BAYESIAN ROBUSTNESS: A NEW LOOK FROM GEOMETRY

Carlos C. Rodríguez
Department of Mathematics and Statistics
State University of New York at Albany
Albany NY 12222, USA
Email: carlos@math.albany.edu

ABSTRACT. The geometric concept of the Lie derivative is introduced as the natural way of quantifying the intrinsic robustness of a hypothesis space. Prior and posterior probability measures are interpreted as differential forms defined invariantly on the hypothesis space. Rates of change with respect to local deformations of the model are computed by means of Lie derivatives of tensors defined on the model (like the Information metric, prior, posterior, etc.). In this way a field theory of inference is obtained. The class of deformations preserving the state of total ignorance is introduced and characterized by a partial differential equation. For location models this equation is the familiar $\nabla \cdot \xi = 0$. A simple condition for the robustness of prior (or posterior) distributions is found: There is robustness when the deformation is along level surfaces of the prior (or posterior) density. These results are then applied to the class of entropic priors. It is shown that the hyper parameter controls the sensitivity with respect to local deformations. It is also shown that entropic priors are only sensitive to deformations that change the intrinsic form of the model around the initial guess.

1. Introduction

The Robustness, of a statistical procedure, is commonly defined as the stability with respect to small changes in the assumptions. This notion has immediate intuitive appeal and it has even been equated to the Holy Grail of Statistics (see [5]).

There is general agreement about the desirability of a consistent theory of Statistical Robustness (Bayesian and non-Bayesian) and the large number of articles and books dedicated to the subject testify it. The technical definition of robustness is still controversial, however. For a serious criticism to the definitions of Hampel [3, p. 1980] and Huber [4, p. 10] see [6, p. 17].

In this paper, the geometric concept of the Lie derivative is introduced as a technical tool for quantifying robustness. The great arsenal of modern geometry tools provide a flexible, rigorous, and powerful framework for developing Statistical inference in general and Bayesian robustness in particular.

The geometrization of statistics is possible in part due to the fact that statistical models have a natural manifold structure. Fisher information endows the models with a Riemannian metric and the Kullback number (entropy) generates this and many other natural geodesic metrics on the model (see [1] and [7]).

The main idea is to exploit the rich geometric structure available in the hypothesis space for the quantification of robustness. Once differential geometry is permitted to be the operational framework, a number of consequences for robustness are straight forward and inevitable. This paper concentrates on the quantification of the robustness of probability distributions defined on the model. The same technique can be used for quantifying the stability of any tensorial quantity defined on the space.

In this approach, there are important differences with traditional methods. First, everything is *intrinsic* to the model. There is no need for postulating super models, nonparametric neighborhoods, or anything outside the given hypothesis space. The model is an enclosed universe that is assumed to include *all* the relevant probability measures for the observed data. The possibility of encoding deformations of the model without reference to an outside, is a remarkable achievement of modern geometry. Second, n does not have to go to infinity for the methods to make sense. In fact they even make sense in the absence of all data. We can quantify the sensitivity of a prior distribution with respect to deformations of the model independently of the observations.

The paper is divided into four sections. In section one we introduce the notation and provide a summary of the main definitions and results from geometry that will be needed later. In section two we introduce probability distributions over the parameter space as differential forms defined on the model and their Lie derivatives are computed by using the methods of section one. We also compute explicitly the sensitivity of the class of entropic priors with respect to deformations of the model. Finally, in section three, we workout the example of inference in the one dimensional gaussian model with entropic priors. We conclude in section four with some comments on the possible future developments of these methods.

2. Local Deformations, Lie Derivatives of Tensors and Volume Elements

We collect here some classical results from the geometry of vector fields on manifolds. The material in this section can be found on most books on modern geometry. We follow the presentation and notation of [2, chap. 23].

Regular (finite dimensional) parametric statistical models will be denoted by $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. They are Riemannian manifolds. The parameterization $\Theta \subset \Re^k$ plays the role of a coordinate system. The tangent space at $P \in \mathcal{P}$ is modeled by the linear space generated by the partial derivatives (w.r. to θ) of the log-likelihoods. In this way the tangent space at P is a subspace of $L^2(P)$ and it inherits the inner product from it. It turns out that the Riemannian metric on the tangent space at P_θ, $g_{ij}(\theta)$, coincides with the Fisher information matrix at θ, see [1] and [7] for detailed definitions.

