

Global Geometry of Bayesian Statistics

Atsuhide Mori

Department of Mathematics, Osaka Dental University, Osaka 573-1121, Japan; mori-a@cc.osaka-dent.ac.jp

Received: 15 January 2020; Accepted: 18 February 2020; Published: 20 February 2020



Abstract: In the previous work of the author, a non-trivial symmetry of the relative entropy in the information geometry of normal distributions was discovered. The same symmetry also appears in the symplectic/contact geometry of Hilbert modular cusps. Further, it was observed that a contact Hamiltonian flow presents a certain Bayesian inference on normal distributions. In this paper, we describe Bayesian statistics and the information geometry in the language of current geometry in order to spread our interest in statistics through general geometers and topologists. Then, we foliate the space of multivariate normal distributions by symplectic leaves to generalize the above result of the author. This foliation arises from the Cholesky decomposition of the covariance matrices.

Keywords: information geometry; poisson structure; symplectic structure; contact structure; foliation; Cholesky decomposition

1. Introduction

Suppose that a smooth manifold U is embedded in the space of positive probability densities defined on a fixed domain. Then, the relative entropy defines a separating premetric $D : U \times U \rightarrow \mathbb{R}_{\geq 0}$ on U . Here a premetric on U is a non-negative function on $U \times U$ vanishing along the diagonal set $\Delta \subset U \times U$, and it is separating if it vanishes only at Δ . Its jet of order 3 at Δ induces a family of differential geometric structures on U , which is the main subject of the information geometry. There is a large body of literature on the information geometry (see [1], [2] and references therein). It is worth noting that another “canonical” choice of premetric other than the above D is discussed in [3].

In the case where U is the space of univariate normal distributions, the half plane $\mathbb{H} = \mathbb{R} \times \mathbb{R}_{>0} \ni (m, s)$ presents U , where m denotes the mean and s the standard deviation. Since the convolution of two normal densities is a normal density, it induces a product $*$ on $\mathbb{H}$ called the convolution product. On the other hand, since the pointwise product of two normal densities is proportional to a normal density, it induces another product $\cdot$ on $\mathbb{H}$ called the Bayesian product. Their expressions are

$$(m, s) * (m', s') = \left(m + m', \sqrt{s^2 + s'^2} \right), \quad (m, s) \cdot (m', s') = \left(\frac{ms'^2 + m's^2}{s^2 + s'^2}, \frac{ss'}{\sqrt{s^2 + s'^2}} \right). \quad (1)$$

In the previous work of the author ([4]), a Fourier-like transformation is defined as the diffeomorphism $\hat{F} : \mathbb{H} \rightarrow \mathbb{H}$ sending (m', s') to $(M, S) = \left(-\frac{m'}{s'^2}, \frac{1}{s'} \right)$. It is an involution interchanging the two operations $*$ and $\cdot$. Accordingly, the stereograph $f : \mathbb{H} \times \mathbb{H} \rightarrow \mathbb{R}_{>0}$ of the above D is defined by

$$\begin{aligned} f : (m, s, M, S) &\mapsto \frac{1}{2} \left(\frac{M}{S} + sS \frac{m}{s} \right)^2 + \frac{1}{2} (sS)^2 - \frac{1}{2} - \ln(sS) \\ &= \frac{1}{2} \left\{ \frac{(m - m')^2 + s^2}{s'^2} - 1 + \ln \frac{s^2}{s'^2} \right\} = D((m, s), (m', s')). \end{aligned} \quad (2)$$

The flow $(m, s, M, S) \mapsto (e^t m, e^t s, e^{-t} M, e^{-t} S)$ ($t \in \mathbb{R}$) preserves f as well as the graph $F \subset \mathbb{H} \times \mathbb{H}$ of $\hat{F}$. The same symmetry appears in the contact/symplectic geometry related to the algebraic geometry of Hilbert modular cusps. Moreover, there exists a contact Hamiltonian flow whose restriction to the graph F presents a certain Bayesian inference. Its application appears in [5].

In this paper, we describe Bayesian statistics in the language of current geometry in order to share the problems among general geometers and topologists. Then, generalizing the above result of the author, we foliate the space of multivariate normal distributions by using the Cholesky decomposition of the covariance matrices and define on each leaf the Fourier-like transformation, the stereograph of the relative entropy, and the contact Hamiltonian flow presenting a Bayesian inference. The ultimate aim of this research is to construct a Bayesian statistical model of space-time on which everything learns by changing its inner distribution along the leaf.

2. Results

2.1. Symplectic/Contact Geometry

Current geometry does not heavily use tensor calculus. Instead, it uses (exterior) differential forms, which can be integrated along cycles, pulled-back under smooth maps, and differentiated without affine connections. In symplectic/contact geometry, the readers must be familiar with differential forms. Then, this subsection is the minimal summary of definitions. For the details, refer to [6].

A (positive) symplectic form on an oriented $2n$ -manifold is a closed 2-form ω satisfying $\omega^n > 0$, where $\omega^n = \omega \wedge \cdots \wedge \omega$. If the orientation is reversed, the 2-form ω becomes a negative symplectic form. In either case, a symplectic form ω identifies a vector field X with an exact 1-form dH through the one-to-one correspondence defined by Hamilton's equation $\iota_X \omega = -dH$. Here ι denotes the interior product. Then, X is called a Hamiltonian vector field of the primitive function H (+constant). The flow generated by X preserves the symplectic form ω . Namely, the Lie derivative $\mathcal{L}_X \omega (= \iota_X d\omega + d\iota_X \omega)$ vanishes. A Lagrangian submanifold is an n -manifold which is immersed in a symplectic $2n$ -manifold so that the pull-back of the symplectic form vanishes. The word "symplectic" is a calque of "complex". Indeed, there exists an almost complex structure J which is compatible with a given symplectic structure, i.e., for which the composition $\omega(\cdot, J\cdot)$ is a Riemannian metric. In the case where J is integrable, ω is called a Kähler form of the complex manifold.

On the other hand, a (positive) contact form on an oriented $(2n - 1)$ -manifold N is a 1-form η satisfying $\eta \wedge (d\eta)^{n-1} > 0$. A (co-oriented) contact structure on N is the conformal class of a contact form. It can be presented as the oriented hyperplane field $\ker \eta$. The product manifold $\mathbb{R}(\ni t) \times N$ carries the exact symplectic form $d(e^t \eta)$. Take a function h on N . Let X be the Hamiltonian vector field of the function $e^t h$ defined on the product manifold $\mathbb{R} \times N$. Then, the push-forward Y of X under the projection of $\mathbb{R} \times N$ to the second factor is well-defined. The vector field Y is called the contact Hamiltonian vector field of the function h on N . The pair of the equations $\eta(Y) = h$ and $\eta \wedge \mathcal{L}_Y \eta = 0$ uniquely determines Y . A Legendrian submanifold is an $(n - 1)$ -manifold which is immersed in a contact $(2n - 1)$ -manifold so that the pull-back of the contact form vanishes.

