

Hellinger Information Matrix and Hellinger Priors

Arkady Shemyakin

Mathematics Department, University of St. Thomas, 2115 Summit Ave, St. Paul, MN 55105, USA;
a9shemyakin@stthomas.edu

Abstract: Hellinger information as a local characteristic of parametric distribution families was first introduced in 2011. It is related to the much older concept of the Hellinger distance between two points in a parametric set. Under certain regularity conditions, the local behavior of the Hellinger distance is closely connected to Fisher information and the geometry of Riemann manifolds. Nonregular distributions (non-differentiable distribution densities, undefined Fisher information or densities with support depending on the parameter), including uniform, require using analogues or extensions of Fisher information. Hellinger information may serve to construct information inequalities of the Cramer–Rao type, extending the lower bounds of the Bayes risk to the nonregular case. A construction of non-informative priors based on Hellinger information was also suggested by the author in 2011. Hellinger priors extend the Jeffreys rule to nonregular cases. For many examples, they are identical or close to the reference priors or probability matching priors. Most of the paper was dedicated to the one-dimensional case, but the matrix definition of Hellinger information was also introduced for higher dimensions. Conditions of existence and the nonnegative definite property of Hellinger information matrix were not discussed. Hellinger information for the vector parameter was applied by Yin et al. to problems of optimal experimental design. A special class of parametric problems was considered, requiring the directional definition of Hellinger information, but not a full construction of Hellinger information matrix. In the present paper, a general definition, the existence and nonnegative definite property of Hellinger information matrix is considered for nonregular settings.

Keywords: Hellinger information; Hellinger priors; Bayes risk



Citation: Shemyakin, A. Hellinger Information Matrix and Hellinger Priors. *Entropy* **2023**, *25*, 344.
<https://doi.org/10.3390/e25020344>

Academic Editors: Geert Verdoolaeghe and Michal Branicki

Received: 20 December 2022

Revised: 16 January 2023

Accepted: 6 February 2023

Published: 13 February 2023



Copyright: © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Information geometry plays an important role in parametrical statistical analysis. Fisher information is the most common information measure that is instrumental in the construction of lower bounds for quadratic risk (information inequalities of Cramer–Rao type), optimal experimental designs (E-optimality) and noninformative priors in Bayesian analysis (Jeffreys prior). These applications of Fisher information require certain regularity conditions on the distributions of the parametric family, which include the existence and integrability of the partial derivatives of the distribution density function with respect to the components of vector parameter and independence of the density support on the parameter. If these regularity conditions are not satisfied, Cramer–Rao lower bounds might be violated, and Jeffreys prior might not be defined.

There exist a number of ways to define information quantities (for the scalar parameter case) and matrices (for the vector parameter case) in the nonregular cases when Fisher information might not exist. One of such suggestions is the Wasserstein information matrix [1], which has been recently applied to the construction of objective priors [2]. The Wasserstein matrix does not require the differentiability of the distribution density function, but it cannot be extended to the case of discontinuous densities. This latter case requires a more general definition of information.

Our approach is based on analyzing the local behavior of parametric sets using finite differences of pdf values at two adjacent points instead of derivatives at a point, which allows us to include differentiable densities as a special case but also to treat non-differentiable

densities including jumps and other types of nonregular behavior (for classification of nonregularities, see [3]). A logical approach is to use (in lieu of Fisher information) the Hellinger information closely related to the definition of Hellinger distance between adjacent points of the parametric set.

Hellinger information for the case of scalar parameter was first defined in [4] and suggested for the construction of noninformative Hellinger priors. Section 2 is dedicated to the revision of this definition and the relationship of Hellinger information to the information inequalities for the scalar parameter proven in [5,6]. It contains some examples of noninformative Hellinger priors comparing the priors obtained in [4] with more recent results.

Lin et al. [7] extended the definition of Hellinger information to a special multiparameter case, where all components of a parameter expose the same type of nonregularity. This is effective in the resolution of some optimization problems in experimental design. However, the most interesting patterns of local behavior of Hellinger distance bringing about differences in the behavior of matrix lower bounds of the quadratic risk and multidimensional noninformative priors are observed when the parametric distribution family has different orders of nonregularity [3] for different components of the vector parameter. Thus, the main challenge in the construction of a Hellinger information matrix in the general case consists in the necessity to consider different magnitudes of increments in different directions of the vector parametric space.