A vector field ξ on the manifold $\mathcal{P}$ is a mapping that assigns to each $P \in \mathcal{P}$ a tangent vector at P. For a pictorial representation think of the model as a k-dimensional (curved) surface and the vector field as the velocity field of a fluid moving on the surface. If the field ξ is smooth (as a map between manifolds) the theory of ordinary differential equations warranties the existence and uniqueness of the following associated autonomous system of differential equations:

$$\begin{aligned} \frac{d\theta^i}{dt} &= \xi^i\left(\theta^1(t), \ldots, \theta^k(t)\right), \quad i = 1, \ldots, k \\ \theta^i|_{t=t_0} &= \theta^i_0 \end{aligned} \tag{1}$$

where θ^i and ξ^i denote the components of P_θ and ξ in the coordinate system Θ and θ_0 is the initial condition. The solution to this system is known as the integral curve of ξ passing through P_{θ_0}. We denote it by $F_t(\theta_0) = \theta(t)$. For a given t the map, $F_t : \theta_0 \mapsto \theta(t)$, defined in a neighborhood of the point θ_0, represents the new position after time t of a particle of fluid which is initially at θ_0. The theory of ordinary differential equations assures that for t small enough the maps F_t are diffeomorphisms (i.e. one to one and with continuous differential both ways). More precisely they form a local one parameter group of diffeomorphisms with group operation $F_t \circ F_s = F_{t+s}$, inverse $F_t^{-1} = F_{-t}$ and identity F_0. Each transformation F_t defines (at least locally) a change of coordinates from θ_0 to $\theta(t)$ i.e. a local re-labeling of the elements of $\mathcal{P}$. In this way if T is a quantity defined in terms of the labels $\theta(t)$ (but providing intrinsic information about the points $P \in \mathcal{P}$) it will have an expression $(F_t T)$ in terms of the labels θ_0, satisfying the rules of transformation for tensors. We have the following:

Definition 1 *The Lie derivative of a tensor T along a vector field ξ is the tensor $L_\xi T$ given by*

$$L_\xi T = \left[\frac{d}{dt}(F_t T)\right]_{t=0} \tag{2}$$

Again, in visual terms the Lie derivative of T along ξ gives the rate of change of T as it is seen when moving with the fluid. Or equivalently, standing at θ_0 we see the components of T change due to the (time dependent) deformation of the space given by F_t and the Lie derivative is just the rate of change (with respect to time) of what we see. By applying the rules of transformations for tensors and using the smoothness of the F_t's (by Taylor's theorem $F_t(\theta_0) = \theta_0 + t\xi(\theta_0) + o(t)$), we can find the components of the Lie derivative of the tensor $T^{i_1\cdots i_p}_{j_1\cdots j_q}$. They are given by:

$$\begin{aligned} L_\xi T^{i_1\cdots i_p}_{j_1\cdots j_q} &= \xi^s \frac{\partial T^{i_1\cdots i_p}_{j_1\cdots j_q}}{\partial \theta^s} + T^{i_1\cdots i_p}_{k j_2\cdots j_q}\frac{\partial \xi^k}{\partial \theta^{j_1}} + \ldots + T^{i_1\cdots i_p}_{j_1\cdots j_{q-1}k}\frac{\partial \xi^k}{\partial \theta^{j_q}} \\ &\quad - T^{l i_2\cdots i_p}_{j_1\cdots j_q}\frac{\partial \xi^{i_1}}{\partial \theta^l} - \ldots - T^{i_2\cdots i_{p-1} l}_{j_1\cdots j_q}\frac{\partial \xi^{i_p}}{\partial \theta^l}. \end{aligned} \tag{3}$$

where here, as in the rest of the paper, the standard implicit summation over repeated indices is assumed. As special cases of this formula we have:

I). **Scalar field**. $T = f$

$$L_\xi f = \xi \cdot \nabla f = \xi^i \frac{\partial f}{\partial \theta^i} \tag{4}$$

II). **Vector field**. $T = \eta^i$

$$L_\xi \eta^i = [\xi, \eta]^i = \xi^j \frac{\partial \eta^i}{\partial \theta^j} - \eta^j \frac{\partial \xi^i}{\partial \theta^j} \tag{5}$$

where $[\xi, \eta]$ denotes the commutator between the vector fields.

