Intrinsically Optimal Equivariant Estimator for the Univariate Linear Normal Models

Gloria García , Marta Cubedo * , Josep Maria Oller

Posted Date: 13 October 2025

doi: 10.20944/preprints202510.0742.v1

Keywords: Equivariant estimator; Information metric; Riemannian risk; Intrinsic bias

 doi:10.20944/preprints202510.0742.v1

Article

Intrinsically Optimal Equivariant Estimator for the Univariate Linear Normal Models

Gloria García [1], Marta Cubedo [2,*] and Josep M. Oller [2]

[1] Department of Education and Professional Training, Gencat,08021 Barcelona, Spain; ggarci76@xtec.cat
[2] University of Barcelona, Avda. Diagonal 643, 08028 Barcelona, Spain; joller@ub.edu
* Correspondence: mcubedo@ub.edu

Abstract

In this paper we prove the existence and unicity of the minimum risk equivariant estimator (MIRE) for the univariate linear normal models, under the action of the subgroup of the affine group which leaves the column space of the design matrix invariant and using the framework of the intrinsic analysis of statistical estimation which uses the square of the Rao distance as the loss function, which expectation behaves very different to the corresponding of the square error loss when we work with small samples, and supplying also an intrinsic bias measure for any equivariant estimator. This estimator is studied and compared with the standard maximum likelihood estimator (MLE) in terms of its intrinsic risk and bias, showing explicitly its differences, apparent with small samples, fortunately small for large sample sizes. Moreover, an approximate and very simple estimator is also suggested, which reduces the majority of bad behaviour of MLE, of intrinsic risk and bias, for small samples.

Keywords: equivariant estimator; information metric; riemannian risk; intrinsic bias

PACS: 62F00, 60K35, 62A00

1. Introduction

The Fisher information matrix of a regular parametric family of probability distributions induces a natural Riemannian structure on the parameter space [1–3]. This structure provides the foundation for the intrinsic analysis of statistical estimation, see [4,5] and, beyond statistics, it has also been shown to reveal connections with physical laws [6].

In this intrinsic framework, estimators are assessed through tools that do not depend on any particular parametrization of the model. Consequently, non-intrinsic criteria such as the squared error loss should be replaced by intrinsic measures. A natural choice is the squared Riemannian (Rao) distance, which serves as an intrinsic loss function. The associated risk can behave very differently from that based on squared error loss, especially in small-sample regimes. Similarly, the classical definition of bias must be reformulated in terms of a vector field determined by the geometry of the model, whose squared norm provides a natural intrinsic bias measure.

The estimator itself should also be intrinsic, i.e., independent of parametrization. Consider, for instance, the exponential distribution under two parametrizations,

$$f(x,\theta) = \theta \exp(-\theta x)\, \mathbb{1}_{\mathbb{R}^+}(x) \qquad \text{and} \qquad f(x,\lambda) = \frac{1}{\lambda} \exp(-x/\lambda)\, \mathbb{1}_{\mathbb{R}^+}(x)$$

where $\mathbb{1}_{\mathbb{R}^+}(x)$ is the positive real numbers indicator. The UMVU estimator computed under each parametrization $\widehat{\theta}$ and $\widehat{\lambda}$ are not related by $\widehat{\theta} = 1/\widehat{\lambda}$. This apparent inconsistency arises because UMVU estimators rely on non-intrinsic notions such as unbiasedness and variance. As a result, the UMVU estimator is parametrization-dependent and cannot be regarded as intrinsically defined.

When the statistical model is invariant under the action of a transformation group on the sample space, it is natural to restrict attention to equivariant estimators, i.e., estimators U satisfying

 doi:10.20944/preprints202510.0742.v1

$$U(g(x)) = \overline{g}(U(x)) \quad \text{for all} \quad g \in \mathcal{G}_n$$

where $\mathcal{G}_n$ is the transformation group acting on the space of all samples of size n and $\overline{g} \in \overline{\mathcal{G}}$ is the induced transformation on the parameter space Θ. This restriction ensures logical consistency: only equivariant estimators remain coherent under data transformations. It is therefore a requirement that should be imposed prior to any attempt to minimize an intrinsic risk function.

As an illustration, consider the estimation of the unconstrained mean μ of a p–variate normal distribution. The MLE of μ (which really defines and intrinsic estimator) based on a sample of size n, is just the sample mean vector $\overline{X}_n$. However, for $p \geq 3$, the James–Stein estimator is known to have smaller quadratic risk [7]. In our view, this does not undermine the MLE. Although the James–Stein estimator enjoys a lower risk under the non-intrinsic quadratic loss, it lacks the essential property of equivariance, which is crucial from the intrinsic perspective.

In this work we focus on classical statistical estimation, as is standard in the absence of prior knowledge. Nonetheless, intrinsic Bayesian approaches based on non-informative priors are also possible [5,8], and represent an interesting avenue for future research. Here, we adopt the squared Rao distance as a natural loss function, since it is the conceptually simplest intrinsic analogue of the quadratic loss, despite potential computational challenges. For a broader background on statistical estimation and its intrinsic developments, see [9–12].

Specifically, we consider the univariate linear normal model with a fixed design matrix. First, we explicitly characterize the class of equivariant estimators under the action of a suitable subgroup of the affine group–namely, those affine transformations of the data that leave the column space of the design matrix unchanged. This subgroup is the largest one that preserves the model's structure. Next, we prove the existence and uniqueness of the estimator within this class that minimizes intrinsic risk, i.e., the equivariant estimator that minimizes the mean squared Rao distance. We also derive an explicit expression for the intrinsic bias of any equivariant estimator.

Furthermore, we compare the intrinsic bias and risk of the proposed estimator with those of the MLE, highlighting their differences in small samples and showing how these differences diminish as the sample size grows. Finally, we propose a computable approximation of the optimal estimator, which mitigates most of the intrinsic bias and risk issues that affect the MLE in finite samples.

2. Equivariant estimators for linear models

Let us consider the univariate linear normal model,

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \tag{1}$$

where $\boldsymbol{y}$ is a $n \times 1$ random vector, $\boldsymbol{X}$ is a $n \times m$ matrix of known constants with $0 < \text{rank}(\boldsymbol{X}) = m \leq n$, β is a $m \times 1$ vector of unknown parameters to be estimated and $\boldsymbol{e}$ is the fluctuation or error of $\boldsymbol{y}$ about $\boldsymbol{X}\boldsymbol{\beta}$. We assume that the errors are unbiased, independent, with the same variance σ^2 and following a n–variate normal distribution, that is $\boldsymbol{e} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$, where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. Therefore $\boldsymbol{y}$ distributes according to an element of the parametric family of probability distributions $\mathcal{P} = \{N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n) \mid (\boldsymbol{\beta}, \sigma) \in \Theta\}$ with parameter space $\Theta = \mathbb{R}^m \times \mathbb{R}_+$, a $(m+1)$ – dimensional simply connected real manifold. Hereafter, we shall identify the elements of $\mathbb{R}^m$ with $m \times 1$ column vectors, when necessary.

Denote by

$$O_E(n) = \{\boldsymbol{H} \in O(n) \mid \boldsymbol{y} \in E \Rightarrow \boldsymbol{H}\boldsymbol{y} \in E\}$$

where E is a subspace of $\mathbb{R}^n$ and $O(n)$ is the group of $n \times n$ orthogonal matrices with entries from $\mathbb{R}$. Define F as the subspace of $\mathbb{R}^n$ spanned by the columns of $\boldsymbol{X}$, that is $F = \text{Col}(\boldsymbol{X})$. Observe that $O_F(n) = O_{F^\perp}(n)$, $\boldsymbol{I}_n \in O_F(n)$, $\text{Col}(\boldsymbol{X}) = \text{Col}(\boldsymbol{H}\boldsymbol{X}) \ \ \forall \boldsymbol{H} \in O_F(n)$ and if $\boldsymbol{H} \in O_F(n)$ then $\boldsymbol{H}^t \in O_F(n)$. Moreover, every $\boldsymbol{H} \in O_F(n)$ induces, in F and $F^\perp$, two isomorphisms isomorphisms preserving the Euclidean norm in each subspace.

 doi:10.20944/preprints202510.0742.v1

$\mathcal{P}$ is invariant under the action of the subgroup of the affine group in $\mathbb{R}^n$ given by the family of transformations

$$g_{(a,\boldsymbol{H},\boldsymbol{c})}(\boldsymbol{y}) = a\boldsymbol{H}\boldsymbol{y} + \boldsymbol{c}\ ,\ \boldsymbol{y} \in \mathbb{R}^n \tag{2}$$

where $a > 0, \boldsymbol{H} \in O_F(n), \boldsymbol{c} \in F$.