2.2. Bayesian Statistics

Suppose that any point y of a smooth manifold M equipped with a volume form $dvol$ presents a positive probability density or probability $\rho_y : W \rightarrow \mathbb{R}_{>0}$ defined on a (possibly discrete) measurable space W , where ρ_y depends smoothly on y , and $\rho_y \neq \rho_{y'}$ for $y \neq y' \in M$. Let $\mathcal{V}$ be the space of positive volume forms with finite total volume on M . Take an arbitrary element $V \in \mathcal{V}$ and consider it as the initial state of the mind M of an agent. Here W stands for (a part of) the world for the agent. Finding a datum $w \in W$ in his world, the agent can take the value $\rho_y(w)$ as a smooth positive function

$\bar{\rho}_w : y \mapsto \rho_y(w)$ on M , which is called the likelihood of the datum w . Then, he can multiply the volume form V by $\bar{\rho}_w > 0$ to obtain a new element of $\mathcal{V}$. This defines the updating map

$$\varphi : W \times \mathcal{V} \ni (w, V) \mapsto \bar{\rho}_w V \in \mathcal{V}. \quad (3)$$

The “psychological” probability density p_V on the mind M defined by $p_V dvol = V / \int_M V$ is accordingly updated into the density $p_{\varphi(w, V)} \propto p_V \bar{\rho}_w$, which is called the conditional probability density given the datum w . Practically, Bayes’ rule on conditional probabilities is expressed as

$$P(y \in \Delta_y | w \in \Delta_w) = \frac{P(w \in \Delta_w | y \in \Delta_y)}{P(w \in \Delta_w)} \cdot P(y \in \Delta_y). \quad (4)$$

Here P denotes the probability of an event, Δ_y (respectively, Δ_w) a small portion of M (respectively, W). Since the state of the world does not depend on the mind of the agent, the probability $P(w \in \Delta_w)$ is independent of y , and therefore approximates a constant on M . On the other hand, the conditional probability $P(w \in \Delta_w | y \in \Delta_y)$ of the datum w approximates a function of y which is clearly proportional to the above likelihood. This implies that the factor $\frac{P(w \in \Delta_w | y \in \Delta_y)}{P(w \in \Delta_w)}$ in the right-hand side of Bayes’ rule (Equation (4)) is approximately proportional to the likelihood. Thus, Bayes’ rule (Equation (4)) implies the updating of p_V via the formula (Equation (3)). The Bayesian product · mentioned in the introduction appears in this context. Namely, the variable of the first factor is the mean y of a normal distribution on W . The density of the normal distribution at the datum w can be considered as a function of y , which is proportional to a normal distribution on M . Thus the Bayesian product of normal distributions on M presents the updating of the density of the predictive mean in the mind of the agent.

The aim of Bayesian updating is practical to many people. Indeed, the aim of the above updating is the estimation of the mean. Nevertheless, it is quite natural that a geometer multiplies a volume form by a positive function once he is given them. In this regard, we can say that the aim of Bayesian updating is a geometric setting of a dynamical system. In particular, a Bayesian updating in the conjugate prior is at first, simply the iteration of a Bayesian product.

2.3. The Information Geometry

Suppose that a manifold U is embedded in the subset $\{V \in \mathcal{V} \mid \int_M V = 1\} = \{p_V dvol \mid V \in \mathcal{V}\}$. Hereafter, we identify the element $p_V dvol \in U$ with the “psychological” probability density p_V . We call U a conjugate prior for the updating map φ if the cone $\tilde{U} = \{e^t V \mid t \in \mathbb{R}, V \in U\}$ satisfies $\varphi(W \times \tilde{U}) \subset \tilde{U}$. (Whether there exists a preferred conjugate prior or not, how to determine the initial state of the mind is another interesting problem. For example, one may fix the asymptotic behavior of the state of mind according to the aim of the Bayesian inference and search for the optimal decision of the initial state. See [7] for an approach to this problem via the information geometry.)

Now we define the “distance” $\tilde{D} : \tilde{U} \times \tilde{U} \rightarrow \mathbb{R}$ on $\tilde{U}$, which satisfies none of the axioms of distance, by

$$\tilde{D}(V_1, V_2) = \int_M V_1 \ln \frac{V_1}{V_2} \quad (\text{the relative entropy}). \quad (5)$$

From the convexity of $-\ln$, we see that the restriction $D = \tilde{D}|_{U \times U}$ ($\geq -\ln(\int_M V_2) = 0$) is a separating premetric on U , which is called the Kullback–Leibler divergence in information theory. This implies that the germ of D along the diagonal set Δ of $U \times U$ represents the zero section of the cotangent sheaf of U , that is, for any point $x = (x^1, \dots, x^n)$ of any chart of U , the Taylor expansion of $D(x + \frac{1}{2}dx, x - \frac{1}{2}dx)$ has no linear terms. Thus the differential $dD : TU \times TU \rightarrow \mathbb{R}$ also vanishes on the diagonal set Δ' of $TU \times TU$. We regard the 1-form on TU represented by the germ of dD along Δ' as a quadratic tensor, and denote it by g (note that $g_x : T_x U \times T_x U \rightarrow \mathbb{R}$ is linear). It appears as two times the

quadratic terms $\frac{1}{2} \sum_{i,j} g_{ij} dx^i dx^j$ ($g_{ij} = g_{ji}$) in the above Taylor expansion. Of course, it also appears in the Taylor expansion of $D\left(x - \frac{1}{2}dx, x + \frac{1}{2}dx\right)$. Thus it can also be considered as the quadratic terms of the symmetric sum $D\left(x + \frac{1}{2}dx, x - \frac{1}{2}dx\right) + D\left(x - \frac{1}{2}dx, x + \frac{1}{2}dx\right)$. The symmetric matrix $[g_{ij}]$ is called the Fisher information in information theory. From the non-negativity of D , we may assume generically that g is a Riemannian metric. We would like to notice that this construction of Riemannian metric by means of symmetric sum also works over $\tilde{U}$. Indeed, we have

$$\tilde{D}(e^{t+\frac{1}{2}dt}V, e^{t-\frac{1}{2}dt}V) + \tilde{D}(e^{t-\frac{1}{2}dt}V, e^{t+\frac{1}{2}dt}V) = e^t dt^2 + O(dt^3). \quad (6)$$