III). **Covector field**. $T = T_j = \frac{\partial f}{\partial \theta^j}$

$$\begin{aligned} (L_\xi T)_j &= \xi^k \frac{\partial T_j}{\partial \theta^k} + T_k \frac{\partial \xi^k}{\partial \theta^j} \\ &= d\,(L_\xi f)_j = L_\xi (df)_j \end{aligned} \tag{6}$$

i.e. Lie derivatives commute with differentials.

IV). **Bilinear form**. $T = g_{ij}$

$$L_\xi g_{ij} = \xi^s \frac{\partial g_{ij}}{\partial \theta^s} + g_{kj} \frac{\partial \xi^k}{\partial \theta^i} + g_{ik} \frac{\partial \xi^k}{\partial \theta^j} = u_{ij} \tag{7}$$

This is known as the strain tensor.

V). **Volume element**.

$$\begin{aligned} T = T_{i_1 \ldots i_k} &= \sqrt{|g|}\, \epsilon_{i_1 \ldots i_k} \\ &= \sqrt{|g|}\, d\theta^{i_1} \wedge \ldots \wedge d\theta^{i_k} \\ &= \pm\sqrt{|g|}\, d\theta^1 \wedge \ldots \wedge d\theta^k \quad \text{if no two indices are equal.} \end{aligned} \tag{8}$$

Where $\epsilon_{i_1 \ldots i_k}$ is the *Levi-Civita* tensor defined as $+1, -1, 0$ depending on the indices forming an even, odd or no permutation of the first k integers. We denote by $|g|$ the absolute value of the determinant of the metric tensor $g_{ij}(\theta)$.

The volume element plays a central role in Bayesian inference. Geometrically, it gives the *surface area* of a small patch on the k-dimensional surface. For this reason is the analogous of the lebesgue (uniform) measure on flat space. It behaves like a totally anti symmetric (skew-symmetric) tensor under coordinate transformations preserving a given orientation of the space. Volume elements can then be interpreted as differential forms of order k and as *total ignorance priors* in statistics. After some simplifications formula 3 gives,

$$L_\xi \left(\sqrt{|g|}\, d\theta^1 \wedge \ldots \wedge d\theta^k \right) = \frac{1}{2} g^{im} \left(L_\xi g_{im} \right) \sqrt{|g|}\, d\theta^1 \wedge \ldots \wedge d\theta^k \tag{9}$$

Where g^{im} denotes the inverse of the (Fisher information) matrix, g_{im}. Expressions involving g are always functions of θ but this will be kept implicit to simplify the notation. Notice that the effect of taking the Lie derivative of the volume element is to multiply it by one half the *trace* of the strain tensor defined in 7.

VI). **Leibniz' rule**. If T and R are arbitrary tensors and $T \otimes R$ denotes the tensor product between them, then,

$$L_\xi \left(T \otimes R \right) = \left(L_\xi T \right) \otimes R + T \otimes \left(L_\xi R \right) \tag{10}$$

3. Robustness of Probability Distributions Defined on the Parameter Space

Probability measures defined on Θ (e.g. priors and posteriors) can be seen as providing alternative ways of measuring the surface area of patches on the manifold. The parameterization Θ is only a convenient artifact to be able to write the formulas explicitly without having to perform the integration directly over the functional space of probability measures. But the parameterization is arbitrary and therefore it must be immaterial. The formulas should show this invariance under reparameterization up front.

