Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 January 2018 doi:10.20944/preprints201712.0191.v2

Peer-reviewed version available at *Entropy* **2018**, *20*, 139; doi:10.3390/e20020139

Article

Lagrangian Function on the Finite State Space Statistical Bundle

Giovanni Pistone [1]

[1] de Castro Statistics, Collegio Carlo Alberto, Piazza Vincenzo Arbarello 8, 10122 Torino, Italy; giovanni.pistone@carloalberto.org

Abstract: The statistical bundle is the set of couples (Q, W) of a probability density Q and a random variable W such that $\mathbb{E}_Q[W] = 0$. On a finite state space, we assume Q to be a probability density with respect to the uniform probability and give an affine atlas of charts such that the resulting manifold is a model for Information Geometry. Velocity and accelleration of a one-dimensional statistical model are computed in this set up. The Euler-Lagrange equations are derived from the Lagrange action integral. An example of Lagrangian using minus the entropy as potential energy is briefly discussed.

Keywords: Information Geometry; Statistical Bundle; Lagrangian function

1. Introduction

The set-up of classical Lagrangian Mechanics is a finite-dimensional Riemannian manifold. For example, see the monographs by V.I. Arnold [1, Ch. III-IV], R. Abraham and J.E. Mardsen [2, Ch. 3], J.E. Marsden and T.S. Ratiu [3, Ch. 7]. Classical Information geometry, as it was firstly defined in the monograph by S.-I. Amari and H. Nagaoka [4], views parametric statistical models as a manifold endowed with a dually-flat connection. In a recent paper, M. Leok and J. Zhang [5] have pointed out the natural relation between these two topics and have given a wide overview of the mathematical structures involved.

In the present paper, we take up the same research program with two further qualification. First, we assume a non-parametric approach by considering the full set of positive probability functions on a finite set, as it was done, for example, in our review paper [6]. The discussion is restricted here to a finite state space to avoid difficult technical problems. Second, we consider a specific expression of the tangent space of the statistical manifold, which is an Hilbert bundle that we call statistical bundle. Our aim is to emphasize the basic statistical intuition of the geometric quantities involved. Because of that, we choose to use systematically the language of non-parametric differential geometry as it is developed in S. Lang monograph [7].

We use here our version of Information Geometry, see the review paper [6]. Preliminary versions of this paper have been presented at the SigmaPhy2017 Conference held in Corfu, Greece, Jul. 10-14 2017 and at a seminar held at Collegio Carlo Alberto, Moncalieri, on Sep. 5, 2017. In these early versions we did not refer to Leok and Zhang work we where unaware of at that time.

In Sec. 2 we review the definition and properties of the statistical bundle, and of the affine atlas that endows it with both a manifold structure and a natural family of transports between the fibers. In Sec. 3 we develop the formalism of the tangent space of the statistical bundle and derive the expression of the velocity and the acceleration of a one-dimensional statistical model in the given affine atlas. The derivation of the Euler-Lagrange equations, together with a relevant example, is discussed in Sec. 4.

2. Statistical bundle

We consider a finite sample space Ω, with $\#\Omega = N$. The probability simplex is $\Delta(\Omega)$ and $\Delta^\circ(\Omega)$ is its interior. The uniform probability on Ω is denoted μ, $\mu(x) = \frac{1}{N}$, $x \in \Omega$. The *maximal exponential*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 January 2018 doi:10.20944/preprints201712.0191.v2

Peer-reviewed version available at *Entropy* **2018**, *20*, 139; doi:10.3390/e20020139

family $\mathcal{E}(\mu)$ is the set of all strictly positive probability densities of (Ω, μ). The expected value of $f\colon \Omega \to \mathbb{R}$ with respect to the density $P \in \mathcal{E}(\mu)$ is denoted $\mathbb{E}_P[f] = \mathbb{E}_\mu[fP] = \frac{1}{N}\sum_{x\in\Omega} f(x)P(x)$.

In [6,8,9] we made the case for the statistical bundle being the key structure of Information Geometry. The *statistical bundle* with base Ω is

$$S\,\mathcal{E}(\mu) = \left\{(Q,V)\middle|Q \in \mathcal{E}(\mu), \mathbb{E}_Q[V] = 0\right\} .$$

The statistical bundle is a semi-algebraic subset of $\mathbb{R}^{2N}$ i.e., it is defined by algebraic equations and strict inequalities. It is trivially a real manifold. At each $Q \in \mathcal{E}(\mu)$ the fiber $S_Q\,\mathcal{E}(\mu)$ is endowed with the scalar product

$$(V_1, V_2) \mapsto \langle V_1, V_2\rangle_Q = \mathbb{E}_Q[V_1 V_2] = \operatorname{Cov}_Q(V_1, V_2) .$$

We add to this structure a special affine atlas of charts in order to show a structure of affine manifold which is of interest in the statistical applications. The *exponential atlas* of the statistical manifold $S\,\mathcal{E}(\mu)$ is the collection of charts given for each $P \in \mathcal{E}(\mu)$ by

$$s_P\colon S\,\mathcal{E}(\mu) \ni (Q,V) \mapsto (s_P(Q), {}^{\mathrm{e}}\mathbb{U}_Q^P V) \in S_P\,\mathcal{E}(\mu) \times S_P\,\mathcal{E}(\mu) , \tag{1}$$

where (with a slight abuse of notation)

$$s_P(Q) = \log\frac{Q}{P} - \mathbb{E}_P\left[\log\frac{Q}{P}\right], \quad {}^{\mathrm{e}}\mathbb{U}_Q^P V = V - \mathbb{E}_P[V] . \tag{2}$$