Let ∇^0 be the Levi-Civita connection of g . We write the lowest degree terms in the Taylor expansions of $D\left(x + \frac{1}{2}dx, x - \frac{1}{2}dx\right) - D\left(x - \frac{1}{2}dx, x + \frac{1}{2}dx\right)$ as $\frac{1}{3} \sum_{i,j,k} T_{ijk} dx^i dx^j dx^k$ ($T_{ijk} = T_{jik} = T_{ikj}$). This presents the symmetric cubic tensor T , which can be constructed from the anti-symmetric difference $D'(x_1, x_2) = D(x_1, x_2) - D(x_2, x_1)$ similarly as above. One can use it to deform the Levi-Civita connection ∇^0 into the α -connections $\nabla^\alpha = \nabla^0 - \alpha g^* T$ ($\alpha \in \mathbb{R}$) without torsion, where g^* denotes the dual metric. Especially, we call ∇^1 and ∇^{-1} respectively the e-connection and the m-connection. The symmetric tensor T is sometimes called skewness since it presents the asymmetry of D . The information geometry concerns the family of α -connections as well as the Fisher information metric on U . We usually do not extend it over $\tilde{U}$ for the symmetric sum of $\tilde{D}$ lacks asymmetry.

2.4. The Geometry of Normal Distributions

We consider the space U of multivariate normal distributions on $M = \mathbb{R}^n$. A vector $\mu = (\mu_i)_{1 \leq i \leq n} \in \mathbb{R}^n$ and an upper triangular matrix $C = [c_{ij}]_{1 \leq i, j \leq n} \in \text{Mat}(n, \mathbb{R})$ with positive diagonal entries parameterize U by declaring that μ presents the mean and $C^T C$ the Cholesky decomposition of the covariance matrix. Further we put

$$\sigma_i = c_{ii}, \quad r_{ij} = \frac{c_{ij}}{c_{ii}} \quad (i, j \in \{1, \dots, n\}), \quad C = \text{diag}(\sigma)[r_{ij}]. \quad (7)$$

The matrix $[r_{ij}]$ is unitriangular, i.e., a triangular matrix whose diagonal entries are all 1. Then, each point $x = (\mu, \sigma, r) \in U = \mathbb{R}^n \times (\mathbb{R}_{>0})^n \times \mathbb{R}^{n(n-1)/2}$ presents the volume form

$$V_x = \frac{1}{\sqrt{(2\pi)^n \sigma_1 \cdots \sigma_n}} \exp\left(-\frac{1}{2} \|C(\sigma, r)^{-T}(y - \mu)\|^2\right) d\text{vol}. \quad (8)$$

Let $\|\cdot\|^2$ denote the sum of the squares of the entries of a matrix as well as a vector. The relative entropy defines the premetric $D(x, x' = (\mu', \sigma', r'))$ by

$$D(x, x') = \frac{\|C(\sigma', r')^{-T}(\mu' - \mu)\|^2}{2} + \frac{\|C(\sigma, r)C(\sigma', r')^{-1}\|^2 - n}{2} - \sum_{i=1}^n \ln \frac{\sigma_i}{\sigma'_i}, \quad (9)$$

Let 1_n denote the unit matrix, and ΔC the difference $C(\sigma + \Delta\sigma, r + \Delta r) - C(\sigma, r)$. We have

$$D(x + \Delta x, x) = \frac{\|C^{-T}\Delta\mu\|^2}{2} + \frac{\|\Delta C C^{-1}\|^2}{2} + \text{tr}(\Delta C C^{-1}) - \ln |1_n + \Delta C C^{-1}|. \quad (10)$$

Let r^{ij} be the entries of the inverse matrix of $[r_{ij}]$. We have

$$(\text{the } ij\text{-entry of } \Delta C C^{-1}) = \begin{cases} \frac{\Delta \sigma_i}{\sigma_i} & (i = j) \\ \frac{\sigma_i + \Delta \sigma_i}{\sigma_j} \sum_{k=i+1}^j r^{kj} \Delta r_{ik} & (i < j) \\ 0 & (i > j) \end{cases} \quad (11)$$

From Equations (10) and (11), we see that the Fisher information metric g is expressed as

$$g = \sum_{k=1}^n \left(\frac{1}{\sigma_k} \sum_{i=1}^k r^{ik} d\mu_i \right)^2 + 2 \sum_{i=1}^n \left(\frac{d\sigma_i}{\sigma_i} \right)^2 + \sum_{l=1}^{n-1} \sum_{k=l+1}^n \left(\frac{\sigma_l}{\sigma_k} \sum_{i=l+1}^k r^{ik} dr_{li} \right)^2. \quad (12)$$

Put

$$\begin{aligned} g_{\mu\mu} &= [g_{\mu_i, \mu_j}] = \left[\sum_{k \geq i, j} \frac{r^{ik} r^{jk}}{\sigma_k^2} \right] = C^{-1} C^{-T}, \\ g_{\sigma\sigma} &= \text{diag} \left(\left(\frac{2}{\sigma_i^2} \right) \right), \\ g_{rr, l} &= [g_{r_{li} r_{lj}}]_{i, j > l} = [\sigma_l^2 g_{\mu_i, \mu_j}]_{i, j > l} \quad (l = 1, \dots, n-1). \end{aligned} \quad (13)$$

Then, the representing matrix of g is the following block diagonal matrix:

$$\text{diag}(g_{\mu\mu}, g_{\sigma\sigma}, g_{rr, 2}, \dots, g_{rr, n}). \quad (14)$$

Lowering the upper indices of the α -connection with $\sum_L g_{KL} \Gamma^{\alpha}_{IJ}{}^K = \Gamma^{\alpha}_{\{I, J\}, K}$, we have

$$\Gamma^0_{\{\mu_i, \mu_j\}, \sigma_k} = -\Gamma^0_{\{\mu_i, \sigma_k\}, \mu_j} = \frac{r^{ik} r^{jk}}{\sigma_k^3}, \quad (15)$$

$$\Gamma^0_{\{\sigma_i, \sigma_i\}, \sigma_i} = \frac{-2}{\sigma_i^3}, \quad (16)$$

$$\Gamma^0_{\{\mu_i, \mu_j\}, r_{ab}} = -\Gamma^0_{\{\mu_i, r_{ab}\}, \mu_j} = \sum_{k=b}^n \frac{r^{bk} (r^{ia} r^{jk} + r^{ik} r^{ja})}{2\sigma_k^2}, \quad (17)$$

$$\Gamma^0_{\{r_{li}, r_{lj}\}, \sigma_l} = -\Gamma^0_{\{r_{li}, \sigma_l\}, r_{lj}} = \sum_{k \geq i, j} \frac{-\sigma_l r^{ik} r^{jk}}{\sigma_k^2}, \quad (18)$$

$$\Gamma^0_{\{r_{li}, r_{lj}\}, \sigma_k} = -\Gamma^0_{\{r_{li}, \sigma_k\}, r_{lj}} = \frac{\sigma_l^2 r^{ik} r^{jk}}{\sigma_k^3} \quad (k \geq i, j), \quad (19)$$

$$\Gamma^0_{\{r_{li}, r_{lj}\}, r_{ab}} = -\Gamma^0_{\{r_{li}, r_{ab}\}, r_{lj}} = \sigma_l^2 \Gamma^0_{\{\mu_i, \mu_j\}, r_{ab}} \quad (a > l), \quad (20)$$

$$\Gamma^0_{\{I, J\}, K} = 0 \quad (\text{for the other choices of } \{I, J\} \text{ and } K). \quad (21)$$