A general definition of the Hellinger information matrix was attempted in [6] as related to information inequalities and in [4] as related to noninformative priors. Important questions were left out, such as the conditions of Hellinger information matrix being positive definite and the existence of non-trivial matrix lower bounds for the quadratic risk in case of the vector parameter. These questions are addressed in Section 3 of the paper. The main results are formulated, and several new examples are considered. General conclusions and possible future directions of study are discussed in Section 4.

2. Hellinger Information for Scalar Parameter

In this section, we address the case of probability measures parametrized by a single parameter. We provide necessary definitions of information measures along with the discussion of their properties, including the new definition of Hellinger information. Then, we consider the applications of Hellinger information to the information inequalities of the Cramér–Frechet–Rao type, the construction of objective priors, and problems of optimal design.

Definition (2) is modified from [8]. Inequality (5) was obtained in [5]. Examples in Section 2.3 are modified from [4].

2.1. Definitions

A family of probability measures $\{P_\theta, \theta \in \Theta \subset \mathbf{R}\}$ is defined on a measurable space $(\mathfrak{X}, \mathfrak{B})$ so that all the measures from the family are absolutely continuous with respect to some σ -finite measure λ on \mathfrak{B} . The square of the Hellinger distance between any two parameter values can be defined in terms of densities $p(x; \theta) = \frac{dP_\theta}{d\lambda}$ as

$$h^2(\theta_1, \theta_2) = \int_{\mathfrak{X}} \left(\sqrt{p(x; \theta_1)} - \sqrt{p(x; \theta_2)} \right)^2 \lambda(dx). \tag{1}$$

This definition of the Hellinger distance (also known as Hellinger–Bhattacharyya distance) in its modern form was given in [9]. We use this definition to construct a new information measure. If for almost all θ from Θ (with regard to measure λ) there exists an $\alpha \in (0, 2]$ (**index of regularity**) such that

$$\lim_{\epsilon \rightarrow 0} \frac{h^{2/\alpha}(\theta, \theta + \epsilon)}{|\epsilon|^2} = J(\theta), \tag{2}$$

we define **Hellinger information** at a point θ as $J(\theta)$. The index of regularity is related to the local behavior of the density $p(x; \theta)$. Using the classification of [3], $1 < \alpha < 2$ corresponds to singularities of the first and the second type, $\alpha = 1$ to densities with jumps, and $0 < \alpha < 1$ to singularities of the third type.

Notice that in the regular situations classified in [3] ($p(x; \theta)$ is twice continuously differentiable with respect to θ for almost all $x \in \mathfrak{X}$ with respect to λ , the density support $\{x : p(x; \theta) > 0\}$ does not depend on parameter θ , Fisher information

$$I(\theta) = E_{\theta} \left(\frac{d}{d\theta} \log p(x; \theta) \right)^2 = \int_{\mathfrak{X}} \left(\frac{d}{d\theta} p(x; \theta) \right)^2 / p(x; \theta) \lambda(dx) \tag{3}$$

is continuous, strictly positive and finite for almost all θ from Θ , it is true that $\alpha = 2, J(\theta) = \frac{1}{4}I(\theta)$. Under the regularity conditions above, the score function $\frac{d}{d\theta} \log p(x; \theta)$ has mean zero and Fisher information as variance. This helps to establish the connection of Fisher information to the limiting distribution of maximum likelihood estimators, its additivity with respect to i.i.d. sample observations, and its role in the lower bounds of risk (information inequalities).

Wasserstein information [1] can be defined for the scalar parameter through the c.d.f. $F(x; \theta) = \int_{-\infty}^x p(x; \theta) d\theta$ as

$$W(\theta) = \int_{\mathfrak{X}} \left(\frac{d}{d\theta} F(x; \theta) / p(x; \theta) \right)^2 \lambda(dx),$$

which does not require differentiability of the density function $p(x; \theta)$. That opens new possibilities for the construction of an objective prior in the case of non-differentiable densities, see [2].