As $s_P(P,V) = (0,V)$, we say that s_P is the chart *centered at* P. If $s_P(Q) = U$, it is easy to derive the exponential form of Q as a density with respect to P, namely $Q = \mathrm{e}^{U - \mathbb{E}_P\left[\log\frac{Q}{P}\right]} \cdot P$. As $\mathbb{E}_\mu[Q] = 1$, then $1 = \mathbb{E}_p\left[\mathrm{e}^{U - \mathbb{E}_P\left[\log\frac{P}{Q}\right]}\right] = \mathbb{E}_p\left[\mathrm{e}^U\right]\mathrm{e}^{-\mathbb{E}_P\left[\log\frac{P}{Q}\right]}$, so that the *cumulant function* K_P is defined on $S_P\,\mathcal{E}(\mu)$ by

$$K_p(U) = \log\mathbb{E}_P\left[\mathrm{e}^U\right] = \mathbb{E}_P\left[\log\frac{P}{Q}\right] = \mathrm{D}(P\,\|Q) ,$$

that is, $K_P(V)$ is the expression in the chart at P of Kullback-Leibler divergence of $Q \mapsto \mathrm{D}(P\,\|Q)$, and we can write

$$Q = \mathrm{e}^{U - K_P(U)} \cdot P = \mathrm{e}_P(U) .$$

The *patch centered at* P is

$$s_P^{-1} = e_P\colon (S_P\,\mathcal{E}(\mu))^2 \ni (U,W) \mapsto (\mathrm{e}_P(U), {}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)} W) \in S\,\mathcal{E}(\mu) .$$

In statistical terms, the random variable $\log(Q/P)$ is the relative point-wise information about Q relative to the reference P, while $s_P(Q)$ is the deviation from its mean value at P. The expression of the other divergence in the chart centered at P is

$$\mathrm{D}(Q\,\|P) = \mathbb{E}_Q\left[\log\frac{Q}{P}\right] = \mathbb{E}_Q[U - K_P(U)] = \mathbb{E}_Q[U] - K_P(U) .$$

The equation above shows that the two divergences are convex conjugate functions in the proper charts, see [10].

The transition maps of the exponential atlas in Eq.s (1) and (2) are

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 January 2018 doi:10.20944/preprints201712.0191.v2

Peer-reviewed version available at Entropy 2018, 20, 139; doi:10.3390/e20020139

$$s_{P_2} \circ e_{P_1}(U, W) = s_{P_2}\left(\mathrm{e}_{P_1}(U), {}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_1 P(U)} W\right) = s_{P_2}\left(\mathrm{e}^{U - K_{P_1}(U)} \cdot P_1, W - \mathbb{E}_{\mathrm{e}_{P_1}(U)}[W]\right) =$$
$$\left(U - K_{P_1}(U) + \log\frac{P_1}{P_2} - \mathbb{E}_{P_2}\left[U - K_{P_1}(U) + \log\frac{P_1}{P_2}\right], W - \mathbb{E}_{\mathrm{e}_{P_1}(U)}[W] - \mathbb{E}_{P_2}\left[W - \mathbb{E}_{\mathrm{e}_{P_1}(U)}[W]\right]\right) =$$
$$\left({}^{\mathrm{e}}\mathbb{U}_{P_1}^{P_2} U + s_{P_2}(P_1), {}^{\mathrm{e}}\mathbb{U}_{P_1}^{P_2} W\right),$$

so that the exponential atlas is indeed affine. Notice that the linear part is ${}^{\mathrm{e}}\mathbb{U}_{P_1}^{P_2}$.

3. The tangent space of the statistical bundle

Let us compute the expression of the velocity at time t of a smooth curve $t \mapsto \gamma(t) = (Q(t), W(t)) \in S\,\mathcal{E}(\mu)$ in the chart centered at P. The expression of the curve is

$$\gamma_P(t) = s_P(\gamma(t)) = \left(s_P(Q(t)), {}^{\mathrm{e}}\mathbb{U}_{Q(t)}^{P} W(t)\right),$$

and hence we have, by denoting the derivative in $\mathbb{R}^N$ by the dot,

$$\frac{d}{dt} s_P(Q(t)) = \frac{d}{dt}\left(\log\frac{Q(t)}{P} - \mathbb{E}_P\left[\log\frac{Q(t)}{P}\right]\right) = \frac{\dot{Q}(t)}{Q(t)} - \mathbb{E}_P\left[\frac{\dot{Q}(t)}{Q(t)}\right] = {}^{\mathrm{e}}\mathbb{U}_{Q(t)}^{P}\frac{\dot{Q}(t)}{Q(t)}, \tag{3}$$

and

$$\frac{d}{dt}\,{}^{\mathrm{e}}\mathbb{U}_{Q(t)}^{P} W(t) = \frac{d}{dt}\left(W(t) - \mathbb{E}_P[W(t)]\right) = \dot{W}(t) - \mathbb{E}_P\left[\dot{W}(t)\right] = {}^{\mathrm{e}}\mathbb{U}_{Q(t)}^{P}\left(\dot{W}(t) - \mathbb{E}_{Q(t)}\left[\dot{W}(t)\right]\right). \tag{4}$$