With respect to the same coordinates, the coefficients of the e-connection are

$$\Gamma^1_{\{\mu_i, \sigma_k\}, \mu_j} = 2\Gamma^0_{\{\mu_i, \sigma_k\}, \mu_j}, \quad (22)$$

$$\Gamma^1_{\{\sigma_i, \sigma_i\}, \sigma_i} = 3\Gamma^0_{\{\sigma_i, \sigma_i\}, \sigma_i}, \quad (23)$$

$$\Gamma^1_{\{\mu_i, r_{ab}\}, \mu_j} = 2\Gamma^0_{\{\mu_i, r_{ab}\}, \mu_j}, \quad (24)$$

$$\Gamma^1_{\{r_{li}, r_{lj}\}, \sigma_l} = 2\Gamma^0_{\{r_{li}, r_{lj}\}, \sigma_l}, \quad (25)$$

$$\Gamma_{\{r_{li}, \sigma_k\}, r_{lj}}^1 = 2\Gamma_{\{r_{li}, \sigma_k\}, r_{lj}}^0 \quad (k \geq i, j), \quad (26)$$

$$\Gamma_{\{r_{li}, r_{ab}\}, r_{lj}}^1 = 2\Gamma_{\{r_{li}, r_{ab}\}, r_{lj}}^0 \quad (a > l), \quad (27)$$

$$\Gamma_{\{I, J\}, K}^1 = 0 \quad (\text{for the other choices of } \{I, J\} \text{ and } K). \quad (28)$$

Those of the m-connection are

$$\Gamma_{\{\mu_i, \mu_j\}, \sigma_k}^{(-1)} = 2\Gamma_{\{\mu_i, \mu_j\}, \sigma_j'}^0 \quad (29)$$

$$\Gamma_{\{\sigma_i, \sigma_j\}, \sigma_i}^{(-1)} = -\Gamma_{\{\sigma_i, \sigma_j\}, \sigma_i'}^0 \quad (30)$$

$$\Gamma_{\{\mu_i, \mu_j\}, r_{ab}}^{(-1)} = 2\Gamma_{\{\mu_i, \mu_j\}, r_{ab}'}^0 \quad (31)$$

$$\Gamma_{\{r_{li}, \sigma_l\}, r_{lj}}^{(-1)} = 2\Gamma_{\{r_{li}, \sigma_l\}, r_{lj}'}^0 \quad (32)$$

$$\Gamma_{\{r_{li}, r_{lj}\}, \sigma_k}^{(-1)} = 2\Gamma_{\{r_{li}, r_{lj}\}, \sigma_k}^0 \quad (k \geq i, j), \quad (33)$$

$$\Gamma_{\{r_{li}, r_{lj}\}, r_{ab}}^{(-1)} = 2\Gamma_{\{r_{li}, r_{lj}\}, r_{ab}}^0 \quad (a > l), \quad (34)$$

$$\Gamma_{\{I, J\}, K}^{(-1)} = 0 \quad (\text{for the other choices of } \{I, J\} \text{ and } K). \quad (35)$$

There is a particular system of coordinates for describing the e-connection. Namely, all the coefficients vanish with respect to the natural parameter $(C^{-1}C^{-T}\mu, \xi)$, where $\xi = (\xi_{ab})_{1 \leq a \leq b \leq n}$ is the upper half of $C^{-1}C^{-T}$. On the other hand, all the coefficients for the m-connection vanish with respect to the expectation parameter (μ, ν) , where $\nu = (\nu_{ab})_{1 \leq a \leq b \leq n}$ is the upper half of $C^T C + \mu\mu^T$.

2.5. The Generalization

This subsection is devoted to the generalization of the result of the author, which is mentioned in the introduction, to the above multivariate setting. We fix the third component r of the coordinate system (μ, σ, r) , and change the presentation of the others. Namely, we take the natural projection $\pi : U = \mathbb{H}^n \times \mathbb{R}^{n(n-1)/2} \rightarrow \mathbb{R}^{n(n-1)/2}$ and replace the coordinates (μ, σ) on the fiber $L(r) = \pi^{-1}(r)$ by (m, s) appearing in the next proposition. The generalization is then straightforward.

Proposition 1. *The fiber $L(r) = \pi^{-1}(r)$ is an affine subspace of U with respect to the e-connection ∇^1 . It can be parameterized by affine parameters $\frac{m_i}{s_i^2}$ and $\frac{1}{s_i^2}$, where $m = [r^{ij}]^T \mu$ and $s = \sqrt{2}\sigma$.*

Proof. The natural parameters $C^{-1}C^{-T}\mu = \left(2 \sum_{k=i}^n r^{ik} \frac{m_k}{s_k^2}\right)_{1 \leq i \leq n}$ and $\xi = \left(-\sum_{k=b}^n r^{ak} r^{bk} \frac{1}{s_k^2}\right)_{1 \leq a \leq b \leq n}$ are affine provided that r^{ij} are constant, and $\frac{m_i}{s_i^2}$ and $\frac{1}{s_i^2}$ are affine. $\square$

Note that $dr(\nabla_{\partial_\mu}^\alpha \partial_\mu)$ is identically zero on some/any fiber $L(r)$ if and only if $\alpha = 1$. The fiber satisfies the following two properties.

Proposition 2. *$L(r)$ is closed under the convolution $*$ and the Bayesian product $\cdot$, and thus inherits them.*

Proof. The covariance of the density at $(\mu, \sigma, r) * (\mu', \sigma', r)$ is

$$[r_{ij}]^T \text{diag}((\sigma_i^2)) [r_{ij}] + [r_{ij}]^T \text{diag}((\sigma_i'^2)) [r_{ij}] = [r_{ij}]^T \text{diag}((\sigma_i^2 + \sigma_i'^2)) [r_{ij}]. \quad (36)$$

This coincides with that of the density at $\left(\mu + \mu', \left(\sqrt{\sigma_i^2 + \sigma_i'^2}\right), r\right)$. Thus $L(r)$ is closed under $*$, and inherits it as $(m, s) * (m', s') = \left(m + m', \left(\sqrt{s_i^2 + s_i'^2}\right)\right)$.