From this point of view, it is necessary to leave tradition and move from the usual interpretation of probability densities as functions with transformation rules governed by the so called *change of variables theorem* to scalar fields with no transformation rules whatsoever. Traditionally, density functions are integrated with *integrals of the second kind* which are just multiple integrals to be handled independently of any metric which may be defined on the space. But if we move to densities as scalar fields, then they have to be integrated with *integrals of the first kind* with respect to the volume element in the desired parameterization. A probability measure on Θ will then be written as a differential form,

$$\pi(\theta)\sqrt{|g|}\, d\theta^1 \wedge \ldots \wedge d\theta^k \tag{11}$$

where $\pi(\theta)$ is a scalar field. Notice that $\pi(\theta)$ is just the Radon-Nikodym derivative of the probability measure defined by 11 with respect to the volume element measure. In other words the density (as a scalar field) is given relative to *total ignorance*. Notice also that 11 is wonderfully invariant. If what it was called θ we now call θ' all we need to do to 11 is to prime the θ's and we get the formula in the new coordinate system.

An example may help to fix the ideas. For example, $\frac{1}{2\pi}e^{-\frac{1}{2}r^2}$ is the probability density scalar field of the standard bivariate gaussian on the euclidean plane parameterized with polar coordinates r, θ. The same density in cartesian coordinates is just $\frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}$. i.e. the point on the euclidean plane with the two labels (x, y) and $[r, \theta]$ has exactly the same density relative to (euclidean) ignorance since $r^2 = x^2 + y^2$.

Noteworthy, this almost trivial change in point of view, helps to clarify an old puzzle of inference: *How come that complete ignorance about a value* $x \in [0, 1]$ *is not complete ignorance about* $y = x^2 \in [0, 1]$*?*. In other words, the *change of variable theorem* transforms the uniform density of x into the non uniform density $\frac{1}{2}y^{-1/2}$ for y. This is regarded as paradoxical, for, it is claimed, indifference about the number x should produce indifference about the number $y = x^2$. When considering *densities as scalar fields* there is no puzzle. The puzzle arises from the insistence, of the change of variables theorem, to keep the underlying measure to be the same (Lebesgue measure on $[0, 1]$ in this case) for x and for y. But, x and $y = x^2$ are just two different numerical labels for events (perhaps measurements of the same thing but in two systems of units) so whatever it was labeled $\frac{1}{2}$, say, with x is relabeled as $\frac{1}{4}$ by y. Therefore, the labels $x = 0.5$ and $y = 0.25$ must have the same chance of occurrence. In fact, they do. But the change of variables theorem hides it by shifting the Jacobian from the volume element, where it belongs, to the density, where it does not belong. Our formula 11, assigns constant density to the numbers in $[0, 1]$ in **all coordinate systems**, linear or non linear transformations of x.

Formula 11 is composed as the product of two invariants. The scalar field density and the volume element. Remember that the volume element is invariant under all reparameterizations preserving orientation. When changing coordinate systems, the two parts remain the same.

3.1. The Robustness of Total Ignorance

The rate of change of the total ignorance prior along a deformation of the model given by a vector field ξ is given by 9. Replacing 7 into 9 and simplifying, we can write

$$L_\xi\left(\sqrt{|g|}\, d\theta^1 \wedge \ldots \wedge d\theta^k\right) = \left(\frac{1}{2}g^{ij}\nabla g_{ij} + \nabla\right)\cdot \xi \left(\sqrt{|g|}\, d\theta^1 \wedge \ldots \wedge d\theta^k\right) \tag{12}$$

Thus, a local deformation of the hypothesis space, does not change the state of total ignorance if it is along a vector field ξ solving the partial differential equation:

$$\left(\frac{1}{2}g^{ij}\nabla g_{ij} + \nabla\right)\cdot\xi = 0 \tag{13}$$

Therefore, if the metric tensor is independent of θ (e.g. for location models), equation 13 reduces to the familiar:

$$\nabla\cdot\xi = 0 \tag{14}$$

Equation 14, and the more general equation, 13, encode the idea of *conservation of ignorance.* Deformations of the model that satisfy them, are precisely those that do not create nor destroy information.