Observe that $\mathcal{G} = \{g_{(a,\boldsymbol{H},\boldsymbol{c})} \mid a > 0, \boldsymbol{H} \in O_F(n), \boldsymbol{c} \in F\}$ induces an action on the parameter space θ given by

$$\overline{g_{(a,\boldsymbol{H},\boldsymbol{c})}}(\boldsymbol{\beta},\sigma) \quad = \quad \left(a\,(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{H}\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{c},\ a\sigma\right) \tag{3}$$

where this result is obtained taking into account that $\boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t$ is the projection matrix into F and thus, if $\boldsymbol{w} \in F$ there exists a unique $\boldsymbol{\eta}$ such that $\boldsymbol{w} = \boldsymbol{X}\boldsymbol{\eta}$ and $\boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{w} = \boldsymbol{w}$.

Since the family $\mathcal{P}$ is invariant under the action of $\mathcal{G}$, it is natural to restrict our attention to the class of equivariant estimators $\boldsymbol{U} = (\boldsymbol{U}_1, U_2)$ of $(\boldsymbol{\beta},\sigma)$ i.e. an estimator satisfying $\boldsymbol{U}\left(g_{(a,\boldsymbol{H},\boldsymbol{c})}(\boldsymbol{y})\right) = \overline{g_{(a,\boldsymbol{H},\boldsymbol{c})}}(\boldsymbol{U}(\boldsymbol{y}))$ for all $g_{(a,\boldsymbol{H},\boldsymbol{c})} \in \mathcal{G}$.

Proposition 1. *Let $\boldsymbol{U}$ be an equivariant estimator of $(\boldsymbol{\beta},\sigma)$. Then $\boldsymbol{U}$ belongs to the family $\{\boldsymbol{U}_\lambda, \lambda \in \mathbb{R}_+\}$ where,*

$$\boldsymbol{U}_\lambda(\boldsymbol{y}) = \left((\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y},\ \lambda\left\{\boldsymbol{y}^t\left(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\right)\boldsymbol{y}\right\}^{1/2}\right)$$

Proof. Let $\boldsymbol{U} = (\boldsymbol{U}_1, U_2)$. The equivariance condition for $\boldsymbol{U}$ involves,

$$\begin{array}{rcl} \boldsymbol{U}_1(a\boldsymbol{H}\boldsymbol{y} + \boldsymbol{c}) & = & a\,(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{H}\boldsymbol{X}\boldsymbol{U}_1(\boldsymbol{y}) + (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{c} \\ U_2(a\boldsymbol{H}\boldsymbol{y} + \boldsymbol{c}) & = & a\,U_2(\boldsymbol{y}) \end{array}$$

for any $a > 0$, $\boldsymbol{H} \in O_F(n)$, $\boldsymbol{c} \in F$.

Any $\boldsymbol{y} \in \mathbb{R}^n$ can be written in a unique form as $\boldsymbol{y} = \boldsymbol{y}_F + \boldsymbol{y}_{F^\perp}$ where $\boldsymbol{y}_F \in F$ and $\boldsymbol{y}_{F^\perp} \in F^\perp$. Specifically

$$\boldsymbol{y}_F = \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y} \quad \text{and} \quad \boldsymbol{y}_{F^\perp} = (\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t)\boldsymbol{y} \tag{4}$$

If we choose $\boldsymbol{c} = -a\boldsymbol{H}\boldsymbol{y}_F$ in the previous expressions, we obtain

$$\boldsymbol{U}_1(a\boldsymbol{H}\boldsymbol{y}_{F^\perp}) \quad = \quad a\,(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{H}\boldsymbol{X}\boldsymbol{U}_1(\boldsymbol{y}) - a(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{H}\boldsymbol{y}_F \tag{5}$$

$$U_2(a\boldsymbol{H}\boldsymbol{y}_{F^\perp}) \quad = \quad a\,U_2(\boldsymbol{y}) \tag{6}$$

First we focus on $\boldsymbol{U}_1$. If we let $\boldsymbol{H}^* = (\boldsymbol{I}_n - 2\boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t)$, we have $\boldsymbol{H}^* \in O_F(n)$ with

$$\boldsymbol{H}^*\boldsymbol{v} = -\boldsymbol{v} \quad \forall \boldsymbol{v} \in F \quad \text{and} \quad \boldsymbol{H}^*\boldsymbol{v} = \boldsymbol{v} \quad \forall \boldsymbol{v} \in F^\perp$$

Now observe that (5) is satisfied for any $a > 0$ and $\boldsymbol{H}$ in $O_F(n)$, in particular for $\boldsymbol{I}_n$ and $\boldsymbol{H}^*$. Therefore

$$\begin{array}{rcl} \boldsymbol{U}_1(a\boldsymbol{y}_{F^\perp}) & = & a\boldsymbol{U}_1(\boldsymbol{y}) - a(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}_F \\ \boldsymbol{U}_1(a\boldsymbol{H}^*\boldsymbol{y}_{F^\perp}) & = & -a\boldsymbol{U}_1(\boldsymbol{y}) + a(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}_F \end{array}$$

But $a\boldsymbol{H}^*\boldsymbol{y}_{F^\perp} = a\boldsymbol{y}_{F^\perp}$ which leads to

$$0 = \boldsymbol{U}_1(\boldsymbol{y}) - (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}_F$$

From (4),

$$\boldsymbol{U}_1(\boldsymbol{y}) = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}\,, \quad \boldsymbol{y} \in \mathbb{R}^n$$

 doi:10.20944/preprints202510.0742.v1

Next, we consider U_2. Let $a = 1$ and $\boldsymbol{H} = \boldsymbol{I}_n$ in (6), it follows

$$U_2(\boldsymbol{y}_{F^\perp}) = U_2(\boldsymbol{y}) \tag{7}$$

Accordingly, it is enough to determine U_2 on $F^\perp$. Let us take a unit vector $\boldsymbol{z} \in F^\perp$ such that $U_2(\boldsymbol{z}) = \lambda > 0$. Then (6).

$$U_2(\boldsymbol{H}\boldsymbol{z}) = U_2(\boldsymbol{z})$$

for any $\boldsymbol{H} \in O_F(n)$. Observe that any arbitrary unit vector in $F^\perp$ can be written as $\boldsymbol{H}\boldsymbol{z}$ for a proper $\boldsymbol{H} \in O_F(n)$. Therefore, for any $\boldsymbol{y}_{F^\perp} \in F^\perp$, $\boldsymbol{y}_{F^\perp} \neq \boldsymbol{0}$, we have

$$U_2(\boldsymbol{y}_{F^\perp}) = \|\boldsymbol{y}_{F^\perp}\|\, U_2\left(\frac{\boldsymbol{y}_{F^\perp}}{\|\boldsymbol{y}_{F^\perp}\|}\right) = \|\boldsymbol{y}_{F^\perp}\|\, U_2(\boldsymbol{H}\boldsymbol{z}) = \|\boldsymbol{y}_{F^\perp}\|\, U_2(\boldsymbol{z}) = \lambda\, \|\boldsymbol{y}_{F^\perp}\|$$

Observe also that if $\boldsymbol{y} = \boldsymbol{0}$ then $\boldsymbol{y}_{F^\perp} = 0$ and, choosing $\boldsymbol{c} = \boldsymbol{0}$, we have that $U_2(\boldsymbol{0}) = a\, U_2(\boldsymbol{0})$, $\forall a \in \mathbb{R}_+$. This implies $U_2(\boldsymbol{0}) = 0$.

