Notes on Fisher distance in statistics, thermodynamics, quantum informatics

Lajos Diósi

Wigner Center, Budapest

3 Nov 2019, Jerusalem

Supported by
Foundational Questions Institute & Fetzer Franklin Fund National Research Development and Innovation Office of Hungary

Grant number K12435

1. Ronald Fisher
   - Fisher-information 1925
   - I never understood, but a week ago
   - Fisher's statistical distance
   - I never understood

2. From Fisher to Riemannian metric in thermodynamics
   - I noticed a Feldmann-Levine-Salamon paper 1986
   - and wrote "him" a letter
   - Encounter 8yy later, "he" is she 1994
   - What (the hell) is the thermodynamic length ℓ_{12}?

3. From Fisher to quantum informatics
   - Statistical distinguishability distance of quantum states
   - A best(?) answer to what (...) the statistical distance is

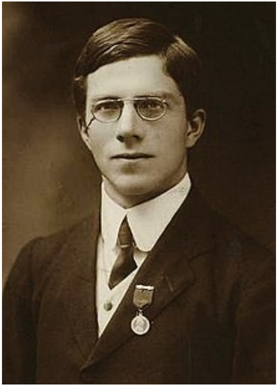
Sir Ronald Aylmer Fisher FRS[3] (17 February 1890 – 29 July 1962) was a British statistician and geneticist. For his work in statistics, he has been described as "a genius who almost single-handedly created the foundations for modern statistical science"[4] and "the single most important figure in 20th century statistics".[5] In genetics, his work used mathematics to combine Mendelian genetics and natural selection; this contributed to the revival of Darwinism in the early 20th-century revision of the theory of evolution known as the modern synthesis. For his contributions to biology, Fisher has been called "the greatest of Darwin's successors".[6]

From 1919 onward, he worked at the Rothamsted Experimental Station for 14 years;[7] there, he analysed its immense data from crop experiments since the 1840s, and developed the analysis of variance (ANOVA). He established his reputation there in the following years as a biostatistician.

He is known as one of the three principal founders of population genetics. He outlined Fisher's principle, the Fisherian runaway and sexy son hypothesis theories of sexual selection. His contributions to statistics include the maximum likelihood, fiducial inference, the derivation of various sampling distributions, founding principles of the design of experiments, and much more.

Fisher held strong views on race. Throughout his life, he was a prominent supporter of eugenics, an interest which led to his work on statistics and genetics.[8] Notably, he was a dissenting voice in the 1950 UNESCO statement *The Race Question*, insisting on racial differences.[9]

Sir
Ronald Fisher
FRS



Fisher in 1913

Born	Ronald Aylmer Fisher 17 February 1890 East Finchley, London, England, United Kingdom
Died	29 July 1962 (aged 72)

Fisher-information 1925

Fisher: Maximum Likelihood Method of parameter estimation
Probability $p_k(\theta)$, sample $\bar{k}$, unknown continuous parameter θ
MaxLik estimate $\theta^\star$: maximize $p_{\bar{k}}(\theta)$ $\Leftrightarrow$ maximize $L_k(\theta)$=ln$p_{\bar{k}}(\theta)$:

$$\left.\frac{\partial L_{\bar{k}}(\theta)}{\partial\theta}\right|_{\theta=\theta^\star}=0$$

Confidence interval $\Delta\theta$ of the estimate satisfies

$$-\left.\frac{\partial^2 L_{\bar{k}}(\theta)}{\partial\theta^2}\right|_{\theta=\theta^\star}(\Delta\theta)^2=1$$

Expectation value

$$-\sum_k p_k(\theta)\frac{\partial^2 L_k(\theta)}{\partial\theta^2}$$

is called Fisher-information.

I never understood, but a week ago

Fisher-information 1925

$$F(\theta) = -\sum_k p_k(\theta) \frac{\partial^2 L_k(\theta)}{\partial\theta^2} \equiv \sum_k \frac{1}{p_k(\theta)} \left(\frac{\partial p_k(\theta)}{\partial\theta}\right)^2$$

Shannon-information 1948:

$$S(\theta) = -\sum_k p_k(\theta) L_k(\theta) = -\sum_k p_k(\theta) \ln p_k(\theta)$$

I never understood why we keep calling $F(\theta)$ information?
Finally (a week ago) I looked into Fisher 1925:
He called it information, we follow him, but we never tell our students that his information has little to do with Shannon's later (and standard) concept of information.

the amount of information provided by a combination of two or more independent observations is thus merely the sum of the amounts of information in each piece separately.

