Rényi Relative Entropy from Homogeneous Kullback-Leibler Divergence Lagrangian

Goffredo Chirco[0000−0001−9538−4956]

Istituto Nazionale di Fisica Nucleare - Sezione di Napoli, Complesso Universitario di Monte S. Angelo ed. 6, Via Cintia, 80126 Napoli, Italia.
goffredo.chirco@na.infn.it

Abstract. We study the *homogeneous* extension of the Kullback-Leibler divergence associated to a *covariant* variational problem on the statistical bundle. We assume a finite sample space. We show how such a divergence can be interpreted as a Finsler metric on an extended statistical bundle, where the time and the time score are understood as extra random functions defining the model. We find a relation between the homogeneous generalisation of the Kullback-Leibler divergence and the Rényi relative entropy, the Rényi parameter being associated to the time-reparametrization lapse of the model. We investigate such intriguing relation with an eye to applications in physics and quantum information theory.

Keywords: Kullback-Leibler Divergence · General Covariance · Rényi Relative Entropy · Non-parametric Information Geometry · Statistical Bundle.

Information Geometry on the Statistical Bundle

The probability simplex on a finite sample space Ω, with $\#\Omega = N$, is denoted by $\Delta(\Omega)$, and $\Delta^\circ(\Omega)$ its interior. The uniform probability function is μ, $\mu(x) = \frac{1}{N}$, $x \in \Omega$. We regard $\Delta^\circ(\Omega)$ as the *maximal exponential family* $\mathcal{E}(\mu)$, in the sense that each strictly positive density q can be written as $q \propto \mathrm{e}^v$, where v is defined up to a constant. The expected value of v with respect to the density q is $\mathbb{E}_q[v]$.

A non metric, non-parametric presentation of Information Geometry (IG) [1] is realised via a *joint* geometrical structure given by the probability simplex together with the set of q-integrable functions $v \in L(q)$: the couple (q, v) forming a *statistical vector bundle*

$$S\mathcal{E}(\mu) = \left\{ (q,v) \,\middle|\, q \in \mathcal{E}(\mu), v \in L^2_0(q) \right\} \ , \tag{1}$$

with base $\Delta^\circ(\Omega)$. For each $q \in \Delta^\circ(\Omega)$, $L^2(q)$ is the vector space of real functions of Ω endowed with the inner product $\langle u, v\rangle_q = \mathbb{E}_q[u\,v]$, and it holds $L^2(q) = \mathbb{R} \oplus L^2_0(p)$ [15, 17]. A (differential) geometry for the statistical bundle $S\mathcal{E}(\mu)$ is naturally provided by an *exponential atlas* of charts given for each $p \in \mathcal{E}(\mu)$ by

$$s_p \colon S\Delta^\circ(\Omega) \ni (q,v) \mapsto \left(\log\frac{q}{p} - \mathbb{E}_p\left[\log\frac{q}{p}\right], {}^{\mathrm{e}}\mathbb{U}^p_q v \right) \in S_p\Delta^\circ(\Omega) \times S_p\Delta^\circ(\Omega)$$

where ${}^{\mathrm{e}}\mathbb{U}_q^p$ denotes the *exponential transport*, defined for each $p, q \in \Delta^\circ(\Omega)$ by

$$
{}^{\mathrm{e}}\mathbb{U}_q^p : S_q\Delta^\circ(\Omega) \ni v \mapsto v - \mathbb{E}_p[v] \in S_p\Delta^\circ(\Omega)
$$

As $s_p(p, v) = (0, v)$, we say that s_p is the chart *centered at* p.

We can write then

$$
q = \exp(v - K_p(v)) \cdot p = e_p(v) \ , \tag{2}
$$

and see the mapping $s_p(q) = e_p(v)^{-1}$ as a section of the bundle.

The *cumulant function* $K_p(v) = \mathbb{E}_p\left[\log \frac{p}{q}\right] = \mathrm{D}(p \,\|\, q)$ is the expression in chart of the Kullback-Leibler divergence $p \mapsto \mathrm{D}(p \,\|\, q)$. The other divergence $\mathrm{D}(q \,\|\, p) = \mathbb{E}_q\left[\log \frac{q}{p}\right] = \mathbb{E}_q[v] - K_p(v)$ is the convex conjugate of the cumulant in the chart centered at p.