Finally, from (7) and (4) we have

$$U_2(\boldsymbol{y}) = \lambda \left\{\boldsymbol{y}^{\mathrm{t}}\left(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{t}}\right)\boldsymbol{y}\right\}^{1/2}$$

□

Observe that the standard maximum likelihood estimator, MLE, for the present model is an equivariant estimator, with $\lambda = \frac{1}{\sqrt{n}}$.

3. Minimum Riemannian risk estimators

In the framework of intrinsic analysis, where the loss function is the square of the Rao distance, the Riemannian distance induced by the information metric in the parameter space Θ, once the class of the equivariant estimators has been determined a natural question arises: which is the equivariant estimator that minimizes the risk?

First of all, we summarize the basic geometric results corresponding to the model (1) which are going to be used hereafter. We are going to use a *standardized* version of the information metric, given by the usual *information metric* corresponding to this linear model divided by a constant factor n, i.e. the number of rows of matrix $\boldsymbol{X}$. This metric is given by

$$ds^2 = \frac{1}{n\,\sigma^2}\left(d\boldsymbol{\beta}^{\mathrm{t}}\,\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X}\,d\boldsymbol{\beta} + 2n\,(d\sigma)^2\right), \tag{8}$$

which is, up to a linear coordinate change, the *Poincaré hyperbolic metric* of the upper half space $\mathbb{R}^m \times \mathbb{R}^+$, see [13]. The Riemannian curvature $\kappa = -\frac{1}{2}$ is constant and negative and the unique geodesic, parameterized by the arc–length, which connects two points $\theta_1 = (\boldsymbol{\beta}_1, \sigma_1)$ and $\theta_2 = (\boldsymbol{\beta}_2, \sigma_2) \in \Theta$, when $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$, is given by:

$$\begin{aligned} \boldsymbol{\beta}(s) &= 2n\,K^{-2}\,\tanh\left(\frac{s}{\sqrt{2}} + \epsilon\right)(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1/2}\boldsymbol{C} + \boldsymbol{D} \\ \sigma(s) &= K^{-1}\,\mathrm{sech}\left(\frac{s}{\sqrt{2}} + \epsilon\right) \end{aligned} \tag{9}$$

where s is the arc–length, $\boldsymbol{C}$ and $\boldsymbol{D}$ are $m \times 1$ vectors whose components, and also ϵ, are convenient real integration constants, such that $\boldsymbol{\beta}(0) = \boldsymbol{\beta}_1$, $\boldsymbol{\beta}(\rho_{12}) = \boldsymbol{\beta}_2$, $\sigma(0) = \sigma_1$, $\sigma(\rho_{12}) = \sigma_2$ being ρ_{12} the

 doi:10.20944/preprints202510.0742.v1

Riemannian distance between θ_1 and θ_2. Finally, K is given by $K^2 = 2n\ \boldsymbol{C}^{\mathrm{t}}\boldsymbol{C}$. When $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, the geodesic is given by

$$\boldsymbol{\beta}(s) = \boldsymbol{D} \qquad \sigma(s) = B\, e^{\pm\frac{s}{\sqrt{2}}} \tag{10}$$

where B is a positive integration constant.

The Rao distance ρ between the points θ_1 and θ_2 is

$$\rho_{12} \equiv \rho(\theta_1,\theta_2) = \sqrt{2}\ \ln\left(\frac{1+\delta(\theta_1,\theta_2)}{1-\delta(\theta_1,\theta_2)}\right) = 2\sqrt{2}\,\operatorname{arctanh}(\delta(\theta_1,\theta_2)) \tag{11}$$

where

$$\delta(\theta_1,\theta_2) = \left(\frac{d_M^2(\boldsymbol{\beta_1},\boldsymbol{\beta_2}) + 2\dfrac{(\sigma_1-\sigma_2)^2}{\sigma_1\sigma_2}}{d_M^2(\boldsymbol{\beta_1},\boldsymbol{\beta_2}) + 2\dfrac{(\sigma_1+\sigma_2)^2}{\sigma_1\sigma_2}}\right)^{1/2} \tag{12}$$

and

$$d_M^2(\boldsymbol{\theta_1},\boldsymbol{\theta_2}) = \frac{1}{n\ \sigma_1\sigma_2}\ (\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2) \tag{13}$$

or, equivalently,

$$\rho(\theta_1,\theta_2) = \sqrt{2}\,\operatorname{arccosh}\left(\frac{1}{4}\ d_M^2(\boldsymbol{\theta_1},\boldsymbol{\theta_2}) + \frac{\sigma_1^2+\sigma_2^2}{2\ \sigma_1\sigma_2}\right) \tag{14}$$

Let $\exp_{\theta_1}^{-1}(\theta_2)$ be the inverse of the exponential map corresponding to Levi-Civita connection and $W^1,\ldots,W^m,W^{m+1}$ its components corresponding to the basis field $(\partial/\partial\beta^1)_{\theta_1},\ldots,(\partial/\partial\beta^m)_{\theta_1},(\partial/\partial\sigma)_{\theta_1}$. Then, we have

$$\begin{aligned} W^i &= \frac{\rho_{12}/\sqrt{2}}{\sinh(\rho_{12}/\sqrt{2})}\ (\beta_2^i-\beta_1^i)\ \frac{\sigma_1}{\sigma_2}, \qquad i=1,\ldots,m \\ W^{m+1} &= \frac{\rho_{12}/\sqrt{2}}{\sinh(\rho_{12}/\sqrt{2})}\left(\cosh(\rho_{12}/\sqrt{2}) - \frac{\sigma_1}{\sigma_2}\right)\sigma_1 \end{aligned} \tag{15}$$

It is well know that the Riemannian distance induced by the information metric is invariant under equivariant estimator transformations. We shall supply a direct and alternative proof for the linear model setting.

Proposition 2. *The Rao distance ρ given by (11) is invariant under the action of the induced group by $\mathcal{G}$ on the parameter space, $\overline{\mathcal{G}}$. In other words*

$$\rho(\overline{g_{(a,\boldsymbol{H},c)}}\ \theta_1, \overline{g_{(a,\boldsymbol{H},c)}}\ \theta_2) = \rho(\theta_1,\theta_2)$$

Proof: Observe that

$$\boldsymbol{H}\boldsymbol{X}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2) \in F\ , \forall \boldsymbol{H} \in O_F(n)$$

and taking into account that $\boldsymbol{X}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{t}}$ is the projection matrix into F, we have

$$\boldsymbol{X}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{t}}\boldsymbol{H}\boldsymbol{X}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2) = \boldsymbol{H}\boldsymbol{X}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)$$

Therefore

$$(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}\boldsymbol{H}^{\mathrm{t}}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{t}}\boldsymbol{H}\boldsymbol{X}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2) = (\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)$$

and the invariance of δ and ρ trivially follows. □

 doi:10.20944/preprints202510.0742.v1

Proposition 3. *$\overline{\mathcal{G}}$ acts transitively on Θ.*

Proof: The transitivity follows observing that a is an arbitrary positive real number and $\boldsymbol{X}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{t}}$ is the projection matrix into F with rank $\boldsymbol{X} = m$. □