It is a common case for a sample of n observations to be distributed into a finite number of classes, the numbers "expected" in each class being functions of one or more unknown parameters, if p is the probability of an observation falling into any one class, the amount of information in the sample is

$$S\left\{\frac{1}{m}\left(\frac{\partial m}{\partial \theta}\right)^2\right\}$$

where $m = np$, is the expectation in any one class. The variance of an efficient statistic derived from a large sample may, of course, be calculated from this expression.

7. *Efficiency of the maximum likelihood solution.*

We shall now prove that when an efficient statistic as defined above exists one may be found by the method of maximum likelihood.

If f stand for the probability that any particular type of observation should occur, and ϕ for the probability that any particular type of sample should occur, then

$$\log \phi = C + S(\log f)$$

when C is a constant which does not involve the parameters, the summation extending over all observations.

Fisher's statistical distance

Fisher-information is a metric of distiguishability rather than information:

$$(\Delta\ell)^2 = F(\theta)(\Delta\theta)^2 = \sum_k \frac{1}{p_k(\theta)}\left(\frac{\partial p_k(\theta)}{\partial\theta}\right)^2(\Delta\theta)^2.$$

$\Delta\ell$: Fisher statistical distance, reparametrization invariant.
$p_k(\theta)$ and $p_k(\theta+\Delta\theta)$ are statistically "well" distinguishable if

$$\Delta\ell \rangle 1.$$

Cramer-Rao bound 1945-46:
Given $p_k(\theta)$, the variance $\Delta\theta$ of any (unbiassed) estimate is bounded by $F(\theta)(\Delta\theta)^2 \geq 1$. Attainable if statistics $\to\infty$.

$\Rightarrow$ Definition: neighbouring $p_k(\theta)$ and $p_k(\theta+\Delta\theta)$ are distinguishable if their distance

$$\Delta\ell = 1.$$

I never understood ...

Fisher statistical distance

$$\Delta\ell = \sqrt{F(\theta)}\Delta\theta$$

is exact distinguishability measure (due to Cramer-Rao bound) between neighbouring $p_k(\theta)$'s.

What about distant ones $p_k(\theta_1)$ and $p_k(\theta_2)$?
Their Fisher-distance:

$$\ell_{12} = \int_{\theta_1}^{\theta^2} \sqrt{F(\theta)} d\theta$$

What (the hell) is the meaning of ℓ_{12}?

From Fisher to Riemann metric in thermodynamics

Weinhold, Ruppeiner, Salamon et al, D. et al. 1976-

Fisher statistical distance ℓ, multiparameter case, $\theta=\{\theta^a\}$:

$$(d\ell)^2 = g_{ab}(\theta)d\theta^a d\theta^b$$

$$g_{ab}(\theta)=\sum_k \frac{1}{p_k(\theta)}\frac{\partial p_k(\theta)}{\partial\theta^a}\frac{\partial p_k(\theta)}{\partial\theta^b}$$

Riemannian metric (of distinguishablity) on space $\{\theta\}$ of $p_k(\theta)$'s.
Apply it to Gibbs-distributions $p_E(\theta)$ of the extensives $\{E_a\}$, parametrized by the intensives $\{\theta^a\}$, taking $V\to\infty$:

$$g_{ab}(\theta)=\frac{\partial^2\phi(\theta)}{\partial\theta^a\partial\theta^b}$$

$$\phi(\theta_1,\theta_2,\dots) \;=\; \text{thermodynamic potential}$$

Spread (in Fisher metric) of thermodynamic fluctuations $\delta\theta^a$ is 1:

$$\langle(\delta\ell)^2\rangle = g_{ab}(\theta)\langle\delta\theta^a\delta\theta^b\rangle = \frac{\partial^2\phi(\theta)}{\partial\theta^a\partial\theta^b}\langle\delta\theta^a\delta\theta^b\rangle = 1.$$

ℓ: <u>thermodynamic length</u>, $\min\ell$: <u>thermodynamic distance</u>

I noticed a Feldmann-Levine-Salamon paper 1986

Journal of Statistical Physics, Vol. 42, Nos. 5/6, 1986

A Geometrical Measure for Entropy Changes

Tova Feldmann,[1] R. D. Levine,[1] and Peter Salamon[2]

Received September 4, 1984; revision received September 6, 1985

The geometrical approach to statistical mechanics is used to discuss changes in entropy upon sequential displacements of the state of the system. An interpretation of the angle between two states in terms of entropy differences is thereby provided. A particular result of note is that any state can be resolved into a state of maximal entropy (both states having the same expectation values for the constraints) and an orthogonal component. A cosine law for the general case is also derived.