Put $u = (\sigma_i^{-2})$ and $y = [r^{ij}]^T x$. The density at (μ, σ, r) is proportional to $\exp\left(y(x)^T \text{diag}(u)y(x) - 2m^T \text{diag}(u)y(x)\right)$. From this we see that $(\mu, \sigma, r) \cdot (\mu', \sigma', r) = (\mu'', \sigma'', r'')$ implies $r'' = r$, $u'' = u + u'$ and $\text{diag}(u'')m'' = \text{diag}(u)m + \text{diag}(u')m'$. Thus $L(r)$ is closed under $\cdot$, and inherits it as $(m, s) \cdot (m', s') = \left(\frac{m_i s_i'^2 + m_i' s_i^2}{s_i^2 + s_i'^2}, \frac{s_i s_i'}{\sqrt{s_i^2 + s_i'^2}}\right)$. $\square$

Proposition 3. The fiber $L(r)$ with the induced metric from g admits a Kähler complex structure.

Proof. The restriction of g is $2 \sum_{i=1}^n \frac{dm_i^2 + ds_i^2}{s_i^2}$. We define the complex structure $J : TL(r) \rightarrow TL(r)$ by $J(\partial_{m_i}) = \partial_{s_i} (\Leftrightarrow J(\partial_{s_i}) = -\partial_{m_i})$. Then, the 2-form $\omega = 2 \sum_{i=1}^n \frac{dm_i \wedge ds_i}{s_i^2}$ satisfies $\omega(\cdot, J\cdot) = g(\cdot, \cdot)$. $\square$

We write the restriction $D|_{L(r)}$ of the premetric D using the coordinates (m, s) as follows, where we omit r for the expression does not depend on r .

$$D|_L((m, s), (m', s')) = \frac{1}{2} \sum_{i=1}^n \left\{ \left(\frac{m_i'}{s_i'} - \frac{m_i}{s_i} \right)^2 + \frac{s_i^2}{s_i'^2} - 1 - \ln \frac{s_i^2}{s_i'^2} \right\}. \quad (37)$$

We take the product $U_1 \times U_2$ of two copies of U . Then, the products $L_1(r) \times L_2(R)$ of the fibers foliate $U_1 \times U_2$. We call this the *primary* foliation of $U_1 \times U_2$. For each $(r, R) \in \mathbb{R}^{n(n-1)}$, we have the coordinate system (m, s, M, S) on the leaf $L_1(r) \times L_2(R)$. From the Kähler forms $\omega_1 = 2 \sum_{i=1}^n \frac{dm_i \wedge ds_i}{s_i^2}$ and $\omega_2 = 2 \sum_{i=1}^n \frac{dM_i \wedge dS_i}{S_i^2}$ respectively on $L_1(r)$ and $L_2(R)$, we define the symplectic forms $\omega_1 \pm \omega_2$ on $L_1(r) \times L_2(R)$, which induce the mutually opposite orientations in the case where n is odd. Hereafter, we consider the pair of regular Poisson structures defined by these symplectic structures on the primary foliation, and fix the primitive 1-forms $\lambda_{\pm} = 2 \sum_{i=1}^n \left(\frac{dm_i}{s_i} \pm \frac{dM_i}{S_i} \right)$. The corresponding pair of Poisson bi-vectors is $\Pi_{\pm} = \frac{1}{2} \sum_{i=1}^n \left(s_i^2 \partial_{m_i} \wedge \partial_{s_i} \pm S_i^2 \partial_{M_i} \wedge \partial_{S_i} \right)$ defined on $U_1 \times U_2$. We take the $2n$ -dimensional submanifolds $F_{\epsilon, \delta} = \left\{ \frac{m_i}{s_i} + \frac{M_i - \epsilon_i}{S_i} = 0, s_i S_i = \delta_i (i = 1, \dots, n) \right\}$ of the leaf $L_1(r) \times L_2(R)$ for $\epsilon \in \mathbb{R}^n$ and $\delta \in (\mathbb{R}_{>0})^n$. The *secondary* foliation of $U_1 \times U_2$ foliates any leaf $U(r) \times U(R)$ by the $3n$ -dimensional submanifolds $F_{\epsilon} = \bigcup_{\delta \in (\mathbb{R}_{>0})^n} F_{\epsilon, \delta}$ for $\epsilon \in \mathbb{R}^n$. The *tertiary* foliation of $U_1 \times U_2$ foliates all leaves F_{ϵ} of the secondary foliation by the $2n$ -dimensional submanifolds $F_{\epsilon, \delta}$ for $\delta \in (\mathbb{R}_{>0})^n$.

Proposition 4. With respect to the Kähler form $d\lambda_{-}$, the tertiary leaves $F_{\epsilon, \delta}$ are Lagrangian correspondences.

Proof. The tangent space $TF_{\epsilon, \delta}$ is generated by $s_i \partial_{m_i} - S_i \partial_{M_i}$ and $2m_i \partial_{m_i} + s_i \partial_{s_i} - S_i \partial_{S_i}$. We have $d\lambda_{-}(s_i \partial_{m_i} - S_i \partial_{M_i}, \partial_{m_i}) = 0$ and $d\lambda_{-}(s_i \partial_{m_i} - S_i \partial_{M_i}, s_i \partial_{s_i} - S_i \partial_{S_i}) = 2 \left(\frac{s_i^2}{s_i^2} - \frac{(-S_i)^2}{S_i^2} \right) = 0$. Thus $F_{\epsilon, \delta}$ are Lagrangian submanifolds of $(L_1(r) \times L_2(R), d\lambda_{-})$. $\square$

The restrictions of $\lambda_{\pm}|_N$ to the hypersurface $N = \left\{ (m, s, M, S) \in L_1 \times L_2 \mid \prod_{i=1}^n (s_i S_i) = 1 \right\}$ are contact forms. Let $\eta_{\pm}$ denote them.

Proposition 5. For any ε and δ with $\prod_{i=1}^n \delta_i = 1$, the submanifold $F_{\varepsilon, \delta} \subset N$ is a disjoint union of n -dimensional submanifolds $\{s = \text{const}\} \subset F_{\varepsilon, \delta}$ which are integral submanifolds of the hyper plane field $\ker \eta_+$ on N .