Since ρ, and thus ρ^2, is invariant under the action of $\overline{\mathcal{G}}$ and $\overline{\mathcal{G}}$ acts transitively on Θ, the distribution of $\rho^2(\boldsymbol{U}_\lambda(\boldsymbol{y}), \theta)$ does not depend on θ, and therefore, the risk of any equivariant estimator remains constant and independent of the target parameter provided that this risk is finite. More precisely, observe that if we let

$$\boldsymbol{z} = \frac{1}{\sigma}\,(\boldsymbol{y}_F - \boldsymbol{X}\boldsymbol{\beta})\,, \quad \mathcal{V} = \|\boldsymbol{z}\| \quad \text{and} \quad \mathcal{W} = \frac{1}{\sigma}\,\|\boldsymbol{y}_{F^\perp}\| \tag{16}$$

from (1) and (4) we clearly have that $\boldsymbol{z} \sim N_n(\boldsymbol{0}, \boldsymbol{X}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{t}})$ with a rank m idempotent covariance matrix, and $\mathcal{V}^2$ and $\mathcal{W}^2$ are independent random variables following a chi-square distribution with m and $(n-m)$ degrees of freedom, equal to the dimensions of F and $F^\perp$ since $\mathcal{V}^2$ and $\mathcal{W}^2$ are quadratic forms based on the projection matrices on these subspaces of $\mathbb{R}^n$ and $\boldsymbol{y}_F$ (or $\boldsymbol{z}$) and $\boldsymbol{y}_{F^\perp}$ are independent random vectors. Therefore, since $\boldsymbol{X}\boldsymbol{U}_{\lambda 1}(\boldsymbol{y}) = \boldsymbol{y}_F$ and $U_{\lambda 2}(\boldsymbol{y}) = \lambda\|\boldsymbol{y}_{F^\perp}\|$, we have that

$$d_M^2(\boldsymbol{U}_\lambda(\boldsymbol{y}), \theta) = \frac{1}{n\,U_{\lambda 2}(\boldsymbol{y})\,\sigma}\,(\boldsymbol{U}_{\lambda 1}(\boldsymbol{y}) - \boldsymbol{\beta})^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X}(\boldsymbol{U}_{\lambda 1}(\boldsymbol{y}) - \boldsymbol{\beta}) = \frac{\mathcal{V}^2}{n\,\lambda\,\mathcal{W}} \tag{17}$$

$$\delta(\boldsymbol{U}_\lambda(\boldsymbol{y}), \theta) = \left(\frac{\dfrac{\mathcal{V}^2}{n\,\lambda\,\mathcal{W}} + 2\,\dfrac{(\lambda\,\mathcal{W}-1)^2}{\lambda\,\mathcal{W}}}{\dfrac{\mathcal{V}^2}{n\,\lambda\,\mathcal{W}} + 2\,\dfrac{(\lambda\,\mathcal{W}+1)^2}{\lambda\,\mathcal{W}}}\right)^{1/2} \tag{18}$$

and

$$\rho(\boldsymbol{U}_\lambda(\boldsymbol{y}), \theta) = 2\sqrt{2}\,\operatorname{arctanh}\left(\left(\frac{\mathcal{V}^2 + 2n\,(\lambda\,\mathcal{W}-1)^2}{\mathcal{V}^2 + 2n\,(\lambda\,\mathcal{W}+1)^2}\right)^{1/2}\right) \tag{19}$$

or

$$\rho(\boldsymbol{U}_\lambda(\boldsymbol{y}), \theta) = \sqrt{2}\,\operatorname{arccosh}\left(\frac{1}{4}\,\frac{\mathcal{V}^2}{n\,\lambda\,\mathcal{W}} + \frac{1}{2}\left(\lambda\,\mathcal{W} + \frac{1}{\lambda\,\mathcal{W}}\right)\right) \tag{20}$$

which have a distribution which depend only on $\mathcal{V}^2$ and $\mathcal{W}^2$, independent random variables with fixed distribution, whatever the value of θ.

Since the risk of any equivariant estimator remains constant on the parameter space, it's enough to examine it at one point, for instance at the point $(\boldsymbol{0}, 1) \in \Theta$. Let us denote the expectation with respect to the n–variate linear normal model $N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_n)$ by $E_{(\boldsymbol{\beta},\sigma)}$ and by E the $E_{(\boldsymbol{0},1)}$. We can prove the following propositions.

Proposition 4. *If $n \geq m+1$ we have:*

$$E_{(\boldsymbol{\beta},\sigma)}\left(\rho^2(\boldsymbol{U}_\lambda(\boldsymbol{y}), (\boldsymbol{\beta}, \sigma))\right) = E\left(\rho^2(\boldsymbol{U}_\lambda(\boldsymbol{y}), (\boldsymbol{0}, 1))\right) < \infty\,, \forall\lambda > 0$$

for any $(\boldsymbol{\beta}, \sigma) \in \Theta$.

Proof: From (14) and (13), since

$$1 + \frac{1}{2!}\,\frac{\rho^2}{2} + \frac{1}{4!}\,\frac{\rho^4}{4} \leq \cosh(\frac{\rho}{2}) = \frac{1}{4}\,d_M^2(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) + \frac{\sigma_1^2 + \sigma_2^2}{2\,\sigma_1\sigma_2} \tag{21}$$

 doi:10.20944/preprints202510.0742.v1

we have

$$0 \leq \rho^2(\theta_1,\theta_2) \leq 12\left(\sqrt{1+\frac{d_M^2(\boldsymbol{\theta_1},\boldsymbol{\theta_2})}{6}+\frac{1}{3}\frac{(\sigma_1-\sigma_2)^2}{\sigma_1\sigma_2}}-1\right) \tag{22}$$

developing the square of the difference and taking into account that the standard Euclidean norm of a vector is less or equal to the absolute value of the sum of its components, we obtain

$$0 \leq \rho^2(\theta_1,\theta_2) \leq 2\sqrt{6}\, d_M(\boldsymbol{\theta_1},\boldsymbol{\theta_2}) + 4\sqrt{3}\left(\sqrt{\frac{\sigma_1}{\sigma_2}}+\sqrt{\frac{\sigma_2}{\sigma_1}}\right)+4\sqrt{3}-12 \tag{23}$$

Notice that both bounds (22) and (23) are invariant under the action of the induced group on the parameter space.

As we mentioned before, from [14], it is enough to prove that the risk is finite at $(\mathbf{0},1) \in \Theta$. Taking into account (16) it follows, from (23), that

$$0 \leq \rho^2(\boldsymbol{U}_\lambda(\boldsymbol{y}),(\mathbf{0},1)) \leq 2\sqrt{6}\sqrt{\frac{\mathcal{V}^2}{\lambda\mathcal{W}\,n}}+4\sqrt{3}\left(\sqrt{\lambda\mathcal{W}}+\frac{1}{\sqrt{\lambda\mathcal{W}}}\right)+4\sqrt{3}-12 \tag{24}$$

Observe that if Q has a chi-square distribution with k degrees of freedom

$$E(Q^\alpha) = 2^\alpha \frac{\Gamma(\frac{k}{2}+\alpha)}{\Gamma(\frac{k}{2})} \quad \text{providedthat} \quad \frac{k}{2}+\alpha > 0$$

Therefore, since $\mathcal{V}^2$ and $\mathcal{W}^2$ are independent random variables following a central chi-square distribution with m and $(n-m)$ degrees of freedom, we have

$$E(\mathcal{V}) = \sqrt{2}\,\frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})},\quad E(\sqrt{\mathcal{W}}) = \sqrt[4]{2}\,\frac{\Gamma(\frac{2(n-m)+1}{4})}{\Gamma(\frac{n-m}{2})} \quad \text{provided that} \quad n-m > -1/2$$

and

$$E(\frac{1}{\sqrt{\mathcal{W}}}) = \frac{1}{\sqrt[4]{2}}\,\frac{\Gamma(\frac{2(n-m)-1}{4})}{\Gamma(\frac{n-m}{2})} \quad \text{provided that } n-m > 1/2$$

Then, taking the average, it follows that

$$\begin{aligned} E\left(\rho^2(\boldsymbol{U}_\lambda(\boldsymbol{y}),(0,1))\right) &< \frac{2\sqrt{6}}{\sqrt{\lambda}\,n}\,E(\mathcal{V})\,E(\frac{1}{\sqrt{\mathcal{W}}}) + \\ &\quad + 4\sqrt{3}\left(\sqrt{\lambda}\,E\,(\sqrt{\mathcal{W}})+\frac{1}{\sqrt{\lambda}}\,E\,(\frac{1}{\sqrt{\mathcal{W}}})\right)+4\sqrt{3}-12 \\ &< +\infty,\ \ \forall\lambda>0 \text{ and } n-m>1/2 \end{aligned}$$

Since n and m are positive integers being $n > m$, we conclude that the risk is finite if $n \geq m+1$. □

Proposition 4 is a sufficient condition for the existence of the Riemannian risk of the equivariant estimator $\boldsymbol{U}_\lambda$, thus

$$\Phi(\lambda) = E\left(\rho^2(\boldsymbol{U}_\lambda(\boldsymbol{y}),(\mathbf{0},1))\right),\ \lambda > 0, \tag{25}$$

is well defined for $n \geq m+1$, which we shall assume hereafter.