3.2. Robustness of Priors and Posteriors

To compute the Lie derivative of an arbitrary distribution over the parameter space we apply Leibniz' rule 10, to the differential form 11,

$$\begin{aligned} L_\xi\left(\pi\sqrt{|g|}\,d\theta^1\wedge\ldots\wedge d\theta^k\right) &= \pi L_\xi\left(\sqrt{|g|}\,d\theta^1\wedge\ldots\wedge d\theta^k\right) + \left(\sqrt{|g|}\,d\theta^1\wedge\ldots\wedge d\theta^k\right)L_\xi\pi \\ &= \left(\frac{1}{2}g^{im}L_\xi g_{im} + L_\xi\log\pi\right)\pi\sqrt{|g|}\,d\theta^1\wedge\ldots\wedge d\theta^k \qquad (15)\end{aligned}$$

Where we have used the fact that tensor multiplication by a scalar field is just regular multiplication and equations 5, and 9. Thus, robustness is obtained when π, ξ and the metric g_{im} are connected through the partial differential equation:

$$L_\xi\log\pi = -\frac{1}{2}g^{im}L_\xi g_{im} \tag{16}$$

This is again an equation expressing conservation of information. There is invariance along ξ when the gradient of the log-likelihood (of the prior or posterior π) projected onto ξ exactly eliminates the sources of information created by the deformation. But, if the deformation ξ does not artificially create information, i.e. if it preserves the state of complete ignorance then, by 9 and 4, the general equation 16 simplifies to,

$$\xi\cdot\nabla\pi = 0 \tag{17}$$

This is not surprising. Since the gradient, $\nabla\pi$, is always orthogonal to the level surfaces $\{\theta : \pi(\theta) = c\}$ we can rewrite 17 as,

Theorem 1 *Let ξ be an ignorance preserving vector field. Then, a probability distribution on Θ with scalar field density π, is robust with respect to deformations along ξ iff π puts constant probability mass on the integral curves of ξ.*

In other words, prior (or posterior) probabilities do not change, only when the deformations remain inside the level surfaces of the density.

3.3. Robustness of Entropic Priors

The name and the derivation of Entropic Priors for the manifold of discrete distributions are due to Skilling (see [11]). The generalization to arbitrary regular parametric models appears in the same volume in [7], see also [9], [8], [10].

Entropic priors are defined by their scalar field density. In the coordinate system of the θ's they are given by

$$\pi(\theta) = \frac{1}{c} e^{-\alpha I(\theta : \theta_0)} \tag{18}$$

where, $I(\theta : \theta_0)$ is the Kullback number between the distributions labeled by θ and a given initial value θ_0. The parameter $\alpha \geq 0$ has to be large enough, so that the constant of integration, c, is finite. Equation 18 has an easy interpretation: The chance of θ decreases exponentially fast with the Kullback distance from θ_0 and the parameter α controls the sensitivity to changes in the distance. Since the density is given as a scalar field, this interpretation holds in **all coordinate systems** i.e. for all parameterizations of the model.

To compute the sensitivity of entropic priors, with respect to deformations of the model, we need only to replace 18 into 15. If we denote by Π the entropic prior probability measure, we have:

$$\frac{dL_\xi \Pi}{d\Pi} = \frac{1}{2} g^{im} L_\xi g_{im} - \alpha \xi \cdot \nabla I(\theta : \theta_0) \tag{19}$$

where the left hand side denotes the Radon-Nikodym derivative of the (signed) measure $L_\xi \Pi$ with respect to Π. If ξ preserves ignorance, the first term of the sum in 19 is zero and

$$\frac{dL_\xi \Pi}{d\Pi} = -\alpha \xi \cdot \nabla I(\theta : \theta_0). \tag{20}$$

Equation 20 contains a lot of information about the nature of entropic priors. Firstly, notice that the parameter α controls the size of the derivative. In other words, the smaller α is, the more robust the inferences are. Jeffreys priors appear as tautological winners: *Ignorance priors are robust with respect to deformations preserving ignorance.* Besides this tautological robustness, obtained when $\alpha = 0$, we have robustness when ξ is orthogonal to ∇I. In other words, when the integral curves of ξ are located on the surface of *entropy spheres* centered at θ_0 i.e. $\{\theta : I(\theta : \theta_0) = \text{const.}\}$. This justifies the following:

Definition 2 *A vector field defined on the statistical model is said to be* information preserving at θ_0 *if it does not change ignorance and has integral curves contained in the level surfaces of* $I(\theta : \theta_0)$.