Proof. We have $\lambda_+(s_i \partial_{m_i} - S_i \partial_{M_i}) = 1 - 1 = 0$. $\square$

For each point $(\varepsilon, \delta) \in \mathbb{H}^n$, we have the diffeomorphism $\hat{F}_{\varepsilon, \delta} : \mathbb{H}^n \rightarrow \mathbb{H}^n$ sending $(m', s') \in \mathbb{H}^n$ to $(M, S) \in \mathbb{H}^n$ with $(m', s', M, S) \in F_{\varepsilon, \delta}$. We put $h_i = -\ln \frac{s_i S_i}{\delta_i}$ and

$$f_{\varepsilon, \delta}(m, s, M, S) = \frac{1}{2} \sum_{i=1}^n \left\{ \left(\frac{M_i - \varepsilon_i}{S_i} + e^{-h_i} \frac{m_i}{s_i} \right)^2 + e^{-2h_i} - 1 + 2h_i \right\}. \quad (38)$$

Then, we have $D|_L((m, s), (m', s')) = f_{\varepsilon, \delta}((m, s), \hat{F}_{\varepsilon, \delta}(m', s'))$. For any $\zeta \in (\mathbb{R}_{>0})^{2n}$, we define the diffeomorphism $\varphi_{\varepsilon, \zeta} : (m, s, M, S) \mapsto (\zeta_{2i-1} m_i, (\zeta_{2i-1} s_i), (\varepsilon_i + \zeta_{2i}(M_i - \varepsilon_i)), (\zeta_{2i} S_i))$, which preserves the 1-forms $\lambda_{\pm}$. It is easy to prove the following proposition.

Proposition 6. In the case where $\zeta_{2i-1} \zeta_{2i} = 1$ for $i = 1, \dots, n$, the diffeomorphism $\varphi_{\varepsilon, \zeta}$ preserves $f_{\varepsilon, \delta}$.

For each $\varepsilon \in \mathbb{R}^n$, we take the set $f_{\varepsilon} = \{(f_{\varepsilon, \delta}, F_{\varepsilon, \delta}) \mid \delta \in (\mathbb{R}_{>0})^n\}$, and consider it as a structure of the secondary leaf F_{ε} .

Proposition 7. For any $\zeta \in (\mathbb{R}_{>0})^n$, the diffeomorphism $\varphi_{\varepsilon, \zeta}$ preserves the set f_{ε} for any $\varepsilon \in \mathbb{R}^n$. In the case where ζ satisfies $\prod_{i=1}^n (\zeta_{2i-1} \zeta_{2i}) = 1$, the diffeomorphism $\varphi_{\varepsilon, \zeta}$ also preserves the hypersurface N .

Proof. $F_{\varepsilon, \delta}$ maps to $F_{\varepsilon, (\zeta_{2i-1} \zeta_{2i} \delta_i)}$, and $f_{\varepsilon, (\zeta_{2i-1} \zeta_{2i} \delta_i)} \circ \varphi_{\varepsilon, \zeta} = f_{\varepsilon, \delta}$ holds. Further $\prod_{i=1}^n \delta_i = 1$ implies $\prod_{i=1}^n (\zeta_{2i-1} \zeta_{2i} \delta_i) = 1$ provided that $\prod_{i=1}^n (\zeta_{2i-1} \zeta_{2i}) = 1$. $\square$

For any $\delta \in (\mathbb{R}_{>0})^n$, the diffeomorphism $\hat{F}_{0, \delta}$ interchanges the operation $(m, s) * (m', s') = \left(m + m', \left(\sqrt{s_i^2 + s_i'^2} \right) \right)$ with the operation $(m, s) \cdot (m', s') = \left(\frac{m_i s_i'^2 + m_i' s_i^2}{s_i^2 + s_i'^2}, \frac{s_i s_i'}{\sqrt{s_i^2 + s_i'^2}} \right)$.

Proposition 8. If $(m, s, M, S), (m', s', M', S') \in F_{0, \delta}$, then

$$\begin{cases} ((m, s) \cdot (m', s'), (M, S) * (M', S')) \in F_{0, \delta} \\ ((m, s) * (m', s'), (M, S) \cdot (M', S')) \in F_{0, \delta} \end{cases}. \quad (39)$$

Proof. Putting $(m'', s'') = (m, s) \cdot (m', s')$ we have $M_i + M_i' = -\frac{\delta_i m_i}{s_i^2} - \frac{\delta_i m_i'}{s_i'^2} = -\frac{\delta_i m_i''}{s_i''^2}$ and $S_i^2 + S_i'^2 = \frac{\delta_i^2}{s_i^2} + \frac{\delta_i^2}{s_i'^2} = \frac{\delta_i^2}{s_i''^2}$, hence the first equation. The second equation is similarly proved. $\square$

A curve $(m(t), s(t)) \in \mathbb{H}^n$ is a geodesic with respect to the e-connection ∇^1 if and only if $\frac{m_i}{s_i^2}$ and $\frac{1}{s_i^2}$ are affine functions of t for $i = 1, \dots, n$.

Definition 1. We say that an e-geodesic $(m(t), s(t)) \in \mathbb{H}^n$ is intensive if it admits an affine parameterization such that $\frac{1}{s_i^2}$ are linear for $i = 1, \dots, n$.

Note that any e-geodesic is intensive in the case where $n = 1$.

Proposition 9. Given an intensive e-geodesic $(m(t), s(t))$, we can change the parameterization of its image $(M(t), S(t)) = \left(\left(\varepsilon_i - \frac{m_i(t)\delta_i}{s_i^2} \right), \left(\frac{\delta_i}{s_i} \right) \right)$ under the diffeomorphism $\hat{F}_{\varepsilon, \delta}$ to obtain an intensive e-geodesic.

Proof. Put $\frac{m_i}{s_i^2} = a_i t + b_i$ and $\frac{1}{s_i^2} = c_i t$. We have $\frac{M_i - \varepsilon}{S_i^2} = -\frac{a_i}{c_i \delta_i} - \frac{b_i}{c_i \delta_i} \cdot \frac{1}{t}$ and $\frac{1}{S_i^2} = \frac{1}{c_i \delta_i^2} \cdot \frac{1}{t}$. They are respectively an affine function and a linear function of $\frac{1}{t}$. $\square$

We have the hypersurface $N = \left\{ \prod_{i=1}^n s_i S_i = 1 \right\} \subset \mathbb{H}^n$ which carries the contact forms $\eta_{\pm} = 2 \sum_{i=1}^n \left(\frac{dm_i}{s_i} \pm \frac{dM_i}{S_i} \right) \Big|_N$. This is defined on any leaf $L_1(r) \times L_2(R) \approx \mathbb{H}^{2n}$ of the primary foliation of $U_1 \times U_2$. Now we state the main result.