Proposition 5. *There exists a unique minimizer $\lambda_{nm} > 0$ of the Riemannian risk given by Φ.*

Proof:

 doi:10.20944/preprints202510.0742.v1

Let us consider the Riemannian risk at $\lambda = e^{\frac{s}{2\sqrt{2}}}$ as a function of s, that is

$$F(s) = \Phi(e^{\frac{s}{2\sqrt{2}}}), \quad s \in \mathbb{R}$$

The particular selection of λ, from which F follows, relies on the Riemannian structure of Θ induced by the information metric. The Riemannian curvature is constant and equal to $-\frac{1}{2}$ and taking into account (10) we have that

$$\gamma(s) = \left((\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{t}}\boldsymbol{y},\ e^{\frac{s}{\sqrt{2}}} \left\{ \boldsymbol{y}^{\mathrm{t}} \Big(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{t}} \Big) \boldsymbol{y} \right\}^{1/2} \right)$$

is a geodesic in Θ; precisely a geodesic parameterized by the arc–length, see [13] for further details.

Then, following [15], the real valued function

$$s \mapsto \rho^2(\gamma(s), (\mathbf{0}, 1))$$

is strictly convex. Since almost surely convexity of a stochastic process carries over the mean of a process, the map F is strictly convex as well.

On the other hand, from Fatou's Lemma $F(s) \to +\infty$ as $s \to -\infty$ or $s \to \infty$. This, together with the strict convexity of F yield the existence of a unique minimizer s^* of the function F, which depends on n and m.

Finally, since the map $s \mapsto e^{\frac{s}{2\sqrt{2}}}$ is a strictly monotonous function; must exist a unique λ_{nm}, namely $\lambda_{nm} = e^{\frac{s^*}{2\sqrt{2}}}$, such that $\Phi(\lambda_{nm}) = \min_{\lambda>0} \Phi(\lambda)$. □

In fact, this result guarantees the unicity of the MIRE, although a numerical analysis is required to obtain it explicitly (see next section). It could be useful to develop a simple approximate estimator, that shall be referred hereafter a-MIRE, obtained, luckily, minimizing a convenient upper bound of $\Phi(\lambda)$. Since

$$1 + \frac{1}{2!}\frac{\rho^2}{2} \leq \cosh(\frac{\rho}{2}) = \frac{1}{4}\, d_M^2(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) + \frac{\sigma_1^2 + \sigma_2^2}{2\,\sigma_1\sigma_2} \tag{26}$$

we shall have

$$0 \leq \rho^2(\boldsymbol{U}_\lambda(\boldsymbol{y}), (\mathbf{0}, 1)) \leq \frac{1}{n\,\lambda}\frac{\mathcal{V}^2}{\mathcal{W}} + 2\left(\sqrt{\lambda \mathcal{W}} - \frac{1}{\sqrt{\lambda \mathcal{W}}} \right)^2 \tag{27}$$

and therefore

$$0 \leq \Phi(\lambda) \leq H(\lambda) = \frac{1}{\lambda}\frac{\Gamma(\frac{n-m}{2} - \frac{1}{2})}{\sqrt{2}\,\Gamma(\frac{n-m}{2})} \left(\frac{m}{n} + 2 \right) + 2\sqrt{2}\,\frac{\Gamma(\frac{n-m}{2} + \frac{1}{2})}{\Gamma(\frac{n-m}{2})} \tag{28}$$

the upper bound $H(\lambda)$ it is clearly a convex functions with an absolute minimum attained when λ satisfy

$$\widetilde{\lambda_{nm}} = \left(\sqrt{1 + \frac{m}{2n}} \right) \frac{1}{\sqrt{n - m - 1}} \tag{29}$$

Furthermore, given an arbitrary m, we have

$$\lim_{n\to\infty} \left(\widetilde{\lambda_{nm}} \right)^2 n = 1 \tag{30}$$

and, therefore, a-MIRE is very close to MLE for large values of n. Observe also that it is possible to compute a-MIRE for $n > m + 1$, a condition which is slightly stronger that the result required for the existence of MIRE in proposition (4).

A further aspect is the intrinsic bias of the equivariant estimators. In fact connections between minimum risk, bias and invariance have been established, see [14]. Since the action of the group G is

 doi:10.20944/preprints202510.0742.v1

not commutative, we cannot guarantee the unbiasedness of the MIRE and an additional analysis must be performed. First of all we are going to compute the vector bias, see [4], a quantitative measure of the bias which is compatible with Lehmann results.

Let $\boldsymbol{A}_\theta(\lambda) = \exp_\theta^{-1}(\boldsymbol{U}_\lambda(\boldsymbol{y}))$ and $A_\theta^1(\lambda), \ldots, A_\theta^m(\lambda), A_\theta^{m+1}(\lambda)$ be the components of $\boldsymbol{A}_\theta(\lambda)$ corresponding to the basis field $(\partial/\partial\beta^1)_\theta, \ldots, (\partial/\partial\beta^m)_\theta, (\partial/\partial\sigma)_\theta$. With matrix notation, $\boldsymbol{A}_{\theta 1}(\lambda) = (A_\theta^1(\lambda), \ldots, A_\theta^m(\lambda))^{\mathrm{t}}$. Furthermore, let us define $h(x) \equiv x/\sinh x$ for $x \neq 0$ and $h(0) \equiv 0$; taking into account (16), (19) and from (11) and (15) we have

$$\begin{aligned} \boldsymbol{X}\boldsymbol{A}_{\theta 1}(\lambda) &= f_\lambda(\mathcal{V}, \mathcal{W}))\ \boldsymbol{z}\ \frac{\sigma}{\lambda\mathcal{W}} \\ A_\theta^{m+1}(\lambda) &= f_\lambda(\mathcal{V}, \mathcal{W})\ g_\lambda(\mathcal{V}, \mathcal{W})\ \sigma \end{aligned} \tag{31}$$

where

$$f_\lambda(\mathcal{V}, \mathcal{W}) \equiv h\left(\rho(\boldsymbol{U}_\lambda(\boldsymbol{y}), \theta)/\sqrt{2}\right)$$

and

$$g_\lambda(\mathcal{V}, \mathcal{W}) \equiv \cosh\left(\rho(\boldsymbol{U}_\lambda(\boldsymbol{y}), \theta)/\sqrt{2}\right) - \frac{1}{\lambda\mathcal{W}}$$

Let $\boldsymbol{B}_\theta(\lambda) = E_\theta(\exp_\theta^{-1}(\boldsymbol{U}_\lambda(\boldsymbol{y})))$ be the intrinsic bias vector corresponding to an equivariant estimator $\boldsymbol{U}_\lambda(\boldsymbol{y}) = (\boldsymbol{U}_{\lambda 1}(\boldsymbol{y}), U_{\lambda 2}(\boldsymbol{y}))$ evaluated at the point $\theta = (\boldsymbol{\beta}, \sigma)$ and let $B_\theta^1(\lambda), \ldots, B_\theta^m(\lambda)$, $B_\theta^{m+1}(\lambda)$ be their components. In matrix notation, $\boldsymbol{B}_{\theta 1}(\lambda) = (B_\theta^1(\lambda), \ldots, B_\theta^m(\lambda))^{\mathrm{t}}$. We have

Proposition 6. *If $n \geq m+1$, the bias vector is finite and*

$$\begin{aligned} \boldsymbol{B}_{\theta 1}(\lambda) &= \boldsymbol{0} \\ B_\theta^{m+1}(\lambda) &= E(f_\lambda(\mathcal{V}, \mathcal{W})\ g_\lambda(\mathcal{V}, \mathcal{W}))\ \sigma \end{aligned} \tag{32}$$

where $\mathcal{V}^2$ and $\mathcal{W}^2$ are independent random variables following a chi-square distribution with m and $n-m$ degrees of freedom respectively.