If we define the *velocity* of $t \mapsto Q(t) = \mathrm{e}^{U(t) - K_p(U(t))} \cdot P$ to be

$$\overset{\star}{Q}(t) = \frac{\dot{Q}(t)}{Q(t)} = \frac{d}{dt}\log Q(t) = \dot{U}(t) - dK_P(U(t))[\dot{U}(t)] \in S_{Q(t)}\,\mathcal{E}(\mu),$$

then $t \mapsto (Q(t), \overset{\star}{Q}(t))$ is a curve in the statistical bundle whose expression in the chart centered at P is $t \mapsto (U(t), \dot{U}(t))$. The velocity as defined above is nothing else as the *score function* of the one-dimensional statistical model see e.g., the textbook by B. Efron and T. Haste [11, §4.2]. The variance of the score function i.e. the squared norm of $\overset{\star}{Q}(t)$ in $S_{Q(t)}\,\mathcal{E}(\mu)$ is classically known as *Fisher information* at t.

We define the *second statistical bundle* to be

$$S^2\,\mathcal{E}(\mu) = \left\{(Q, W, X, Y) \middle| (Q, W) \in S\,\mathcal{E}(\mu), X, Y \in S_Q\,\mathcal{E}(\mu)\right\},$$

with charts

$$s_P(Q, V, X, Y) = \left(s_P(Q, V), {}^{\mathrm{e}}\mathbb{U}_Q^P X, {}^{\mathrm{e}}\mathbb{U}_Q^P Y\right),$$

we can identify the second bundle with the tangent space of the first bundle as follows.

For each curve $t \mapsto \gamma(t) = (Q(t), W(t))$ in the statistical bundle, define its *velocity at* t to be

$$\overset{\star}{\gamma}(t) = \left(Q(t), W(t), \overset{\star}{Q}(t), \dot{W}(t) - \mathbb{E}_{Q(t)}\left[\dot{W}(t)\right]\right),$$

because $t \mapsto \overset{\star}{\gamma}(t)$ is a curve in the second statistical bundle and that its expression in the chart at P has the last two components equal to the values given in Eq.s (3) and (4).

In particular, consider the a curve $t \mapsto \chi(t) = (Q(t), \overset{\star}{Q}(t))$. The velocity is

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 January 2018 doi:10.20944/preprints201712.0191.v2

Peer-reviewed version available at *Entropy* **2018**, *20*, 139; doi:10.3390/e20020139

$$\overset{\star}{\chi}(t) = \left(Q(t), \overset{\star}{Q}(t), \overset{\star}{Q}(t), \overset{\star\star}{Q}(t)\right) ,$$

where the *acceleration* $\overset{\star\star}{Q}(t)$ is

$$\overset{\star\star}{Q}(t) = \frac{d}{dt}\frac{\dot{Q}(t)}{Q(t)} - \mathbb{E}_{Q(t)}\left[\frac{d}{dt}\frac{\dot{Q}(t)}{Q(t)}\right] = \frac{\ddot{Q}(t)}{Q(t)} - \left(\overset{\star}{Q}(t)^2 - \mathbb{E}_{Q(t)}\left[\overset{\star}{Q}(t)^2\right]\right) \tag{5}$$

It should be noted that the acceleration has been defined without explicitly mentioning the relevant connection. In fact, the connection here is implicitly defined by the transports ${}^{\mathrm{e}}\mathbb{U}_P^Q$, which is unusual in Differential Geometry, but is quite natural from the probabilistic point of view, see P. Gibilisco and G. Pistone [12]. We shall see below that the non-parametric approach to Information Geometry allows to define a dual transport, hence a dual connection as it was in [4]. because of that. we could have defined other types of acceleration together with the one we have defined. Namely, we could consider an *exponential acceleration* ${}^{e}\mathrm{D}^2Q(t) = \overset{\star\star}{Q}(t)$, a *mixture acceleration* ${}^{m}\mathrm{D}^2Q(t) = \ddot{Q}(t)/Q(t)$, and a *Riemannian acceleration*

$${}^{0}\mathrm{D}^2Q(t) = \frac{1}{2}\left({}^{e}\mathrm{D}^2Q(t) + {}^{m}\mathrm{D}^2Q(t)\right) = \frac{\ddot{Q}(t)}{Q(t)} - \frac{1}{2}\left(\left(\frac{\dot{Q}(t)}{Q(t)}\right)^2 - \mathbb{E}_{Q(t)}\left[\left(\frac{\dot{Q}(t)}{Q(t)}\right)^2\right]\right) , \tag{6}$$

each acceleration being associated with a specific connection, see the review paper [6]. We do not further discuss the different second order geometries associated to the statistical bundle in this paper.

Example 1 (Boltzmann-Gibbs)**.** Let us compare the formalism we have introduced above with standard computations in Statistical Physics. The *Boltzmann-Gibbs distribution* gives to point $x \in \Omega$ the probability $\mathrm{e}^{-(1/\theta)H(x)}/Z(\theta)$, with $Z(\theta) = \sum_{x\in\Omega} \mathrm{e}^{-(1/\theta)H(x)}$ and $\theta > 0$, see Landau and Lifshits [13, Ch. 3]. As a curve in $\mathcal{E}(\mu)$, it is $Q(\theta) = N\mathrm{e}^{-(1/\theta)H}/Z(\theta)$ because of the reference to the uniform probability. The velocity defined above becomes in this case $\overset{\star}{Q}(\theta) = \theta^{-2}(H - \mathbb{E}_\theta[H])$, while the acceleration of Eq. (5) is $\overset{\star\star}{Q}(\theta) = -\theta^{-3}(H - \mathbb{E}_\theta[H])$. Notice that we have the equation $\theta\overset{\star\star}{Q}(\theta) + \overset{\star}{Q}(\theta) = 0$.