Theorem 1. The contact Hamiltonian vector field Y of the restriction of the function $\sum_{i=1}^n \frac{m_i}{s_i}$ to N for the contact form η_+ coincides with that for the other contact form η_- . It is tangent to the tertiary leaves $F_{\varepsilon, \delta}$ and defines flows on them. Here each flow line presents the correspondence between intensive e-geodesics in Proposition 9.

Corollary 1. For any $\delta \in (\mathbb{R}_{>0})^n$, the flow on the leaf $F_{0, \delta}$ presents the iteration of the operation $*$ on the first factor of $U_1 \times U_2$ and that of the operation $\cdot$ on the second factor as is described in Proposition 8 (see Figure 1).

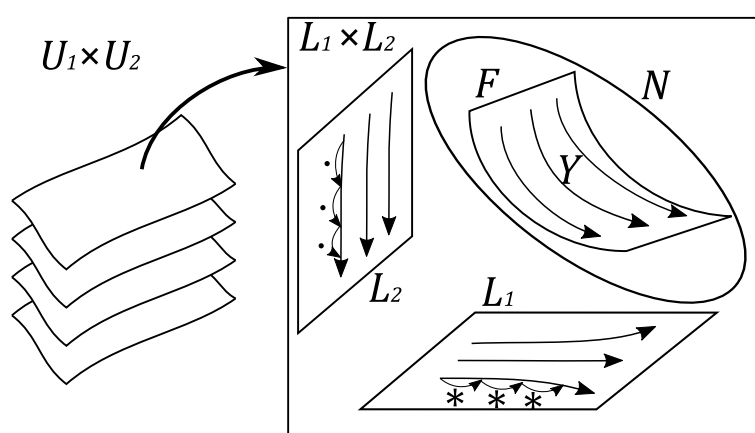


Figure 1. On any leaf of the primary foliation of $U_1 \times U_2$, there is a bi-contact hypersurface N carrying the bi-contact Hamiltonian vector field Y . Because of the dimension, the surface F in the figure presents simultaneously a leaf of the secondary foliation and a leaf of the tertiary foliation of that leaf. The flow on the tertiary leaf $F = F_{0, \delta}$ traces the common lift of the iteration of $*$ on L_1 and the iteration of $\cdot$ on L_2 .

From Corollary 1, we see that the flow model of Bayesian inference studied in [4,5] also works for the multivariate case. We prove the theorem.

Proof. Take the vector field $\tilde{Y} = \frac{1}{4} \sum_{i=1}^n (2m_i \partial_{m_i} + s_i \partial_{s_i} - S_i \partial_{S_i})$ on $U_1 \times U_1$. It satisfies $\iota_{\tilde{Y}} \lambda_{\pm} = \sum_{i=1}^n \frac{m_i}{s_i}$, $\tilde{\lambda}_Y \lambda_{\pm} = \lambda_{\pm}$, and $\iota_{\tilde{Y}} d \ln(s_i S_i) = 0$, and thus its restriction to N is the contact Hamiltonian vector field Y . It also satisfies $\iota_{\tilde{Y}} d \left(\frac{m_i}{s_i} + \frac{M_i - \varepsilon_i}{S_i} \right) = \frac{1}{4} \left(\frac{m_i}{s_i} + \frac{M_i - \varepsilon_i}{S_i} \right)$ ($i = 1, \dots, n$), where the right-hand sides vanish along $F_{\varepsilon, \delta}$. Given a point $(P_0, Q_0) = (m_0, s_0, r_0, M_0, S_0, R_0) \in N \subset L_1(r_0) \times L_2(R_0) \subset U \times U$, we have the integral curve $(m(t), s(t), r(t), M(t), S(t), R(t)) = (e^{2t} m_0, e^t s_0, r_0, M_0, e^{-t} S_0, R_0)$ of Y with initial point (P_0, Q_0) . We can change the parameter of the curve $(m(t), s(t), r_0)$ on the first factor with $t' = e^{-2t}$ to obtain an intensive e-geodesic. $\square$

2.6. The Symmetry

The diffeomorphism $\varphi_{\varepsilon, \zeta}$ with $\prod_{i=1}^n (\zeta_{2i-1} \zeta_{2i}) = 1$ in Proposition 7 also appears in the standard construction of Hilbert modular cusps in [8]. We sketch the construction.

Since the function $H = \ln \prod_{i=1}^n (s_i S_i)$ is linear in the logarithmic space $\mathbb{R}^{2n} \ni ((\ln s_i), (\ln S_i))$, we can take $2n - 1$ points $((\ln \zeta_{2i-1}(k)), (\ln \zeta_{2i}(k))) \in \{H = 0\}$ ($k = 1, \dots, 2n - 1$) so that the quotient of the primary leaf $L(r) \times L(R)$ under the $\mathbb{Z}^{2n-1}$ -action generated by $\varphi_{\varepsilon, \zeta(1)}, \dots, \varphi_{\varepsilon, \zeta(2n-1)}$ is the total space of a vector bundle over $T^{2n-1} \times \mathbb{R}$. Here T^{2n-1} is the quotient of $\{H = 0\}$ under the $\mathbb{Z}^{2n-1}$ -lattice generated by the above points, and the fiber $\mathbb{R}^{2n}$ consists of the vectors $(m, M - \varepsilon)$.

In the univariate case ($n = 1$), on the logarithmic sS -plane, we can take any point $(\ln s, \ln S) = (\ln \zeta_1(1), \ln \zeta_2(1))$ of the line $H = \ln s + \ln S = 0$ other than the origin. That is, we put $\zeta_1(1) = e^t$ and $\zeta_2(1) = e^{-t}$ ($t \neq 0$). Then, the map $\varphi_{\varepsilon, \zeta}$ sends (m, s, M, S) to $(e^t m, e^t s, e^{-t}(M - \varepsilon), e^{-t} S)$. The $\mathbb{Z}$ -action generated by it rolls up the level sets of H , so that the quotient of the logarithmic sS -plane becomes the cylinder $T^1 \times \mathbb{R}$, which is the base space. On the other hand, the mM -plane, which is the fiber, expands horizontally and contracts vertically. This is the inverse monodromy along T^1 . In general, we obtain a similar $\mathbb{R}^{2n}$ -bundle over $T^{2n-1} \times \mathbb{R}$ if we take the $2n - 1$ points of $H = 0$ in general position.