Moreover, the square of the norm of the bias vector is constant and given by

$$\|\boldsymbol{B}_\theta(\lambda)\|_\theta^2 = 2\left(\ E(\ f_\lambda(\mathcal{V}, \mathcal{W})\ \ g_\lambda(\mathcal{V}, \mathcal{W})\ )\ \right)^2 \tag{33}$$

Proof: Observe that if $n \geq m+1$ we have

$$\begin{aligned} \|\boldsymbol{B}_\theta(\lambda)\|_\theta^2 &\leq E_\theta(\|\exp_\theta^{-1}(\boldsymbol{U}_\lambda(\boldsymbol{y}))\|_\theta)^2 \leq E_\theta(\|\exp_\theta^{-1}(\boldsymbol{U}_\lambda(\boldsymbol{y}))\|_\theta^2) \\ &= E_\theta(\rho^2(\boldsymbol{U}_\lambda(\boldsymbol{y}), \theta)) < \infty \end{aligned}$$

where $\|\ \cdot\ \|_\theta$ denotes the Riemannian norm at the tangent space at θ.

On the other hand taking into account (31) and defining $\boldsymbol{z}$ as in (16) observe that $\mathcal{V} = \|\boldsymbol{z}\| = \|-\boldsymbol{z}\|$, $\boldsymbol{z}$ is independent of $\mathcal{W}$ and $\boldsymbol{z}$ and $-\boldsymbol{z}$ has the same distribution. Then we have

$$\begin{aligned} \boldsymbol{X}\boldsymbol{B}_{\theta 1}(\lambda) &= E_\theta\Big(f_\lambda(\mathcal{V}, \mathcal{W})\ \boldsymbol{z}\ \frac{\sigma}{\lambda\mathcal{W}}\Big) \\ &= E_\theta\Big(f_\lambda(\mathcal{V}, \mathcal{W})\ (-\boldsymbol{z})\ \frac{\sigma}{\lambda\mathcal{W}}\Big) = -E_\theta\Big(f_\lambda(\mathcal{V}, \mathcal{W})\ \boldsymbol{z}\ \frac{\sigma}{\lambda\mathcal{W}}\Big) = \boldsymbol{0} \end{aligned}$$

and since $\mathrm{rank}(\boldsymbol{X}) = m$ it follows that $\boldsymbol{B}_{\theta 1}(\lambda) = \boldsymbol{0}$.

$B_\theta^{m+1}(\lambda)$ is obtained directly from (31). The distribution of $\boldsymbol{z}$ and $\mathcal{W}^2$ follow from basic properties of multivariate normal distribution. Finally, the norm of the bias vector field follows from (32) and (8). □

 doi:10.20944/preprints202510.0742.v1

We may remark, finally, that the norm of the bias vector field of any equivariant estimator, $\|\boldsymbol{B}_\theta(\lambda)\|_\theta$, is invariant under the action of the induced group, $\overline{\mathcal{G}}$, on the parameter space and since this group acts transitively on Θ, this quantity must be constant, which is clear from (33).

4. Numerical Evaluation

In this section we are going to compare, numerically, the MIRE estimator with the standard MLE. Observe that both estimators only differ in the estimation of the parameter σ (or σ^2). Precisely, the MIRE and the MLE of σ^2 are, respectively,

$$\widehat{\sigma^2} = \lambda_{nm}^2\, \boldsymbol{y}^{\mathrm{t}}\Big(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{t}}\Big)\boldsymbol{y} \qquad \overset{*}{\sigma}{}^2 = \frac{1}{n}\,\boldsymbol{y}^{\mathrm{t}}\Big(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{t}}\Big)\boldsymbol{y} \tag{34}$$

namely, they only differ by the factor $\xi_{nm} = \lambda_{nm}^2 n$, since $\widehat{\sigma^2} = \xi_{nm}\overset{*}{\sigma}{}^2$.

In order to compare MLE with the MIRE we have computed the factor ξ_{nm}, the intrinsic risk and the square of the norm of the bias vector for each estimators. All the computations have been performed using Mathematica 10.2.

Moreover, we have suggested a rather simple approximation for ξ_{nm}, or λ_{nm}, which allow us to approximate MIRE estimator through (29), i.e.

$$\widetilde{\xi_{nm}} = \left(1 + \frac{m}{2n}\right)\frac{n}{n-m-1} \tag{35}$$

The corresponding estimator, that shall be referred hereafter as a-MIRE, has been also compared with MLE and MIRE in terms of intrinsic risk.

The results are summarized in the following figures which summarize the tables given in the Appendix, see [16] for a convincing argument for the use of plots over tables. In Figure 2 numerical results of ξ_{nm} (left) and $\Phi(\lambda_{nm}) = \Phi(\sqrt{\xi_{nm}/n})$ (right) are displayed graphically. Observe that, for m fixed, as n increases $\Phi(\sqrt{\xi_{nm}/n})$ goes to zero and ξ_{nm} goes to one. The exact numerical values are given in the Appendix, Table 1.

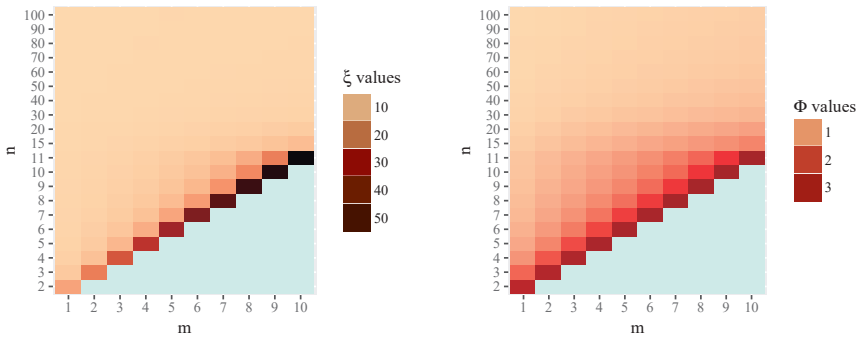


Figure 1. Numerical results for ξ_{nm} (left) and $\Phi(\lambda_{nm})$ (right), with $\xi_{nm} = \lambda_{nm}^2 n$.

