since probability distributions are normalized,

$$\int p(x|\lambda)\, dx = 1$$

it follows that

$$\int \frac{\partial^n}{\partial \lambda^n} p(x|\lambda)\, dx = \frac{\partial^n}{\partial \lambda^n} \int p(x|\lambda)\, dx = \frac{\partial^n}{\partial \lambda^n} 1 = 0\ .$$

Consequently, the mean of the "score function" $\partial \ln p/\partial\lambda$ is zero.

$$\left\langle \frac{\partial}{\partial \lambda} \ln p(x|\lambda) \right\rangle = \int \frac{\partial}{\partial \lambda} p(x|\lambda)\, dx = 0$$

It follows that the Fisher information is the variance of the score.

Again, if we can interchange differentiation and integration, the Fisher information is also equal to the negative mean of the score's second derivative:

$$\begin{aligned}
-\left\langle \frac{\partial^2}{\partial \lambda^2} \ln p(x|\lambda) \right\rangle &= -\int p(x|\lambda) \frac{\partial}{\partial \lambda} \left[\frac{1}{p(x|\lambda)} \frac{\partial p(x|\lambda)}{\partial \lambda} \right] dx \\
&= \int \frac{1}{p(x|\lambda)} \left[\frac{\partial p(x|\lambda)}{\partial \lambda} \right]^2 dx - \int \frac{\partial^2 p(x|\lambda)}{\partial \lambda^2} dx \\
&= \int p(x|\lambda) \left[\frac{\partial \ln p(x|\lambda)}{\partial \lambda} \right]^2 dx - 0 \\
&= I(\lambda)
\end{aligned}$$

Acknowledgments: Thanks to David Sivak, Lawrence Berkeley National Laboratory (Numerous comments, observations and corrections); and Neri Merhav, Technion - Israel Institute of Technology (For pointing our the connection between the Cramér-Rao inequality and thermodynamic uncertainty relations).

Version History v4 (2011-12-17) Minor typo and grammar corrections. Added Fisher information of inverse temperature and relation between Cramér-Rao inequality thermodynamic uncertainty relations; v4 (2012-01-05) Fix several sign errors and other minor corrections (Kudos: David Sivak).

References

[1] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 2nd edition, 2006.

[2] J. W. Gibbs. *Elementary Principles in Statistical Mechanics*. Charles Scribner's Sons, New York, 1902.

[3] D. Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, Oxford, 1987.

[4] D. Brody and N. Rivier. Geometrical aspects of statistical mechanics. *Phys. Rev. E*, 51(2):1006–1011, 1995. doi:10.1103/PhysRevE.51.1006.

[5] G. E. Crooks. Measuring thermodynamic length. *Phys. Rev. Lett.*, 99:100602 (4), 2007. doi:10.1103/PhysRevLett.99.100602.

[6] C. R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81–91, 1945.

[7] J. Burbea and C. R. Rao. Entropy differential metric, distance and divergence measures in probability spaces - a unified approach. *J. Multivariate Anal.*, 12(4):575–596, 1982. doi:10.1016/0047-259X(82)90065-3.

[8] F. Weinhold. Metric geometry of equilibrium thermodynamics. *J. Chem. Phys.*, 63(6):2479–2483, 1975. doi:10.1063/1.431689.

[9] G. Ruppeiner. Thermodynamics: A riemannian geometric model. *Phys. Rev. A*, 20(4):1608–1613, 1979. doi:10.1103/PhysRevA.20.1608.

[10] F. Schlögl. Thermodynamic metric and stochastic measures. *Z. Phys. B*, 59(4):449–454, 1985. doi:10.1007/BF01328857.

[11] P. Salamon, J. Nulton, and E. Ihrig. On the relation between entropy and energy versions of thermodynamic length. *J. Chem. Phys.*, 80(1):436–437, 1984. doi:10.1063/1.446467.

[12] R. Gilmore. Uncertainty relations of statistical mechanics. *Phys. Rev. A*, 31(5):3237–3239, 1985. doi:10.1103/PhysRevA.31.3237.