Following the original construction of Amari's Information Geometry [4], we have defined on the statistical bundle a manifold structure which is both an affine and a Riemannian manifold. The base manifold $\mathcal{E}(\mu)$ is actually an Hessian manifold with respect to any of the convex functions $K_p(U) = \log \mathbb{E}_p\left[\mathrm{e}^U\right]$, $U \in S_p\,\mathcal{E}(\mu)$, see [14]. Many computations are actually performed using the Hessian structure. The following equations are easily checked and frequently used

$$\mathbb{E}_{\mathrm{e}_P(U)}[H] = dK_P(U)[H] ; \tag{7}$$

$${}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)}H = H - dK_P(U)[H] ; \tag{8}$$

$$d^2K_P(U)[H_1, H_2] = \left\langle {}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)}H_1, {}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)}H_2 \right\rangle_{\mathrm{e}_P(U)} ; \tag{9}$$

$$d^3K_p(U)[H_1, H_2, H_3] = \mathbb{E}_{\mathrm{e}_P(U)}\left[{}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)}H_1 \cdot {}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)}H_2 \cdot {}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)}H_3\right] . \tag{10}$$

We have defined a centering operation that can be thought of as a *transport* among fibers,

$${}^{\mathrm{e}}\mathbb{U}_P^Q \colon S_p\,\mathcal{E}(\mu) \to S_q\,\mathcal{E}(\mu) ,$$

whose adjoint is ${}^{\mathrm{m}}\mathbb{U}_q^p V = \frac{q}{p}V$. In fact, is the adjoint of ${}^{\mathrm{e}}\mathbb{U}_p^q$,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 January 2018 doi:10.20944/preprints201712.0191.v2

Peer-reviewed version available at Entropy 2018, 20, 139; doi:10.3390/e20020139

$$\left\langle {}^{\mathrm{e}}\mathbb{U}_P^Q U, V \right\rangle_Q = \mathbb{E}_Q\left[(U - \mathbb{E}_Q[U])V\right] = \mathbb{E}_Q[UV] = \mathbb{E}_P\left[U\left(\frac{Q}{P}V\right)\right] = \left\langle U, {}^{\mathrm{m}}\mathbb{U}_Q^P V \right\rangle_P$$

Moreover, iff $U, V \in S_P\mathcal{E}(\mu)$, then

$$\left\langle {}^{\mathrm{e}}\mathbb{U}_P^Q U, {}^{\mathrm{m}}\mathbb{U}_P^Q V \right\rangle_Q = \left\langle {}^{\mathrm{e}}\mathbb{U}_Q^P\, {}^{\mathrm{e}}\mathbb{U}_P^Q U, V \right\rangle_P = \langle U, V\rangle_P .$$

Example 2 (Entropy flow)**.** This example is taken from [8]. In the scalar field $\mathcal{H}(Q) = -\mathbb{E}_Q[\log Q]$ there is no dependence on the fiber. If $t \mapsto Q(t) = \mathrm{e}^{V(t) - K_P(V(t))} \cdot P$ is a smooth curve in $\mathcal{E}(\mu)$ expressed in the chart centered at P, then we can write

$$\begin{aligned}\mathcal{H}(Q(t)) = -\mathbb{E}_{Q(t)}\left[V(t) - K_P(V(t)) + \log P\right] = \\ K_P(V(t)) - \mathbb{E}_{Q(t)}\left[V(t) + \log P + \mathcal{H}(P)\right] + \mathcal{H}(P) = \\ K_P(V(t)) - dK_P(V(t))[V(t) + \log P + \mathcal{H}(P)] + \mathcal{H}(P) ,\end{aligned} \tag{11}$$

where the argument of the last expectation belongs to the fiber $S_P\mathcal{E}(\mu)$ and we have expressed the expected value as a derivative by using Eq. (7).

Using again Eq. (7), and also Eq. (9) we compute the derivative of the entropy along the given curve as

$$\begin{aligned}\frac{d}{dt}\mathcal{H}(Q(t)) = \frac{d}{dt}K_P(V(t)) - \frac{d}{dt}dK_P(V(t))[V(t) + \log P + \mathcal{H}(P)] = \\ dK_P(V(t))[\dot{V}(t)] - d^2K_P(V(t))[V(t) + \log P + \mathcal{H}(P), \dot{V}(t)] - dK_P(V(t))[\dot{V}(t)] = \\ -\mathbb{E}_{Q(t)}\left[{}^{\mathrm{e}}\mathbb{U}_P^{Q(t)}(V(t) + \log P)\, {}^{\mathrm{e}}\mathbb{U}_P^{Q(t)}\dot{V}(t)\right] .\end{aligned}$$

We use now the equations

$$V(t) + \log P = \log Q(t) + K_P(V(t)) , \quad {}^{\mathrm{e}}\mathbb{U}_P^{Q(t)}\left(\log Q(t) + K_P(V(t))\right) = \log Q(t) + \mathcal{H}(Q(t)) ,$$

and ${}^{\mathrm{e}}\mathbb{U}_P^{Q(t)}\dot{V}(t) = \overset{\star}{Q}(t)$, to obtain

$$\frac{d}{dt}\mathcal{H}(Q(t)) = -\left\langle \log Q(t) + \mathcal{H}(Q(t)), \overset{\star}{Q}(t) \right\rangle_{Q(t)} .$$

We have identified the gradient of the entropy in the statistical bundle,

$$\operatorname{grad}\mathcal{H}(Q) = -(\log Q + \mathcal{H}(Q)) . \tag{12}$$

Notice that the previous computation could have been done using the exponential family $Q(t) = \mathrm{e}_P(tV)$. See in [8] the computation of the gradient flow.