With respect to any of the convex functions $K_p(v)$ the maximal exponential family is a Hessian manifold. The given exponential atlas then provides the statistical bundle with an *affine* geometry, with a dual covariant structure induced by the inner product on the fibers given by the duality pairing between $S_q\,\mathcal{E}(\mu)$ and its dual ${}^*S_q\,\mathcal{E}(\mu)$, and the associated dual affine transports [10].

From Divergences on $\mathcal{E}(\mu) \times \mathcal{E}(\mu)$ to Lagrangians on $S\mathcal{E}(\mu)$

A divergence is a smooth mapping $D\colon \mathcal{E}(\mu) \times \mathcal{E}(\mu) \to \mathbb{R}$, such that for all $q, r \in \mathcal{E}(\mu)$ it holds $D(q, r) \geq 0$ and $D(q, r) = 0$ iff $q = r$. Every divergence can be associated to a Lagrangian function on the statistical bundle via the *canonical mapping* [5]

$$
\mathcal{E}(\mu)^2 \ni (q, r) \mapsto (q, s_q(r)) = (q, w) \in S\mathcal{E}(\mu) \ , \tag{3}
$$

where $r = \mathrm{e}^{w - K_q(w)} \cdot q$, that is, $w = s_q(r) = e_r^{-1}(w)$.

Remark 1. Such a mapping appears to be an instance of a general integral relation between the $\mathcal{G} = \mathcal{E}(\mu) \times \mathcal{E}(\mu)$ intended as a *pair groupoid* and the associated Lie algebroid $\mathfrak{Lie}(\mathcal{G})$ corresponding to the tangent bundle $S\mathcal{E}(\mu)$, see e.g. [11].

The inverse mapping is the *retraction* mapping

$$
S\mathcal{E}(\mu) \ni (q, w) \mapsto (q, e_q(w)) = (q, r) \in \mathcal{E}(\mu)^2 \ . \tag{4}
$$

As the curve $t \mapsto e_q(tw)$ has null exponential acceleration [16], one could say that eq. (4) defines the *exponential* mapping of the exponential connection, while eq. (3) defines the so-called *logarithmic* mapping.

The expression in a chart centered at p of the mapping of eq. (4) is affine:

$$
\begin{gathered}
S_p\,\mathcal{E}(\mu) \times S_p\,\mathcal{E}(\mu) \to \mathcal{E}(\mu) \times \mathcal{E}(\mu) \to S\mathcal{E}(\mu) \to S_p\,\mathcal{E}(\mu) \times S_p\,\mathcal{E}(\mu) \\
(p, u, v) \mapsto (e_p(u), e_p(v)) \mapsto (e_p(u), s_{e_p(u)}(e_p(v))) \mapsto (p, u, (v - u)) \ .
\end{gathered}
$$

The correspondence above maps every divergence D into a *divergence Lagrangian*, and conversely,

$$L(q, w) = D(q, e_q(w)) \ , \quad D(q, r) = L(q, s_q(r)) \ . \tag{5}$$

Notice that, according to our assumptions on the divergence, the divergence Lagrangian defined in eq. (5) is non-negative and zero if, and only if, $w = 0$.

Kullback-Leibler Dissimilarity Functional

Let us consider, as a canonical example, a Lagrangian given by the Kullback-Leibler divergence $\mathrm{D}\left(q \,\|\, r\right)$, which, by means of the canonical mapping in (3), corresponds to the cumulant function $K_q(w)$ on $S_q\,\mathcal{E}\left(\mu\right)$. This is a case of high regularity as we assume the densities q and r positive and connected by an open exponential arc. The Hessian structure of the exponential manifold is reflected in the hyper-regularity of the cumulant Lagrangian.

Let $\mathbb{R} \ni t \mapsto q(t) \in \mathcal{E}\left(\mu\right)$ be a smooth curve on the exponential manifold, intended as a one-dimensional parametric model. In the exponential chart centered at p, the velocity of the curve $q(t)$ is computed as

$$\frac{d}{dt} s_p(q(t)) = \frac{d}{dt}\left(\log\frac{q(t)}{p} - \mathbb{E}_p\left[\log\frac{q(t)}{p}\right]\right) = \frac{\dot q(t)}{q(t)} - \mathbb{E}_p\left[\frac{\dot q(t)}{q(t)}\right] = \\ {}^{\mathrm{e}}\mathbb{U}^{p}_{q(t)} \frac{\dot q(t)}{q(t)} = {}^{\mathrm{e}}\mathbb{U}^{p}_{q(t)} \frac{d}{dt}\log q(t) \ . \tag{6}$$

where partial derivatives are defined in the trivialisations given by the affine charts. By expressing the tangent at each time t in the moving frame at $q(t)$ along the curve, we define the *velocity* of the curve as the *score function* of the one-dimensional parametric model (see e.g. [8, §4.2])

$$\overset{\star}{q}(t) = {}^{\mathrm{e}}\mathbb{U}^{q(t)}_{p} \frac{d}{dt} s_p(q(t)) = \dot u(t) - \mathbb{E}_{q(t)}\left[\dot u(t)\right] = \frac{d}{dt}\log q(t) = \frac{\dot q(t)}{q(t)} \ . \tag{7}$$