From Proposition 7, we see that the leaf F_{ε} of the secondary foliation with the set f_{ε} , as well as the pair of the 1-forms $\lambda_{\pm}$ with the function H , descends to the $\mathbb{R}^{2n}$ -bundle. If there exists further a $\mathbb{Z}^{2n}$ -lattice on the fiber $\mathbb{R}^{2n}$ which is simultaneously preserved by the maps $((m_i), (M_i - \varepsilon_i)) \mapsto ((\zeta_{2i-1}(k)m_i), (\zeta_{2i}(k)(M_i - \varepsilon_i)))$ ($k = 1, \dots, 2n - 1$), we obtain a T^{2n} -bundle over $T^{2n-1} \times \mathbb{R}$. Such a choice of $\zeta(k)$ would be number theoretical. Indeed, this is the case for Hilbert modular cusps. Moreover, we are considering the 1-forms $\lambda_{\pm}$, which descend to the T^{2n} -bundle. See [9] for the standard construction with special attention to these 1-forms.

We should notice that the vector field Y does not descend to the T^{2n} -bundle. However, every actual Bayesian inference along Y eventually stops. Thus, we may take sufficiently large T^{2n} and consider Y as a locally supported vector field to perform the inference in the quotient space.

3. Discussion

Finally, we would like to comment on the transverse geometry of the primary foliation. The author conjectures that it has some relation to the M-theory. See e.g., [10] for a relation between Poisson geometry and matrix theoretical or non-commutative geometrical physics.

The premetric $D((\mu + \Delta\mu, \sigma + \Delta\sigma, r + \Delta r), (\mu, \sigma, r))$ on U can be decomposed as

$$D((\mu + \Delta\mu, \sigma + \Delta\sigma, r), (\mu, \sigma, r)) + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(\frac{\sigma_i + \Delta\sigma_i}{\sigma_j} \sum_{k=i+1}^j r^{kj} \Delta r_{ik} \right)^2. \quad (40)$$

The first term on the right-hand side presents the fiber premetric $D|_L((m + \Delta m, s + \Delta s), (m, s))$, where $\Delta m = [r^{ij}]^T \Delta \mu$ and $\Delta s = \sqrt{2} \Delta \sigma$. If $\Delta \sigma = 0 (\Leftrightarrow \Delta s = 0)$, then we have the (non-information geometrical) Pythagorean-type formula

$$D((\mu + \Delta \mu, \sigma, r + \Delta r), (\mu, \sigma, r)) = D((\mu + \Delta \mu, \sigma, r), (\mu, \sigma, r)) + D((\mu, \sigma, r + \Delta r), (\mu, \sigma, r)). \quad (41)$$

Note that each term in this expression does not depend on μ . The second term on the right-hand side can be expressed as

$$D((\mu, \sigma, r + \Delta r), (\mu, \sigma, r)) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(\frac{\sigma_i}{\sigma_j} \sum_{k=i+1}^j r^{kj} \Delta r_{ik} \right)^2. \quad (42)$$

This presents the discretized version of the following restriction of the Fisher information g :

$$g|_{\{(\mu, \sigma) = \text{const}\}} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(\frac{\sigma_i}{\sigma_j} \sum_{k=i+1}^j r^{kj} dr_{ik} \right)^2. \quad (43)$$

We have the orthonormal frame with respect to this metric which consists of

$$e_{ij} = \frac{\sigma_i}{\sigma_j} \sum_{k=j}^n r_{jk} \partial_{r_{ik}} \quad (1 \leq i < j \leq n). \quad (44)$$

This frame satisfies the relations $[e_{ij}, e_{kl}] = \delta_{il} e_{kj} - \delta_{kj} e_{il}$ of the unitriangular algebra, where $\delta_{\cdot, \cdot}$ denotes Kronecker's delta. Using the dual coframe e^{ij} , the relations can be expressed as

$$de^{ij} = \sum_{k=i+1}^{j-1} e^{ik} \wedge e^{kj}. \quad (45)$$

The transverse section of the primary foliation of $U_1 \times U_2$ is the product of two copies of the unitriangular Lie group, which we would like to call the bi-unitriangular group. We fix the frame (respectively, the coframe) of the transverse section consisting of the above e_{ij} (respectively, e^{ij}) in the first factor U_1 and their copies E_{ij} (respectively, E^{ij}) in the second factor U_2 . The quotient manifold of the bi-unitriangular group by a cocompact lattice inherits a Riemannian metric from the sum of Fisher informations, and carries the following global $(n-2)$ -plectic structure Ω ($d\Omega = 0, \Omega^n > 0$):

$$\Omega = \sum_{i=1}^n e^{i,i+1} \wedge \dots \wedge e^{i,n} \wedge E^{n-i+1,n-i+2} \wedge \dots \wedge E^{n-i+1,n}. \quad (46)$$

We notice that, in the symplectic case where $n = 3$, the quotient manifold admits no Kähler structure (see [11]). However, it is still remarkable that the transverse symplectic 6-manifold is naturally ignored in the Bayesian inference described in this paper. Conjecturally, a similar model would help us to treat events in parallel worlds (or planes) in the same “psychological” procedure.

Funding: This research received no external funding.

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

References

1. Amari, S. *Information Geometry and Its Applications*; Springer: Tokyo, Japan, 2016.
2. Ay, N.; Jost, J.; Le, H.V.; Schwachhöfer, L. *Information Geometry*; Springer: Gewerbesrasse, Switzerland, 2017.
3. Felice, N.; Ay, N. Towards a Canonical Divergence within Information Geometry. *arXiv* **2018**, arXiv:1806.11363.
4. Mori, A. Information geometry in a global setting. *Hiroshima Math. J.* **2018**, *48*, 291–306. [[CrossRef](#)]

5. Mori, A. A concurrence theorem for alpha-connections on the space of t-distributions and its application. *Hokkaido Math. J.* **2020**, in press.
6. Geiges, H. *An Introduction to Contact Topology*; Cambridge University Press: Cambridge, UK, 2008.
7. Snoussi, H. Bayesian information geometry: Application to prior selection on statistical manifolds. In *Advances in Imaging and Electron Physics 146*; Hawkes, P., Ed.; Elsevier: San Diego, CA, USA, 2007; pp. 163–207.
8. Hirzebruch, F. Hilbert modular surfaces. *Enseign. Math.* **1973**, *19*, 183–281.
9. Massot, P.; Niederkrüger, K.; Wendl, C. Weak and strong fillability of higher dimensional contact manifolds. *Invent. Math.* **2013**, *192*, 287–373. [[CrossRef](#)]
10. Kuntner, N.; Steinacker, H. On Poisson geometries related to noncommutative emergent gravity. *J. Geom. Phys.* **2012**, *62*, 1760–1777. [[CrossRef](#)]
11. Cordero, L.; Fernández, M.; Gray, A. Symplectic manifolds with no Kähler structure. *Topology* **1986**, *25*, 375–380. [[CrossRef](#)]



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).