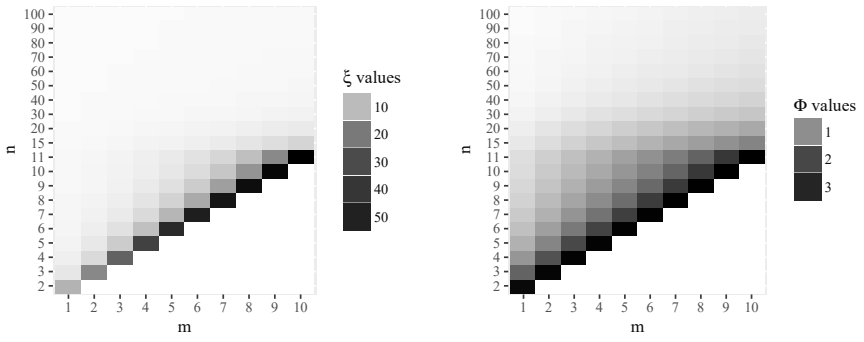


Figure 2. Numerical results for ξ_{nm} (left) and $\Phi(\lambda_{nm})$ (right), with $\xi_{nm} = \lambda_{nm}^2 n$.

 doi:10.20944/preprints202510.0742.v1

Figure 4 shows graphically the percentage of intrinsic risk increment for MLE (left) and a-MIRE (right), that is

$$100\ \frac{\left(\Phi(\frac{1}{\sqrt{n}}) - \Phi(\lambda_{nm})\right)}{\Phi(\lambda_{nm})} \qquad \text{and} \qquad 100\ \frac{\left(\Phi(\widetilde{\lambda_{nm}}) - \Phi(\lambda_{nm})\right)}{\Phi(\lambda_{nm})} \tag{36}$$

respectively, for some values of n and m. The exact numerical results are given in the Appendix, Table 2.

Observe that if we approximate the MIRE by the MLE for a certain value of m, the relative difference of risk decreases as n increases. Notice that these relative differences of risks are rather moderate or small (about 10-15%) only if $n > 10m$. Let us remark also the non appropriate behavior of the MLE for small values of $n - m$. This regards to the intrinsic risk increment: when we use MLE instead of MIRE this increment oscillates between $80,5\%$ and $251,1\%$ when $n - m = 1$ with m from 1 to 10. On the other hand, the behavior of the a-MIRE is reasonably good, with its intrinsic risk very similar to the risk of the MIRE estimator: here the percentage on risk increment is less than 1% for all studied cases and also this percentage decreases as n increases. In fact this percentage is lower than 1‰ when $n - m \geq 4$, which indicates the extraordinary degree of approximation, being therefore the a-MIRE a reasonable and useful approximation of MIRE. In particular as an example, for a two way analysis of variance, with a and b levels for factors A and B respectively, with a single replicate for treatment we shall have

$$\widetilde{\lambda} = \left(\sqrt{1 + \frac{a+b-1}{2ab}}\right) \frac{1}{\sqrt{ab-a-b}} \tag{37}$$

with $a > 1$ and $b > 1$, while the corresponding quantity for MLE is $\lambda^* = \frac{1}{\sqrt{ab}}$.

In particular if $a = 3$ and $b = 5$ we shall have

$$\widetilde{\lambda} = \left(\sqrt{1 + \frac{7}{30}}\right) \frac{1}{\sqrt{7}} = 0,419750$$

and $\lambda^* = \frac{1}{\sqrt{15}} = 0,258199$, which is sensibly different. At this moment, it could be useful to recall that the quadratic loss and the squared of the Riemannian distance behaves very different, see [5].

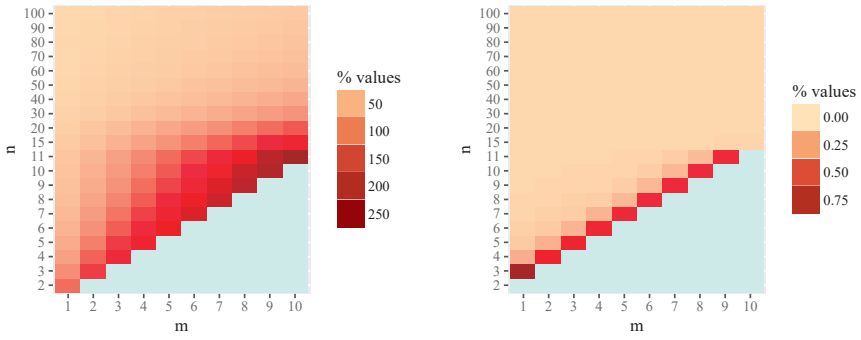


Figure 3. Percentage of intrinsic risk increment for MLE (left) and a-MIRE (right).

 doi:10.20944/preprints202510.0742.v1

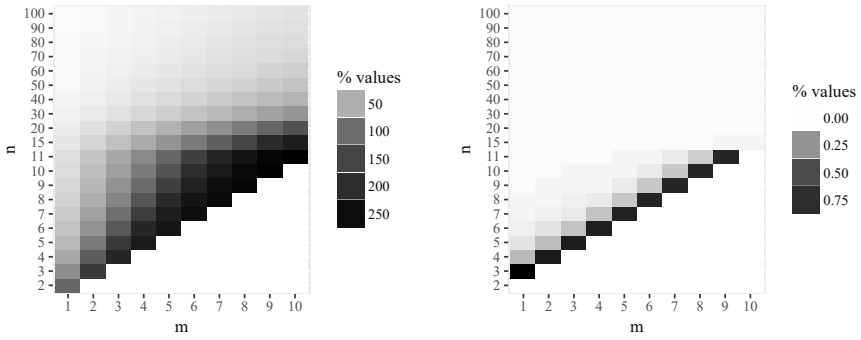


Figure 4. Percentage of intrinsic risk increment for MLE (left) and a-MIRE (right).

Figure 6 displays graphically the numerical results of the $m+1$ component of the bias vector divided by σ, i.e. the unique non zero physical component of this vector field, for MIRE (left) and MLE (right), $\frac{1}{\sigma} B_\theta^{m+1}(\lambda_{nm})$ and $\frac{1}{\sqrt{n}\,\sigma} B_\theta^{m+1}(\frac{1}{\sqrt{n}})$ respectively, for some values of n and m. Observe the sign differences, meaning that MIRE overestimates, in average, σ while MLE underestimates this quantity, in average as well. This follows from the equation of the geodesics of the present model (9). The exact numerical values are given in the Appendix, Table 3.

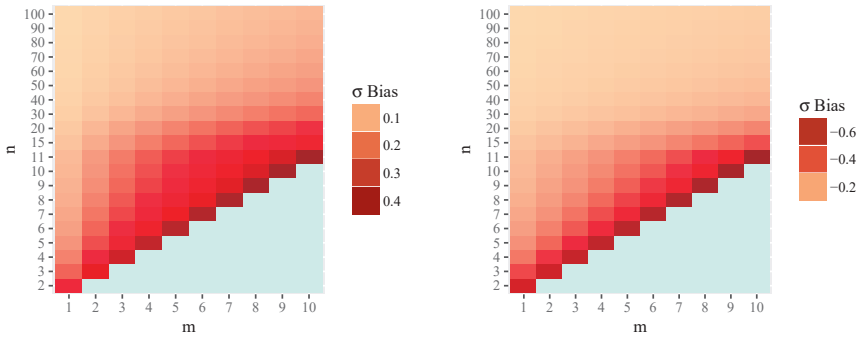


Figure 5. Numerical results for the unique non–null physical component of bias vector field $\frac{1}{\sigma} B_\theta^{m+1}(\lambda)$, for MIRE (left, with $\lambda = \lambda_{nm}$) and MLE (right, with $\lambda = 1/\sqrt{n}$).

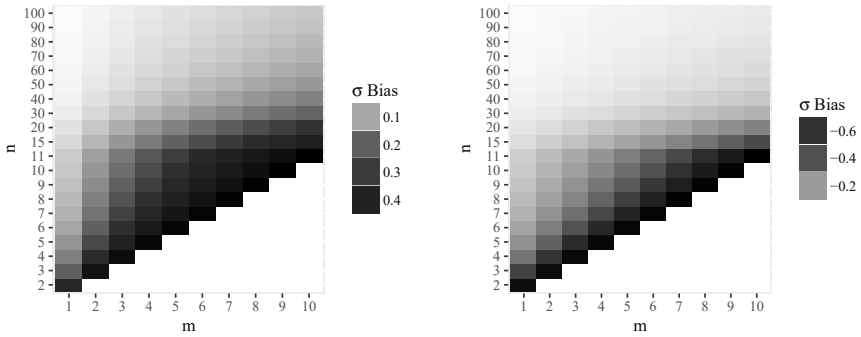


Figure 6. Numerical results for the unique non–null physical component of bias vector field $\frac{1}{\sigma} B_\theta^{m+1}(\lambda)$, for MIRE (left, with $\lambda = \lambda_{nm}$, positive values) and MLE (right, with $\lambda = 1/\sqrt{n}$, negative values).