In the next section, we extend the computation illustrated in the example above to scalar fields on the statistical bundle.

4. Lagrangian function

A *Lagrangian function* is a smooth scalar field on the statistical bundle

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 January 2018 doi:10.20944/preprints201712.0191.v2

Peer-reviewed version available at *Entropy* **2018**, *20*, 139; doi:10.3390/e20020139

$$L\colon S\,\mathcal{E}\,(\mu) \ni (Q,W) \mapsto L(Q,W) \in \mathbb{R}\,.$$

At each fixed density $Q \in \mathcal{E}\,(\mu)$, the partial mapping

$$S_Q\,\mathcal{E}\,(\mu) \ni W \mapsto L(Q,W) \tag{13}$$

is a defined on the vector space $S_q\,\mathcal{E}\,(\mu)$, hence we can use the ordinary derivative, which is called in this case *fiber derivative*,

$$d_2L(Q,W)[H_2] = \left.\frac{d}{dt}L(Q,W+tH_2)\right|_{t=0}, \quad H_2 \in S_Q\,\mathcal{E}\,(\mu)\ . \tag{14}$$

Example 3 (Running example I)**.** If

$$L(Q,W) = \frac{1}{2}\langle W,W\rangle_Q + \kappa\,\mathcal{H}\,(Q)\ ,\quad \kappa \geq o\,, \tag{15}$$

then $d_2L(Q,W)[H_2] = \langle W,H_2\rangle_Q$. The example is suggested by the form of the classical Lagrangian function in mechanics, where the first term is the kinetic energy and $-\kappa\,\mathcal{H}\,(Q)$ is the potential energy.

As the statistical bundle $S\,\mathcal{E}\,(\mu)$ is non-trivial, the computation of the partial derivative of the Lagrangian with respect to the first variable requires some care. We want to compute the expression of the total derivative in a chart of the affine atlas defined in Eq.s (1) and (1).

Let $t \mapsto \gamma(t) = (Q(t),W(t))$ a smooth curve in the statistical bundle. In the chart centered at P we have

$$Q(t) = \mathrm{e}^{U(t)-K_P(U(t))}\cdot P = \mathrm{e}_P(U(t)),\quad W(t) = {}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U(t))}V(t)\,,$$

with $t \mapsto \gamma_P(t) = (U(t),V(t))$ being a smooth curve in $(S_P\,\mathcal{E}\,(\mu))^2$. Let us compute the velocity of variation of the Lagrangian L along the curve γ.

$$\frac{d}{dt}L(\gamma(t)) = \frac{d}{dt}L(Q(t),W(t)) = \frac{d}{dt}L(\mathrm{e}_P(U(t)),{}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U(t))}V(t)) = \frac{d}{dt}L_P(U(t),V(t))\,,$$

with $L_P(U,V) = L(\mathrm{e}_P(U),{}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)}V)$. It follows that

$$\frac{d}{dt}L(Q(t),W(t)) = d_1L_P(U(t),V(t))[\dot{U}(t)] + d_2L_P(U(t),V(t))[\dot{V}(t)]\,. \tag{16}$$

If we write $Q = \mathrm{e}_P(U)$ and $W = {}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)}V$, then we have

$$d_2L_P(U,V)[H_2] = \left.\frac{d}{dt}L_P(U,V+tH_2)\right|_{t=0} = \left.\frac{d}{dt}L(Q,W+t\,{}^{\mathrm{e}}\mathbb{U}_P^QH_2)\right|_{t=0} = d_2L(Q,W)[{}^{\mathrm{e}}\mathbb{U}_P^QH_2]\,, \tag{17}$$

where d_2L is the fiber derivative of L. As $\dot{U}(t) = {}^{\mathrm{e}}\mathbb{U}_{Q(t)}^P\overset{\star}{Q}(t)$ and ${}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U(t))}\dot{V}(t) = \overset{\star}{W}(t)$, it follows from Eq.s (16) and (17), that

$$\frac{d}{dt}L(Q(t),W(t)) = d_1L_P(U(t),V(t))[{}^{\mathrm{e}}\mathbb{U}_{Q(t)}^P\overset{\star}{Q}(t)] + d_2L(Q(t),W(t))[\overset{\star}{W}(t)]\,.$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 January 2018 doi:10.20944/preprints201712.0191.v2

Peer-reviewed version available at Entropy 2018, 20, 139; doi:10.3390/e20020139

In the equation above the first term in the RHS does not depend on P because the LHS and the second term of the RHS do not depend on P. Hence we define the first partial derivative of the Lagrangian function to be

$$d_1(Q,W)[H_1] = d_1 L_P(U,V)[{}^{\mathrm{e}}\mathbb{U}^P_{\mathrm{e}_P(U)} H_1] \,, \quad H_1 \in S_Q\,\mathcal{E}\,(\mu)\;, \tag{18}$$

so that the derivative of L along γ becomes

$$\frac{d}{dt} L(Q(t),W(t)) = d_1 L(Q(t),W(t))[\overset{\star}{Q}(t)] + d_2 L(Q(t),W(t))[\overset{\star}{W}(t)]\,. \tag{19}$$

In particular, if $W(t) = \overset{\star}{Q}(t)$, then

$$\frac{d}{dt} L(Q(t),\overset{\star}{Q}(t)) = d_1 L(Q(t),\overset{\star}{Q}(t))[\overset{\star}{Q}(t)] + d_2 L(Q(t),\overset{\star}{Q}(t))[\overset{\star\star}{Q}(t)]\,,$$

see Eq. (5).