The mapping $q \mapsto (q, \overset{\star}{q})$ is a lift of the curve to the statistical bundle whose expression in the chart centered at p is $t \mapsto (u(t), \dot u(t))$.

We shall now understand the cumulant Lagrangian as a divergence between q and r at any time, which compares the two probabilities at times infinitesimally apart. This amounts to consider the two probabilities as two one-dimensional parametric models constrained via the canonical mapping defined with respect to the score velocity vector, namely $r(t) = e_{q(t)}(\overset{\star}{q}(t))$.

By summing these divergences in time, we define a *dissimilarity* functional as the integral of the Kullback-Leibler cumulant Lagrangian $L\colon S\mathcal{E}\left(\mu\right) \times \mathbb{R} \to \mathbb{R}$ along the model

$$\begin{aligned}(q, r) \mapsto A[q] &= \int L(q(t), \overset{\star}{q}(t), t)\ dt \qquad (8)\\ &= \int \mathrm{D}\left(q(t) \,\|\, e_{q(t)}(\overset{\star}{q}(t)), t\right)\ dt = \int K_{q(t)}(\overset{\star}{q}(t))\ dt \ .\end{aligned}$$

As shown in [16, 5], via $A[q]$, one can consistently define a *variational principle* on the statistical bundle, leading to a *non-parametric* expression of the Euler-Lagrange equations in the statistical bundle. Indeed, if q is an extremal of the action integral, one gets

$$\frac{D}{dt} \operatorname{grad}_{\mathrm{e}} L(q(t), \overset{\star}{q}(t), t) = \operatorname{grad} L(q(t), \overset{\star}{q}(t), t) \ , \tag{9}$$

where grad indicates the natural gradient at $q(t)$ on the manifold, $\operatorname{grad}_{\mathrm{e}}$ the natural fiber-derivative with respect to the score, and $\frac{D}{dt}$ the mixture covariant derivative, as defined in [5].

The variational problem on the statistical bundle provides a natural setting for *accelerated* optimization on the probability simplex, with the divergence working as a kinetic energy *regulariser* for the scores, leading to faster converging and more stable optimization algorithms (see e.g. [9]).

More generally, the cumulant Lagrangian in (8) can be used to formulate a generalised geodesic principle on the statistical bundle, in terms of a class of local non-symmetric, non-quadratic generalizations of the Riemannian metrics. To this aim, beside smoothness and convexity, we need in first place the divergence Lagrangian to be *positive homogenous* of the first order in the scores. This brings new structures into play.

Re-parametrization Invariance of the Dissimilarity Action

Homogeneous Lagrangians (more precisely, positively homogeneous of degree one) lead to actions that are invariant under time re-parametrizations. Consider the action in eq. (8) and introduce a formal time parameter τ, such that $t = f(\tau)$ and $q(t) \to q(f(\tau)) = q(\tau)$. This elevates the time integration variable t to the rank of an independent dynamical variable, with f an *arbitrary* function, for which we can assume $\dot{f}(\tau) > 0$.

As a consequence, the re-parametrised action reads

$$\begin{aligned} A[(q,t)] = \int d\tau \, L(q(\tau), \overset{\star}{q}(\tau)/\dot{f}(\tau))) \, \dot{f}(\tau) \\ = \int d\tau \; \mathrm{D}\left(q(\tau) \, \| \, e_q(\overset{\star}{q}(\tau)/\dot{f}(\tau))\right) \, \dot{f}(\tau) \ , \end{aligned}$$

where $dt \to df(\tau) = \frac{d}{d\tau} f \, d\tau = \dot{f} \, d\tau$.