HUNGARIAN ACADEMY OF SCIENCES
CENTRAL RESEARCH INSTITUTE FOR PHYSICS
LETTERS: H—1525 BUDAPEST 114. P.O.B. 49. CABLE: MTA KFKI TELEX: 22-47-22 PHONE: 166-540

Prof. Tova Feldman
Dept. of Phys. Chem.
The Hebrew University
Jerusalem 91904

1986. szeptember 18

Tisztelt és Kedves Kolléga!

Nagy örömmel olvastam, hogy – munkatársaival – a termodinamika Weinhold-Ruppeiner-... típusú geometriai elméletével foglalkozik. Szomorú vagyok viszont azért, mert erről csak a budapesti konferencia poszterje után szereztem tudomást Lukács Béla kollégámtól, aki jelen volt de Önnel /Veled/ már nem tudott találkozni. Én idén márciusban a Technionon voltam vendég és egy napra a Héber Egyetemet is meglátogattam /prof. Amin tanszékét/. Nagyon-nagyon sajnálom, hogy nem tudtunk egymásról, mert nekünk is több cikkünk jelent meg a Riemann-metricizációról. Ezeket most mellékelem, ha nem veszi tolakodásnak.

A Weinhold-Ruppeiner távolság és a Wootters távolság kapcsolatára mi is utaltunk, örülök hogy JSP 42, 1127 /1986/ cikkükben ez bővebben is ki van fejtve. Én elsősorban kvantummechanikai méréselmélettel foglalkozom és annak idején

Nagy örömmel olvastam, hogy – munkatársaival – a termodinamika Weinhold-Ruppeiner-... típusú geometriai elméletével foglalkozik. Szomorú vagyok viszont azért, mert erről csak a budapesti konferencia poszterje után szereztem tudomást Lukács Béla kollégámtól, aki jelen volt de Önnel /Veled/ már nem tudott találkozni. Én idén márciusban a Technionon voltam vendég és egy napra a Héber Egyetemet is meglátogattam /prof. Amin tanszékét/. Nagyon-nagyon sajnálom, hogy nem tudtunk egymásról, mert nekünk is több cikkünk jelent meg a Riemann-metricizációról. Ezeket most mellékelem, ha nem veszi tolakodásnak.

A Weinhold-Ruppeiner távolság és a Wootters távolság kapcsolatára mi is utaltunk, örülök hogy JSP 42, 1127 /1986/ cikkükben ez bővebben is ki van fejtve. Én elsősorban kvantummechanikai méréselmélettel foglalkozom és annak idején éppen a Wootters féle statisztikus távolság alkalmazásait keresve, a Gibbs eloszlásokon keresztül éppen a Ruppeiner távolsághoz jutottam.

Boldog lennék, ha – legalább irásban – kapcsolatba kerülnénk egymással. Kérem írja meg, ha angolul könnyebb.

Végezetül boldog ÚjÉvet kívánok!

Diósi Lajos

Diósi Lajos

לשנה טובה תכתבו

Encounter 8yy later, “he” is she 1994

Conf. on Modern Developments in Thermodynamics
(1994 Oct 2-7, Irsee, Germany)

Dinner time, talking to Peter Salamon in English, I don’t remember why not in Hungarian, but English did matter, since in Hungarian there’re no genders:

P: Didn’t you talk to Tova?

I: Does he attend the conference?

P: Yes, but Tova is she, not he!

I: Really? Are you sure Tova is femail in Hebrew?

P: Be sure, it is.

What (the hell) is the thermodynamic length ℓ_{12}?

Thermodynamic distance $\Delta\ell = \sqrt{g_{ab}(\theta)\Delta\theta^a\Delta\theta^b}$ (with $g_{ab} = \partial_a\partial_b\phi$) is exact distinguishability measure between neighbouring equilibrium thermodynamic states θ and $\theta + \Delta\theta$, against their fluctuations $\delta\theta$.

What about distant states θ_1 and θ_2?
Thermodynamic length between them (parameter invariant):

$$\ell_{12} = \int_P \sqrt{g_{ab}(\theta)d\theta^a d\theta^b}$$

Attempts to interpret ℓ_{12}:
Bring state from θ_1 to θ_2 in finite time, under very specific conditions of Salamon, or of D. et al! For entropy production $S(prod)$ we got:

$$\min S(prod) = \text{const} \times (\min \ell_{12})^2$$

Under general conditions: we don't know what (...) ℓ_{12} is good for!