Example 4 (Running example II)**.** With the Lagrangian of Eq. (15), we have

$$L_P(U,V) = \frac{1}{2}\left\langle {}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)} V, {}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)} V \right\rangle_{\mathrm{e}_P(U)} - \kappa\,\mathbb{E}_{\mathrm{e}_P(U)}\left[U - K_P(U) + \log P\right] = \\ \frac{1}{2} d^2 K_P(U)[V,V] + \kappa\left(K_P(U) - dK_P(U)[U + \log P + \mathcal{H}\,(P)] + \mathcal{H}\,(P)\right)\;,$$

see Eq.s (9) and (11). The first partial derivative is

$$\begin{aligned}
&d_1 L_P(U,V)[H_1] = \\
&\frac{1}{2} d^3 K_P(U)[V,V,H_1] + \kappa\left(dK_P(U)[H_1] - d^2 K_P(U)[U + \log P + \mathcal{H}\,(P)\,, H_1] - dK_P(U)[H_1]\right) = \\
&\qquad \frac{1}{2} d^3 K_P(U)[V,V,H_1] - \kappa d^2 K_P(U)[U + \log P + \mathcal{H}\,(P)\,, H_1] = \\
&\qquad \frac{1}{2}\,\mathbb{E}_Q\left[W^2\,{}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)} H_1\right] - \kappa\,\mathbb{E}_Q\left[(\log Q + \mathcal{H}\,(Q))\,{}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)} H_1\right] = \\
&\qquad\qquad \mathbb{E}_Q\left[\left(\frac{1}{2}\left(W^2 - \mathbb{E}_Q\left[W^2\right]\right) - \kappa(\log Q + \mathcal{H}\,(Q))\right){}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)} H_1\right]\;,
\end{aligned}$$

where we have used Eq.s (9) and (10) together with ${}^{\mathrm{e}}\mathbb{U}_P^{\mathrm{e}_P(U)}(U + \log P + \mathcal{H}\,(P)) = \log Q + \mathcal{H}\,(Q)$.

We have found that

$$d_1 L(Q,W)[H_1] = \left\langle \frac{1}{2}\left(W^2 - \mathbb{E}_Q\left[W^2\right]\right) - \kappa(\log Q + \mathcal{H}\,(Q)), H_1 \right\rangle_Q \;, \quad H_1 \in S_Q\,\mathcal{E}\,(\mu)\;, \tag{20}$$

and also

$$d_1 L(Q(t),\overset{\star}{Q}(t))[\overset{\star}{Q}(t)] = \left\langle \frac{1}{2}\left(\overset{\star}{Q}(t)^2 - \mathbb{E}_Q\left[\overset{\star}{Q}(t)^2\right]\right) - \kappa(\log Q + \mathcal{H}\,(Q)), \overset{\star}{Q}(t) \right\rangle_Q \;.$$

Using the fiber derivative computed in the first part of the running example, we find

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 January 2018 doi:10.20944/preprints201712.0191.v2

Peer-reviewed version available at *Entropy* **2018**, *20*, 139; doi:10.3390/e20020139

$$\frac{d}{dt}L(Q(t),\overset{\star}{Q}(t)) = \left\langle \frac{1}{2}\left(\overset{\star}{Q}(t)^2 - \mathbb{E}_Q\left[\overset{\star}{Q}(t)^2\right]\right) - \kappa(\log Q + \mathcal{H}(Q)), \overset{\star}{Q}(t)\right\rangle_Q + \left\langle \overset{\star}{Q}(t), \overset{\star\star}{Q}(t)\right\rangle_Q .$$

Notice that Eq. (12) shows that one of the term in the equations above is grad $\mathcal{H}(Q)$.

5. Action integral

If $[0,1] \ni t \mapsto Q(t)$ is a smooth curve in the exponential manifold, then the *action integral*

$$A(Q) = \int_0^1 L(Q(t),\overset{\star}{Q}(t))\, dt$$

is well defined. We consider the expression of Q in the chart centered at P, $Q(t) = \mathrm{e}^{U(t)-K_P(U(t))} \cdot P$.

Given $\varphi \in C^1([0,1])$ with $\varphi(0) = \varphi(1) = 0$, for each $\delta \in \mathbb{R}$ and $H \in S_P\,\mathcal{E}(\mu)$ we define the perturbed curve

$$Q_\delta(t) = \mathrm{e}^{(U(t)+\delta\varphi(t)H)-K_P(U(t)+\delta\varphi(t)H)} \cdot P .$$

We have $Q_\delta(0) = Q(0)$, $Q_\delta(1) = Q(1)$, and

$$\overset{\star}{Q}_\delta(t) = \dot{U}(t) + \delta\dot{\varphi}(t)H - \mathbb{E}_{Q_\delta(t)}\left[(\dot{U}(t) + \delta\dot{\varphi}(t))H\right] ,$$

whose expression in the chart centered at P is $\dot{U}(t) + \delta\dot{\varphi}(t)H$.