The new Lagrangian

$$\tilde{L}(q, f, v, \dot{f}) = \mathrm{D}\left(q \, \| \, e_q(v/\dot{f})\right) \, \dot{f} = K_q(\overset{\star}{q}/\dot{f}) \, \dot{f} \tag{10}$$

is defined on the *extended* statistical bundle $\tilde{S}\mathcal{E}\left(\mu\right) = S\mathcal{E}\left(\mu\right) \times \mathbb{R}$, with base $\tilde{\mathcal{E}}\left(\mu\right) = \mathcal{E}\left(\mu\right) \times \mathbb{R}$.

In particular, $\tilde{L}(q, f, v, \dot{f})$ is *homogeneous* of degree one in the velocity $(v, \dot{f}) \in \tilde{S}_{(q,f)}\,\mathcal{E}\,(\mu)$, that is

$$\tilde{L}(q, f, \lambda v, \lambda \dot{f}) = \lambda \tilde{L}(q, f, v, \dot{f}) \ . \tag{11}$$

In terms of $\tilde{L}$, the dissimilarity action becomes covariant to time reparametrisation, providing a generalisation of the notion of *information length* (see e.g.[6]), which allows for a generalised geodesic principle on the (extended) statistical bundle.

We shall focus on the very notion of homogeneous cumulant function. Our interest in this sense is twofold. On the one hand, as already noticed in [12], the homogenised cumulant function $\tilde{L}(q, f, v, \dot{f}) = K_q(\mathring{q}/\dot{f})\,\dot{f}$ generalises the Hessian geometry of $\mathcal{E}\,(\mu)$ to a *Finsler* geomety for the extended exponential family $\tilde{\mathcal{E}}\,(\mu)$, with $\tilde{L}$ playing the role of the Finsler metric. On the other hand, as it will be shown in the last section, the reparametrization invariance *symmetry* property of the homogeneous cumulant can be used to newly motivate the definition of Rényi relative entropy, with the Rényi parameter appearing as associated to the time-reparametrization lapse of the model.

Finsler Structure on the Extended Statistical Bundle

A Finsler metric on a differentiable manifold M is a continuous non-negative function $F : TM \to [0, +\infty)$ defined on the tangent bundle, so that for each point $x \in M$, (see e.g. [7])

$F(v + w) \leq F(v) + F(w)$ for every v, w tangent to M at x (*subadditivity*)
$F(\lambda v) = \lambda F(v)\ \forall\, \lambda \geq 0$ (*positive homogeneity*)
$F(v) > 0$ unless $v = 0$ (*positive definiteness*)

In our setting, by definition, for each point $\tilde{q} = (q, f) \in \tilde{\mathcal{E}}\,(\mu)$, and for $\dot{f}(\tau) > 0$, the Lagrangian $\tilde{L}$ is positively homogeneous, continuous, and non-negative on the extended statistical bundle $\tilde{S}\mathcal{E}\,(\mu)$. By labelling $\tilde{v} = (v, \dot{f}) \in \tilde{S}_{(q,f)}\,\mathcal{E}\,(\mu)$, subadditivity requires

$$\tilde{L}(\tilde{q}, \tilde{v} + \tilde{w}) \leq \tilde{L}(\tilde{q}, \tilde{v}) + \tilde{L}(\tilde{q}, \tilde{w}) \tag{12}$$

for every $\tilde{v}, \tilde{w} \in \tilde{S}_{(q,f)}\,\mathcal{E}\,(\mu)$. By noticing that $\tilde{v}+\tilde{w} = (v, \dot{f})+(w, \dot{f}) = (v+w, 2\dot{f})$, the subadditivity of $\tilde{L}$ is easily proved via Cauchy-Schwartz inequality.

We have

$$
\begin{aligned}
\tilde{L}(\tilde{q}, \tilde{v} + \tilde{w}) &= \tilde{L}(q, f, v + w, 2\dot{f}) = 2\dot{f} \log \mathbb{E}_q \left[\exp\left(\frac{v + w}{2\dot{f}} \right) \right] \\
&= 2\dot{f} \log \mathbb{E}_q \left[\exp\left(\frac{v}{2\dot{f}} \right) \exp\left(\frac{w}{2\dot{f}} \right) \right] \\
&\le 2\dot{f} \log \left(\mathbb{E}_q \left[\left(\exp\left(\frac{v}{2\dot{f}} \right) \right)^2 \right]^{1/2} \mathbb{E}_q \left[\left(\exp\left(\frac{w}{2\dot{f}} \right) \right)^2 \right]^{1/2} \right) \\
&= \dot{f} \log \mathbb{E}_q \left[\exp\left(\frac{v}{v_f} \right) \right] + \dot{f} \log \mathbb{E}_q \left[\exp\left(\frac{w}{\dot{f}} \right) \right] \\
&= \tilde{L}(\tilde{q}, \tilde{v}) + \tilde{L}(\tilde{q}, \tilde{w}) \quad .
\end{aligned}
$$

Together with the positive homogeneity expressed in (11) and the positive definiteness inherited by the definition of the KL divergence, we get that for each f, the homogeneous cumulant function $\tilde{L}$ satisfies the properties of a Finsler metric.