However, we are interested in even less regular situations (including uniform densities with support depending on parameter) for which neither Fisher information $I(\theta)$ nor Wasserstein information $W(\theta)$ can be helpful, while Hellinger information $J(\theta)$ may function as their substitute.

2.2. Information Inequalities

We define the quadratic Bayes risk for an estimator $\theta^* = \theta^*(X^{(n)})$ constructed by an independent identically distributed sample $X^{(n)} = (X_1, \dots, X_n)$ of size n from the model considered above with $p(x^{(n)}; \theta) = \frac{dP_{\theta}^{(n)}}{d\lambda^n}$ and prior $\pi(\theta)$ as

$$R(\theta^*) = \int_{\mathfrak{X}^n} \int_{\Theta} \left(\theta^*(x^{(n)}) - \theta \right)^2 p(x^{(n)}; \theta) \pi(\theta) \lambda^n(dx) d\theta.$$

Let us consider an integral version of the classical Cramér–Frechet–Rao inequality, which under certain regularity conditions leads to the following asymptotic lower bound for the Bayes risk in terms of Fisher information:

$$\inf_{\theta^*} R(\theta^*) \geq n^{-1} \int_{\Theta} I^{-1}(\theta) \pi(\theta) d\theta + O(n^{-2}). \tag{4}$$

This lower bound, which can be proven to be tight, was first obtained by [10], also in [11,12] under slightly different regularity assumptions. This bound can be extended to the nonregular case, when Fisher information may not exist. One of these extensions is Hellinger information inequality, providing an asymptotic lower bound

$$\inf_{\theta^*} R(\theta^*) \geq n^{-2/\alpha} C(\alpha) \int_{\Theta} J^{-1}(\theta) \pi(\theta) d\theta + O(n^{-2/\alpha}) \tag{5}$$

obtained in [6] under the assumptions of Hellinger information $J(\theta)$ being strictly positive, almost surely continuous, bounded on any compact subset of Θ , and satisfying condition

$\int_{\Theta} J^{-1}(\theta)\pi(\theta)d\theta < \infty$, where Θ is an open subset of real numbers and the constant $C(\alpha)$ is related to technical details of the proof and is not necessarily tight.

The key identity establishing the role of $J(\theta)$ in the case of i.i.d. samples $x^{(n)} = (x_1, \dots, x_n)$

$$h_n^2(\theta_1, \theta_2) = \int_{\mathcal{X}} \left(\sqrt{p(x^{(n)}; \theta_1)} - \sqrt{p(x^{(n)}; \theta_2)} \right)^2 \lambda^n(dx) = 2 \left[1 - \left(1 - \frac{1}{2} h^2(\theta_1, \theta_2) \right)^n \right] \tag{6}$$

easily follows from the definition and independence of $\{X_i\}$. Similar to the additivity of Fisher information, it allows for a transition from a single observation to a sample.

2.3. Hellinger Priors

The three most popular ways to obtain a non-informative prior might be as follows:

- The Jeffreys rule $\pi(\theta) \propto \sqrt{I(\theta)}$ [13],
- Probability matching priors [14,15],
- Reference priors [16,17].

For many regular parameter families, probability matching and reference priors both satisfy the Jeffreys rule. However, it is not necessary in case of multi-parametric families and the loss of regularity. Let us focus on the nonregular case, when Fisher information may not be defined. Most comprehensive results on reference priors in the nonregular case were obtained in [18].

Define Hellinger prior for the parametric set as in [4,8]:

$$\pi_H(\theta) \propto \sqrt{J(\theta)}. \tag{7}$$

Hellinger priors will often coincide with well-known priors obtained by the approaches described above. However, there are some distinctions. A special role might be played by Hellinger priors in nonregular cases. We provide two simple examples of densities with support depending on the parameter.

Example 1. Uniform $X \sim \text{Unif}(\theta^{-1}, \theta), \theta \in (1, \infty)$.

$$\alpha = 1, \pi_H(\theta) \propto J(\theta) = \frac{\theta^2 + 1}{\theta(\theta^2 - 1)}.$$

The same prior can be constructed as the probability matching prior g or the reference prior [18].