Let us consider the variation in δ of the action integral. We apply Eq. (19) applied to the smooth curve in $S\,\mathcal{E}(\mu)$ given by

$$\delta \mapsto (Q_\delta(t), \overset{\star}{Q}_\delta(t)) ,$$

where t is fixed. As

$$\frac{d}{d\delta}\log Q_\delta(t) = \frac{d}{d\delta}(U(t) + \delta\varphi(t)H) - \mathbb{E}_{Q_\delta(t)}\left[\frac{d}{d\delta}(U(t)+\delta\varphi(t)H)\right] = \varphi(t)(H - \mathbb{E}_{Q_\delta(t)}[H])$$

and

$${}^{\mathrm{e}}\mathbb{U}_P^{Q_\delta(t)} \frac{d}{d\delta}(\dot{U}(t) + \delta\dot{\varphi}(t)H) = \dot{\varphi}(t)(H - \mathbb{E}_{Q_\delta(t)}[H]) ,$$

we obtain

$$\begin{aligned}\frac{d}{d\delta}A(Q_\delta) = \int_0^1 \frac{d}{d\delta}L(Q_\delta(t),\overset{\star}{Q}_\delta(t))\, dt = \\ \int_0^1 \left(\varphi(t) d_1 L(Q_\delta(t),\overset{\star}{Q}_\delta(t))[H - \mathbb{E}_{Q_\delta(t)}[H]] + \dot{\varphi}(t) d_2 L(Q_\delta(t),\overset{\star}{Q}_\delta(t))[H - \mathbb{E}_{Q_\delta(t)}[H]]\right) dt = \\ \int_0^1 \varphi(t)\left(d_1 L(Q_\delta(t),\overset{\star}{Q}_\delta(t))[H - \mathbb{E}_{Q_\delta(t)}[H]] - \frac{d}{dt} d_2 L(Q_\delta(t),\overset{\star}{Q}_\delta(t))[H - \mathbb{E}_{Q_\delta(t)}[H]]\right) dt .\end{aligned}$$

If $t \mapsto Q(t)$ is an critical curve of the action integral, then $\frac{d}{d\delta}A(Q_\delta)\Big|_{\delta=0} = 0$, hence for all φ and H we have

$$\int_0^1 \varphi(t)\left(d_1 L(Q(t),\overset{\star}{Q}(t))[H - \mathbb{E}_{Q(t)}[H]] - \frac{d}{dt} d_2 L(Q(t),\overset{\star}{Q}(t))[H - \mathbb{E}_{Q(t)}[H]]\right) dt = 0 . \qquad (21)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 January 2018 doi:10.20944/preprints201712.0191.v2

Peer-reviewed version available at *Entropy* **2018**, *20*, 139; doi:10.3390/e20020139

This, in turn, implies that for each $t \in [0,1]$ and $H \in S_{Q(t)}\,\mathcal{E}\,(\mu)$ the Euler-Lagrange equation holds:

$$d_1 L(Q(t), \overset{\star}{Q}(t))[H] - \frac{d}{dt} d_2 L(Q(t), \overset{\star}{Q}(t))[H] = 0\,. \tag{22}$$

Example 5 (Running example III)**.** For the Lagrangian of Eq. (15) we can use Eq. (20) in the form

$$d_1 L(Q(t), \overset{\star}{Q}(t))[H - \mathbb{E}_{Q(t)}[H]] = \\ \left\langle \frac{1}{2}\left(\overset{\star}{Q}(t)^2 - \mathbb{E}_{Q(t)}\left[\overset{\star}{Q}(t)^2\right]\right) - \kappa(\log(Q(t)) + \mathcal{H}(Q(t))), H - \mathbb{E}_{Q(t)}[H] \right\rangle_{Q(t)},$$

with $H \in S_P\,\mathcal{E}\,(\mu)$. For the other term, we have

$$d_2 L(Q(t), \overset{\star}{Q}(t))[H - \mathbb{E}_{Q(t)}[H]] = \left\langle \overset{\star}{Q}(t), H - \mathbb{E}_{Q(t)}[H] \right\rangle_{Q(t)} = d^2 K_P(U(t))[\dot{U}(t), H]\,,$$

whose derivative is

$$\frac{d}{dt}\, d^2 K_P(U(t))[\dot{U}(t), HR] = d^3 K_P(U(t))[\dot{U}(t), \dot{U}(t), H] + d^2 K_P(U(t))[\ddot{U}(t), H] = \\ \mathbb{E}_{Q(t)}\left[\overset{\star}{Q}(t)^2 (H - \mathbb{E}_{Q(t)}[H])\right] + \mathbb{E}_{Q(t)}\left[\overset{\star\star}{Q}(t)(H - \mathbb{E}_{Q(t)}[H])\right] = \\ \mathbb{E}_{Q(t)}\left[\left(\overset{\star}{Q}(t)^2 - \mathbb{E}_{Q(t)}\left[\overset{\star}{Q}(t)^2\right]\right)(H - \mathbb{E}_{Q(t)}[H])\right] + \mathbb{E}_{Q(t)}\left[\overset{\star\star}{Q}(t)(H - \mathbb{E}_{Q(t)}[H])\right].$$