Figure 8 shows graphically the numerical results of the percentage of intrinsic risk due to bias for MIRE and MLE, that is

$$100\ \frac{\|\boldsymbol{B}\|_\theta^2(\lambda_{nm})}{\Phi(\lambda_{nm})} \qquad \text{and} \qquad 100\ \frac{\|\boldsymbol{B}\|_\theta^2(\frac{1}{\sqrt{n}})}{\Phi(\lambda_{nm})} \tag{38}$$

respectively, for some values of n and m. Observe that the bias is moderate, with respect the intrinsic risk, in both estimators. The bias of MIRE estimator is smaller than the bias of MLE for small values of m and the opposite for large values. The exact numerical results are given in the Appendix, Table 4.

 doi:10.20944/preprints202510.0742.v1

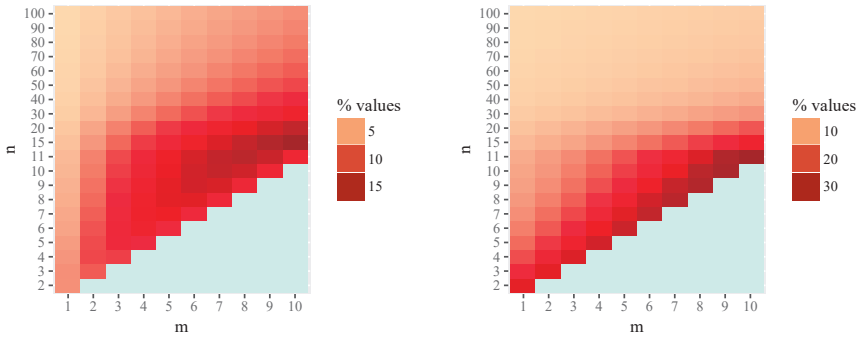


Figure 7. Percentage of intrinsic risk due to bias for MIRE (left) and MLE (right), referred to the risk of MIRE.

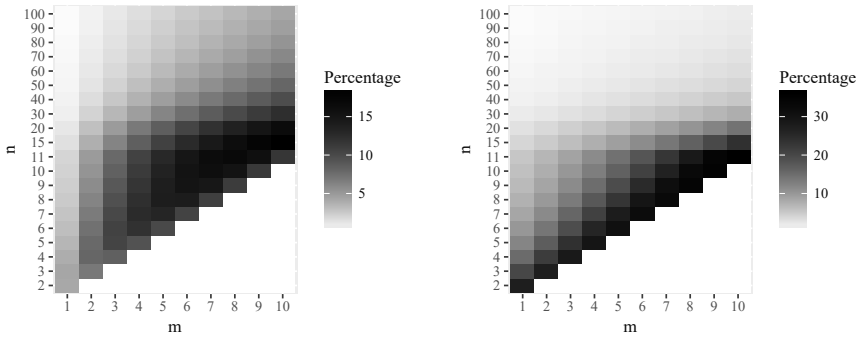
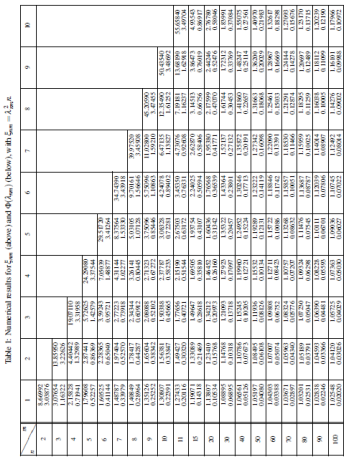


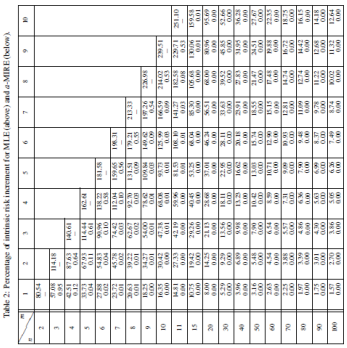
Figure 8. Percentage of intrinsic risk due to bias for MIRE (left) and MLE (right), referred to the risk of MIRE.

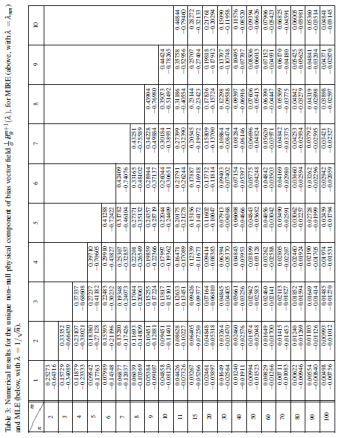
Acknowledgements: We have to thank the referees and the editor comments and suggestions for the improvement of this paper.

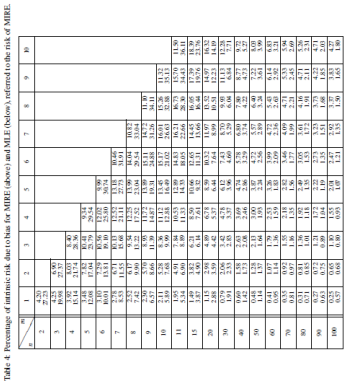
5. Appendix



 doi:10.20944/preprints202510.0742.v1







References

1. Rao, C. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **1945**, *37*, 81–91.
2. Burbea, J.; Rao, C. Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. *J. Multivar. Anal.* **1982**, *12*, 575–596.
3. Burbea, J. Informative geometry of probability spaces. *Expo. Math.* **1986**, *4*, 347–378.
4. Oller, J.; Corcuera, J. Intrinsic Analysis of the Statistical Estimation. *Ann. Stat.* **1995**, *23*, 1562–1581.
5. García, G.; Oller, J. What does intrinsic mean in statistical estimation? *Sort* **2006**, *2*, 125–146.
6. Bernal-Casas, D.; Oller, J. Variational Information Principles to Unveil Physical Laws. *Mathematics* **2024**, *12*, 3941. https://doi.org/10.3390/math12243941.
7. Muirhead, R. *Aspects of Multivariate Statistical Theory*; John Wiley & Sons Inc.: New York, NY, USA, 1982.
8. Bernardo, J.; Juárez, M. Intrinsic estimation. In *Bayesian Statistics 7*; Bernardo, J.; Bayarri, M.; Berger, J.; Dawid, A.; Hackerman, W.; Smith, A.; West, M., Eds.; Oxford University Press: Berlin, Germany, 2003; pp. 465–476.

 doi:10.20944/preprints202510.0742.v1

9. Lehmann, E.; Casella, G. *Theory of Point Estimation*; Springer Science & Business Media: New York, NY, USA, 2006.
10. ichi Amari, S. *Information Geometry and Its Applications*; Springer: Tokyo, 2016.
11. Ay, N.; Jost, J.; Lê, H.V.; Schwachhöfer, L. *Information Geometry*; Springer: Cham, 2017.
12. Nielsen, F. An Elementary Introduction to Information Geometry. *Entropy* **2020**, *22*, 1100. https://doi.org/10.3390/e22101100.
13. Burbea, J.; Oller, J. The information metric for univariate linear elliptic models. *Stat. Decis.* **1988**, *6*, 209–221.
14. Lehmann, E. A general concept of unbiasedness. *Ann. Math. Stat.* **1951**, *22*, 587–592.
15. Karcher, H. Riemannian Center of Mass and Mollifier Smoothing. *Commun. Pure Appl. Math.* **1977**, *30*, 509–541.
16. Gelman, A.; Pasarica, C.; Dodhia, R. Let's practice what we preach: turning tables into graphs. *Am. Stat.* **2002**, *56*, 121–130.