Example 2. Uniform $X \sim \text{Unif}(\theta, \theta^2), \theta \in (1, \infty)$.

$$\alpha = 1, \pi_H(\theta) \propto J(\theta) = \frac{2\theta + 1}{\theta(\theta - 1)}. \tag{8}$$

This prior is different from the reference prior obtained in [18] by rather technical calculations, maximizing the Kullback–Leibler divergence from the prior to the posterior:

$$\pi_{R1}(\theta) \propto \frac{2\theta - 1}{\theta(\theta - 1)} \exp \left\{ \psi \left(\frac{2\theta}{2\theta - 1} \right) \right\}, \tag{9}$$

where $\psi(z) = \frac{d}{dz} \log \Gamma(z), z > 0$ is the polygamma function order 1. Tri Minh Le [19] used a similar approach, maximizing Hellinger distance between the prior and the posterior, and obtained

$$\pi_{R2}(\theta) \propto \frac{(2\theta + 1)^2(2\theta - 1)^5}{\theta^3(\theta - 1) [20\sqrt{\pi}(2\theta - 1)^2 - 2(2\theta + 1)^2]^2}. \tag{10}$$

All three priors in (8), (9), and (10) have distinct functional forms. However, they are very close after appropriate re-normalization on the entire domain, which can be demonstrated graphically and numerically. See Figure 1 and the following comment.

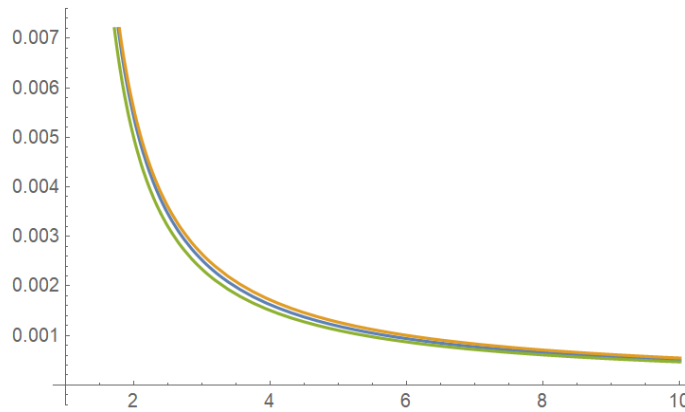


Figure 1. Three priors for Example 2: $\pi_H(\theta), \pi_{R1}(\theta), \pi_{R2}(\theta)$.

For instance, the ratio $\pi_H(\theta) / \pi_{R2}(\theta)$ monotonically increases for θ from 1 to ∞ so that $0.996 < \pi_H(\theta) / \pi_{R2}(\theta) < 1.083$.

2.4. Optimal Design

A polynomial model of the experimental design may be presented as in [20],

$$y_i = \sum_{k=0}^q \theta_k x_i^k + \epsilon_i, i = 1, \dots, n,$$

where x_i are scalars, θ is the unknown vector parameter of interest, and errors ϵ_i are non-negative i.i.d variables with density $p_0(y; \alpha) \sim \alpha c(\alpha) y^{\alpha-1}$ (e.g., Weibull or Gamma). The space of balanced designs is defined as

$$\Xi = \{ \xi = (w_i, x_i), i = 1, \dots, n : \sum_{i=1}^n w_i x_i = 0, x_i \in [-A, A] \},$$

and there exist several definitions of optimal design. Lin et al. [7] suggest using criterion

$$\xi_{opt} = \arg \max_{\xi} J_{\xi}(\theta), J_{\xi}(\theta) = \inf_u J_{\xi}(\theta, u)$$

where $J_{\xi}(\theta, u)$ is Hellinger information in the direction $u \in \mathbf{R}^{q+1}$, defined as

$$\lim_{\epsilon \rightarrow 0} \frac{h(\theta, \theta + \epsilon u)}{|\epsilon|^{\alpha}} = J_{\xi}(\theta, u), \tag{11}$$

which is similar to the definition given in Section 1, but notice the difference with (2) in the treatment of powers: ϵ^{α} versus ϵ^2 in the denominator. Notice also that index of regularity α is assumed to be the same for all components of the vector parameter.