Re-parametrization Invariance gives $(1/\dot{f})$-Rényi divergence

The homogeneous Lagrangian $(q, \overset{\star}{q}, f, \dot{f}) \mapsto \dot{f} K_q(\overset{\star}{q}/\dot{f})$ describes a family of scaled Kullback-Leibler divergence measure, $\dot{f}(\tau)\, \mathrm{D}\left(q(\tau) \,\|\, \tilde{r}(\tau)\right)$, between the probabilities $q(\tau)$ and $\tilde{r}(\tau)$ infinitesimally apart in (sample) space and time, such that $\tilde{r}(\tau)$ can be expressed via retraction (exponential) mapping $\tilde{r}(\tau) = e_{q(\tau)}(\overset{\star}{q}(\tau)/\dot{f}(\tau))$ at any time. The same expression can be easily rewritten in terms of unscaled distributions q and r, by virtue of the canonical (log) mapping $\overset{\star}{q}(t) = s_{q(t)}(r(t))$. Indeed, we have

$$
\begin{aligned}
\dot{f}\, K_q(\overset{\star}{q}/\dot{f}) &= \dot{f} \log \mathbb{E}_q \left[\exp\left(\frac{\overset{\star}{q}}{\dot{f}} \right) \right] = \dot{f} \log \mathbb{E}_q \left[\exp\left(\frac{1}{\dot{f}} \left(\log \frac{r}{q} - \mathbb{E}_q \left[\log \frac{r}{q} \right] \right) \right) \right] \\
&= \dot{f} \log \mathbb{E}_q \left[\left(\frac{r}{q} \right)^{\frac{1}{\dot{f}}} \right] + \mathrm{D}\left(q \,\|\, r\right) = \dot{f} \log \mathbb{E}_\mu \left[r^{\frac{1}{\dot{f}}} q^{1 - \frac{1}{\dot{f}}} \right] + \mathrm{D}\left(q \,\|\, r\right) \quad . \qquad (13)
\end{aligned}
$$

where we use the definition $K_p(v) = \mathbb{E}_p \left[\log \frac{p}{q} \right] = \mathrm{D}\left(p \,\|\, q\right)$ in the second line. We removed the dependence on time in (13) in order to ease the notation.

Therefore, we see that for $\dot{f} = 1/1 - \alpha$, the term $\dot{f} \log \mathbb{E}_\mu \left[r^{1/\dot{f}}\, q^{1 - 1/\dot{f}} \right]$ corresponds to the definition of the α-Rényi divergence [18] with a minus sign

$$
D_\alpha(q || r) = \frac{1}{\alpha - 1} \log \mathbb{E}_\mu \left[r^{1-\alpha}\, q^{\alpha} \right] \;,
$$

while the second term, giving the KL for the unscaled distributions, derives from the centering contribution in the canonical map.

Remark 2. The Rényi parameter can be put in direct relation with the *lapse* factor of the reparametrization symmetry. Since the α parameter in Rènyi's entropy is constant, in this sense relating the lapse $\dot{f}$ to α amounts to restrict to reparametrization function f which are *linear* in time.

Remark 3. The two terms together in expression eq. (13) define a family of free energies, typically espressed as $(1/KT)\, F_\alpha(q, q') = D_\alpha(q||q') - \log Z$ (see e.g. [3]), and the action in eq. (8) can be understood as an integrated free energy associated with the transition from $q(\tau)$ to $r(\tau)$ along a curve in the bundle. Similar structures appear both in physics and quantum information theory in the study of out-of-equilibrium systems, where they provide extra constraints on thermodynamic evolution, beyond ordinary Second Law [2].

Remark 4. While the use of the canonical mapping in our affine setting somehow naturally leads to the Rényi formula for the divergence, along with the *exponential mean* generalization of the entropy formula [13, 14], the interpretation of the relation of Rènyi index and time lapse induced by reparametrization symmetry is open [19]. In our approach, setting $\dot{f} > 0$ just fixes $\alpha < 1$. A detailed thermodynamic analysis of these expressions is necessary for a deeper understanding of the map between α and $\dot{f}$.