This definition makes true the following:

Theorem 2 *Entropic priors are robust with respect to deformations preserving information at the initial guess,* θ_0

It is well known that the Kullback number generates the Riemannian metric (see [7] or [9]). In fact, a simple Taylor expansion of the Kullback number produces:

$$\begin{aligned}\xi \cdot \nabla I(\theta : \theta_0) &= \langle \xi, \theta - \theta_0 \rangle_\theta + o(|\theta - \theta_0|) \\ &= \xi^i v^j g_{ij}(\theta) + o(|\theta - \theta_0|) \end{aligned} \tag{21}$$

where $v = \theta - \theta_0$ is in fact a tangent vector at θ when θ_0 approaches θ. From here, we obtain the following

Theorem 3 *Entropic Priors are robust with respect to local isometries at* θ_0

By local isometries at θ_0 we mean deformations that close to θ_0 do not change the metric. These deformations just send points in spherical orbits around θ_0. The previous theorem shows that entropic priors, as opposed to other classes of priors, are very compatible with the intrinsic Riemannian geometry of the hypothesis space. Entropic priors are sensitive only to deformations that change the *intrinsic form* of the model around θ_0.

4. Example: The Gaussians

The main purpose of this section is to illustrate some of the formulas introduced in this paper with a concrete example. An in depth analysis of the robustness of gaussians, however, is beyond the scope of the present article.

The gaussian distributions form a two dimensional Riemannian space. The metric tensor (Fisher information matrix) in the coordinate system $\theta = (\theta^1, \theta^2) = (\mu, \sigma)$ is diagonal with $g_{11} = 1/\sigma^2$, $g_{22} = 2/\sigma^2$. Thus,

$$g^{ij} \nabla g_{ij} = \sigma^2 \left(0, \frac{-2}{\sigma^3}\right) + \frac{\sigma^2}{2}\left(0, \frac{-4}{\sigma^3}\right) \tag{22}$$

The vector fields $\xi = (\xi^1, \xi^2)$ that preserve ignorance are given, from 13 and 22, by

$$\frac{\partial \xi^1}{\partial \mu} + \frac{\partial \xi^2}{\partial \sigma} - \frac{2}{\sigma}\xi^2 = 0 \tag{23}$$

This can be easily shown to have the general solution:

$$\xi = \left(h(\sigma) + \frac{2}{\sigma}\psi - \frac{\partial \psi}{\partial \sigma}, \frac{\partial \psi}{\partial \mu}\right) \tag{24}$$

where h is an arbitrary differentiable function of σ, and ψ is an arbitrary differentiable function of μ and σ, with continuous second order partial derivatives.

4.1. Entropic Prior on the Gaussians

Consider the entropic prior model on the manifold of gaussians with initial guess $\theta_0 = (0, 1)$ i.e. the standard normal distribution. Straight forward computations show 18 to be,

$$\pi(\mu, \sigma) = \frac{1}{c(\alpha)} \sigma^\alpha e^{-\frac{\sigma^2}{2\alpha}} e^{-\frac{\mu^2}{2\alpha}} \tag{25}$$

Equation 25 is the scalar field density relative to the volume element:

$$\sigma^{-2}\, d\mu \wedge d\sigma \tag{26}$$

The symbolic manipulator MAPLE, shows that, for $\alpha > 1$

$$c(\alpha) = 2^{(\alpha/2-1)}\sqrt{\pi}\alpha^{-\alpha/2}\Gamma\left(\frac{1}{2}(\alpha-1)\right) \tag{27}$$

The integral of 25 with respect to 26 diverges for $\alpha \leq 1$. I believe this to be the reason of why Jeffreys, and many others after him, thinks that the volume element 26 (i.e. $\alpha = 0$) is too uninformative. The prior distribution obtained at the first divergent value of c, when $\alpha = 1$, produces posterior inferences remarkably similar to the popular conjugate prior for this case. Even for two observations. This suggests to extend the definition of uninformative prior to include all the entropic priors with divergent c. The boundary value of α (in this case $\alpha = 1$) can be used to approximate the frequentist methods.