Dropping the generic H, the Euler-Lagrange equation becomes

$$\overset{\star\star}{Q}(t) + \left(\overset{\star}{Q}(t)^2 - \mathbb{E}_{Q(t)}\left[\overset{\star}{Q}(t)^2\right]\right) = \frac{1}{2}\left(\overset{\star}{Q}(t)^2 - \mathbb{E}_{Q(t)}\left[\overset{\star}{Q}(t)^2\right]\right) - \kappa(\log(Q(t)) + \mathcal{H}(Q(t)))\,,$$

that is

$$\overset{\star\star}{Q}(t) + \frac{1}{2}\left(\overset{\star}{Q}(t)^2 - \mathbb{E}_{Q(t)}\left[\overset{\star}{Q}(t)^2\right]\right) = -\kappa(\log(Q(t)) + \mathcal{H}(Q(t)))\,,$$

The equation above has been derived using the exponential affine geometry of the statistical bundle and involves $\overset{\star\star}{Q}(t)$. However, by using Eq.s (5), (6), and (12), we find the equivalent form

$${}^{0}\mathrm{D}^2 Q(t) = \kappa\,\mathrm{grad}\,\mathcal{H}(Q(t))\ .$$

6. Discussion

We have shown that the research program consisting in applying to Statistics concepts taken from Classical Mechanics makes sense, even if no practical application has been produced in this paper. Some simple examples have been discussed in order to show clearly that the language from classical mechanics is indeed suggestive when applied to typical concepts in Statistics such as Fisher score and statistical entropy. The derivation of the Euler-Lagrange equations is classically done in the set-up of the Riemannian geometry, while here we have used the affine structure of Information Geometry. The present provisional results prompt to a generalization to non-finite sample spaces and the development of applications. Finally, the related Hamiltonian formalism remains to be investigated.

Acknowledgments: The Author gratefully thanks Hiroshi Matsuzoe (Nagoya Institute of Technology, JP), Lamberto Rondoni (Politecnico di Torino, IT), Antonio Scarfone (CNR and Politecnico di Torino, IT), Tatsuaki Wada (Ibaraki University, JP), for their interesting comments on early versions of this piece of research. He thanks two anonymous referees for their useful and enlightening comments. He acknowledges the support of de Castro Statistics, Collegio Carlo Alberto, and of GNAMPA-INdAM.

Conflicts of Interest: The author declare no conflict of interest.

References

1. Arnold, V.I. *Mathematical methods of classical mechanics*; Vol. 60, *Graduate Texts in Mathematics*, Springer-Verlag, New York, 1989; pp. xvi+516. Translated from the 1974 Russian original by K. Vogtmann and A. Weinstein, Corrected reprint of the second (1989) edition.
2. Abraham, R.; Marsden, J.E. *Foundations of mechanics*; Benjamin/Cummings Publishing Co., Inc., Advanced Book Program, Reading, Mass., 1978; pp. xxii+m–xvi+806. Second edition, revised and enlarged, With the assistance of Tudor Raţiu and Richard Cushman.
3. Marsden, J.E.; Ratiu, T.S. *Introduction to mechanics and symmetry*, second ed.; Vol. 17, *Texts in Applied Mathematics*, Springer-Verlag, New York, 1999; pp. xviii+582. A basic exposition of classical mechanical systems.
4. Amari, S.; Nagaoka, H. *Methods of information geometry*; American Mathematical Society, 2000; pp. x+206. Translated from the 1993 Japanese original by Daishi Harada.
5. Leok, M.; Zhang, J. Connecting Information Geometry and Geometric Mechanics. *Entropy* **2017**, *19*, 518.
6. Pistone, G. Nonparametric information geometry. In *Geometric science of information*; Nielsen, F.; Barbaresco, F., Eds.; Springer, Heidelberg, 2013; Vol. 8085, *Lecture Notes in Comput. Sci.*, pp. 5–36. First International Conference, GSI 2013 Paris, France, August 28-30, 2013 Proceedings.
7. Lang, S. *Differential and Riemannian manifolds*, third ed.; Vol. 160, *Graduate Texts in Mathematics*, Springer-Verlag, 1995; pp. xiv+364.
8. Pistone, G. Examples of the application of nonparametric information geometry to statistical physics. *Entropy* **2013**, *15*, 4042–4065.
9. Lods, B.; Pistone, G. Information Geometry Formalism for the Spatially Homogeneous Boltzmann Equation. *Entropy* **2015**, *17*, 4323–4363.
10. Pistone, G.; Rogantin, M. The exponential statistical manifold: mean parameters, orthogonality and space transformations. *Bernoulli* **1999**, *5*, 721–760.
11. Efron, B.; Hastie, T. *Computer age statistical inference*; Vol. 5, *Institute of Mathematical Statistics (IMS) Monographs*, Cambridge University Press, New York, 2016; pp. xix+475. Algorithms, evidence, and data science.
12. Gibilisco, P.; Pistone, G. Connections on non-parametric statistical manifolds by Orlicz space geometry. *IDAQP* **1998**, *1*, 325–347.
13. Landau, L.D.; Lifshits, E.M. *Course of Theoretical Physics. Statistical Physics.*, 3rd ed.; Vol. V, Butterworth-Heinemann, 1980.
14. Shima, H. *The geometry of Hessian structures*; World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2007; pp. xiv+246.