From Fisher to quantum informatics

Fisher-distance on the full class of p_k's becomes simple:

$$(d\ell)^2 = \sum_k \frac{(dp_k)^2}{p_k} = 4\sum (dx_k)^2$$

$$x_k = \sqrt{p_k}, \quad \sum_k x_k^2 = 1$$

Euclidean metric on the unit hypersphere! Closed form(!) of distance:

$2\times$ spherical angle

$$\ell_{12} = 2\arccos\Big(\sum_k x_{1k}x_{2k}\Big) = 2\arccos\Big(\sum_k \sqrt{p_{1k}p_{2k}}\Big) \quad \epsilon[0,\pi]$$

With one eye on quantum, represent p_k's by diagonal density matrices

$$\hat\rho = \mathrm{diag}[p_1, p_2, \dots, p_k, \dots].$$

$$\ell_{12} = 2\arccos\left(\mathrm{tr}(\sqrt{\hat\rho_1\hat\rho_2}\right) \quad \epsilon[0,\pi]$$

If quantum theory answers me what (...) Fisher-distance is good for?

Statistical distinguishability distance of quantum states

Bures, Uhlmann, Braunstein &Caves, ... 1970's-

Full Fisher-story has been extended for quantum mechanics

- quantum Fisher-"information", quantum Cramer-Rao bound
- statistical length in space of density matrices $\hat{\rho}$: Bures-distance
- measure of undistingushability: fidelity $\mathcal{F}$

$$\mathcal{F}(|\psi_1\rangle, |\psi_2\rangle) = |\langle\psi_1|\psi_2\rangle|^2 \quad \epsilon[0,1]$$

For mixed states:

$$\mathcal{F}(\hat{\rho}_1, \hat{\rho}_2) = \mathrm{tr}\sqrt{\sqrt{\hat{\rho}_1}\hat{\rho}_2\sqrt{\hat{\rho}_1}} \quad \epsilon[0,1]$$

Uhlmann: If $\hat{\rho}_1, \hat{\rho}_2$ on a subsystem are reduced states of of the pure states $|\psi_1\rangle, |\psi_2\rangle$ then

$$\mathcal{F}(\hat{\rho}_1, \hat{\rho}_2) = \min \mathcal{F}(|\psi_1\rangle, |\psi_2\rangle)$$

where minimization is over all possible $|\psi_1\rangle, |\psi_2\rangle$, called purifications, of $\hat{\rho}_1, \hat{\rho}_2$ resp.

Quantum informatic answer: what statistical distance is good for?

Fisher statistical distance between $\{p_{1k}\}$ and $\{p_{2k}\}$

$$\ell_{12} = 2\arccos\left(\mathrm{tr}(\sqrt{\hat\rho_1\hat\rho_2}\right)$$

where $\hat\rho_1 = \mathrm{diag}[\{p_{1k}\}], \hat\rho_2 = \mathrm{diag}[\{p_{2k}\}]$.
Quantum informatic fidelity of $\hat\rho_1, \hat\rho_2$:

$$\mathcal{F}(\hat\rho_1,\hat\rho_2) = \mathrm{tr}\sqrt{\sqrt{\hat\rho_1}\hat\rho_2\sqrt{\hat\rho_1}} = \mathrm{tr}\sqrt{\hat\rho_1\hat\rho_2}$$

Hence $\cos(\ell_{12}/2) = \mathcal{F}(\hat\rho_1,\hat\rho_2)$. Uhlmann: $\cos(\ell_{12}/2) = \min|\langle\psi_1|\psi_2\rangle|^2$.
The best interpretation of the classical statistical length comes from quantum informatics:

A best(?) answer to what (...) the statistical distance is

If everything is quantum, and classical statistics (e.g. our two probability distributions p_1 and p_2) come from famous intrinsic randomness of quantum systems, and if we have access to the full quantum system, then p_1 and p_2 are at least as distinguishable than any two statevectors satisfying

$$\cos(\ell_{12}/2) = |\langle\psi_1|\psi_2\rangle|^2.$$

Interpretation of a classical concept invokes quantum concepts. Strange!

Neither statistics, nor thermodynamics or quantum informatics have answered me: What (...) is statistical lenght ℓ_{12} is good for in its full generality?

After my 35yy of attempts, I'm still unhappy.

But, on the other hand, and above all:

I'm glad to celebrate

Tova's 90th birthday

and

our 25 years in friendship and collaboration