Remark 5. The proposed result is quite intriguing when considered together with the Finsler characterization of the statistical manifold induced by the homogenized divergence, and its relation with *contact geometry* on the projectivised tangent bundle of the Finsler manifold (see e.g. [4]).

Acknowledgements

The author would like to thank G. Pistone and the anonymous referees for the careful read of the manuscript and the interesting points raised.

Bibliography

[1] Amari, S., Nagaoka, H.: Methods of information geometry. American Mathematical Society (2000), translated from the 1993 Japanese original by Daishi Harada

[2] Bernamonti, A., Galli, F., Myers, R.C., Oppenheim, J.: Holographic second laws of black hole thermodynamics. Journal of High Energy Physics **2018**(7), 111 (2018). https://doi.org/10.1007/JHEP07(2018)111

[3] Brandão, F., Horodecki, M., Ng, N., Oppenheim, J., Wehner, S.: The second laws of quantum thermodynamics. Proceedings of the National Academy of Sciences **112**(11), 3275–3279 (2015). https://doi.org/10.1073/$pnas$.1411728112

[4] Chern, S.-S., Chen, W., Lam, W. K. : Lectures on differential geometry. Singapore ; River Edge, N.J.: World Scientific (1999)

[5] Chirco, G., Malagò, L., Pistone, G.: Lagrangian and hamiltonian mechanics for probabilities on the statistical manifold, arXiv:2009.09431[math.ST]

[6] Crooks, G.E.: Measuring thermodynamic length. Phys. Rev. Lett. **99**, 100602 (Sep 2007). https://doi.org/10.1103/$PhysRevLett$.99.100602

[7] D. Bao, S.-S. Chern, Z.S.: An Introduction to Riemann-Finsler Geometry, in Graduate Texts in Mathematics, Springer, New York, NY, springer-verlag new york, inc. (2000), https://doi.org/10.1007/$978 - 1 - 4612 - 1268 - 3$

[8] Efron, B., Hastie, T.: Computer age statistical inference, Institute of Mathematical Statistics (IMS) Monographs, vol. 5. Cambridge University Press, New York (2016), https://doi.org/10.1017/CBO9781316576533

[9] França, G., Jordan, M.I., Vidal, R.: On dissipative symplectic integration with applications to gradient-based optimization, arXiv:2004.06840[math.OC]

[10] Gibilisco, P., Pistone, G.: Connections on non-parametric statistical manifolds by Orlicz space geometry. IDAQP **1**(2), 325–347 (1998)

[11] Grabowska, K., Grabowski, J., Kuś, M., Marmo, G.: Lie groupoids in information geometry. Journal of Physics A: Mathematical and Theoretical **52**(50), 505202 (nov 2019). https://doi.org/10.1088/$1751 - 8121/ab542e$

[12] Ingarden, R.S.: Information geometry of thermodynamics. In: Višek, J.Á. (ed.) Transactions of the Tenth Prague Conference: Information Theory, Statistical Decision Functions, Random Processes held at Prague Volume A. pp. 421–428. Springer Netherlands, Dordrecht (1988). https://doi.org/10.1007/$978 - 94 - 009 - 3859 - 5_4 4$

[13] Kolmogorov, A.: Sur la notion de la moyenne. Atti R. Accad. Naz. Lincei **12** 388 (1930)

[14] Nagumo, M.: Über eine klasse der mittelwerte. Japanese journal of mathematics : transactions and abstracts **7**, 71–79 (1930)

[15] Pistone, G.: Nonparametric information geometry. In: Nielsen, F., Barbaresco, F. (eds.) Geometric science of information, Lecture Notes in Com-

put. Sci., vol. 8085, pp. 5–36. Springer, Heidelberg (2013), first International Conference, GSI 2013 Paris, France, August 28-30, 2013 Proceedings
[16] Pistone, G.: Lagrangian function on the finite state space statistical bundle. Entropy **20**(2), 139 (2018). https://doi.org/10.3390/e20020139
[17] Pistone, G.: Information geometry of the probability simplex. a short course. Nonlinear Phenomena in Complex Systems **23** (2), 221–242 (2020)
[18] Rényi, A.: On Measures of Entropy and Information, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, University of California Press, 547–561 (1961)
[19] Chirco, G.: *In preparation*