The level curves of 25, are closed curves on the upper half plane (μ, σ) with equations,

$$\frac{\mu^2}{2\alpha} + \frac{\sigma^2}{2\alpha} - \alpha \log \sigma = k \tag{28}$$

where k depends on α. Computer experiments show these curves to be similar to ellipses centered about $(0, 1-\epsilon(\alpha))$ with ϵ tending to zero as α increases. Therefore, there is robustness when the velocity vectors 24 are tangent to the level curves, 28. This happens when,

$$\left(\sigma^2 - \alpha^2\right)\frac{1}{\mu}\frac{\partial\Psi}{\partial\mu} - \sigma\frac{\partial\Psi}{\partial\sigma} - 2\Psi + \sigma h(\sigma) = 0 \tag{29}$$

Preliminary analysis indicates that equation 29 imposes a heavy restriction on the deformations for which there is robustness. The group of isometries of the gaussians, together with theorem 3, could be used to find the desired deformations around $\theta_0 = (0, 1)$. It is possible to show, that the group of direct isometries of the gaussians, is that of the Lobachevskian plane. This group, is known to be isomorphic to the orthochronous connected component of the identity of the Lorentz group for three dimensional space-time (i.e. a space with metric: $x^2 + y^2 - t^2$, see [9]).

5. Conclusions

In retrospect, this paper should be consider a first attempt to demonstrate that it makes sense to use Lie derivatives for quantifying bayesian robustness. No doubt, the geometrization of Inference provides a powerful language for asking questions about statistical procedures. As usual, geometry brings the paraphernalia of visual imagery that embodies the objects of study and allows to *see* the theorems. Looking ahead, to the (immediate) future, we can anticipate that many of the successful applications of modern geometry to physics might be reproduced, for the theory of Inference. Stokes theorem will begin to play a central role in bayesian robustness, for the very simple reason that the Lie derivatives of priors and posteriors are again differential forms ready to be integrated over patches. There is also room for connections and gauges, square roots of laplacians, Lie algebras and index theorems. We need to find the people with the *guts* to do it.

References

[1] Shun-ichi Amari. *Differential-Geometrical Methods in Statistics*, volume 28 of *Lecture Notes in Statistics*. Springer-Verlag, 1985.

[2] B.A. Dubrovin, A.T. Fomenko, and S.P. Novikov. *Modern Geometry–Methods and Applications, Part-I*, volume GTM 93 of *Graduate Texts in Mathematics*. Springer-Verlag, 1984.

[3] F. R. Hampel. A general qualitative definition of robustness. *Ann. Math. Statist.*, 42:1887–1896, 1971.

[4] P. J. Huber. *Robust Statistics*. John Wiley and Sons, 1981.

[5] J. B. Kadane, editor. *Robustness of Bayesian Analyses*. North-Holland, 1981.

[6] J. Pfanzagl. *Contributions to a general asymptotic statistical theory*. Lecture Notes in Statistics. Springer-Verlag, New York, 1982.

[7] Carlos C. Rodríguez. The metrics induced by the kullback number. In John Skilling, editor, *Maximum Entropy and Bayesian Methods*. Kluwer Academic Publishers, 1989.

[8] Carlos C. Rodríguez. Objective bayesianism and geometry. In Paul F. Fougère, editor, *Maximum Entropy and Bayesian Methods*. Kluwer Academic Publishers, 1990.

[9] Carlos C. Rodríguez. Entropic priors. Available in Electronic Form on the Internet "gopher", Oct. 1991. `gopher cscgoph2.albany.edu 2-4-12-1-1`.

[10] Carlos C. Rodríguez. From euclid to entropy. In W. T. Grandy, Jr., editor, *Maximum Entropy and Bayesian Methods*. Kluwer Academic Publishers, 1991.

[11] John Skilling. Classical Max Ent data analysis. In John Skilling, editor, *Maximum Entropy and Bayesian Methods*. Kluwer Academic Publishers, 1989.