3. Hellinger Information Matrix for Vector Parameter

In this section, we concentrate on the multivariate parameter case allowing for different degrees of regularity for different components of the vector parameter. We define the Hellinger information matrix, determine our understanding of matrix information inequalities, formulate main results establishing lower bounds for the Bayes risk in terms of Hellinger information matrix, and provide examples of Hellinger priors illustrating the conditions of Theorems 1 and 2.

Proofs of the main results use the approach developed in [5]; Example 3 was previously mentioned in [8].

3.1. Definitions

Extending definitions of Section 1 to the vector case $\Theta \subset \mathbf{R}^m, m = 1, 2, \dots$, we first introduce, as in [6], the *Hellinger distance matrix* H with elements

$$H_{ij}(\theta, U) = \int_{\mathfrak{X}} \left(\sqrt{p(x; \theta)} - \sqrt{p(x; \theta + \mathbf{u}_i)} \right) \left(\sqrt{p(x; \theta)} - \sqrt{p(x; \theta + \mathbf{u}_j)} \right) \lambda(dx) \quad (12)$$

where increments \mathbf{u}_i are columns of an $m \times m$ matrix U . Define also vectors $\alpha = (\alpha_1, \dots, \alpha_m)'$ (**index of regularity** with components $0 < \alpha_i \leq 2$) and $\delta = (\delta_1, \dots, \delta_m)'$ with components $\delta_i = \epsilon^{2/\alpha_i}, \Delta = \text{Diag}(\delta)$ such that for all $i = 1, \dots, m$ there exist finite non-degenerate limits

$$0 < \lim_{\epsilon \rightarrow 0} |\epsilon|^{-2} H_{ii}^{2/\alpha_i}(\theta, \Delta) < \infty$$

Then, the **Hellinger information matrix** will be defined by its components

$$J_{ij}(\theta) = \lim_{\epsilon \rightarrow 0} |\epsilon|^{-2} H_{ij}^{1/\alpha_i + 1/\alpha_j}(\theta, \Delta). \quad (13)$$

Notice that components of the vector index of regularity α can be different, and therefore, components of the vector of increments δ can have different orders of magnitude with respect to ϵ . As a result, while the elements of matrix $H(\theta, \Delta)$ may expose different local behavior depending on the components of α , the elements of matrix $J(\theta)$ are all finite.

3.2. Information Inequalities

Define the matrix of Bayes risk for an i.i.d. sample $x^{(n)} = (x_1, \dots, x_n)$ with $p(x^{(n)}; \theta) = \frac{dP_{\theta}^{(n)}}{d\lambda^n}$ as

$$\mathbf{R}(\theta^*) = \int_{\mathfrak{X}^{(n)}} \int_{\Theta} \left(\theta^*(x^{(n)}) - \theta \right) \left(\theta^*(x^{(n)}) - \theta \right)' p(x^{(n)}; \theta) \pi(\theta) \lambda^n(dx) d\theta \quad (14)$$

and matrix ordering $A(n) \succeq B(n)$ as asymptotic positive semi-definite property

$$\inf_{|v|=1} \sum_{i=1}^m \sum_{j=1}^m (A(n) - B(n))_{ij} v_i v_j \geq -\delta_n, \lim_{n \rightarrow \infty} \delta_n \|A(n) - B(n)\|^{-1} = 0, \delta_n > 0. \quad (15)$$

Let also

$$\mathbf{E} \left(g(X^{(n)}, \theta) \right) = \int_{\mathfrak{X}^{(n)}} \int_{\Theta} g(x^{(n)}, \theta) p(x^{(n)}; \theta) \pi(\theta) \lambda^n(dx) d\theta$$

denote the expectation over $(\mathfrak{X}^{(n)}, \Theta)$. The following results formulate conditions under which the lower bounds for risk (14) in the sense of (15) are obtained in terms of Hellinger information and the index of regularity.

3.3. Main Results

Theorem 1. Let $\alpha = (2, \dots, 2, \alpha_m)', \alpha_m \in [1, 2)$,

$$J_{ij}(\theta) = I_{ij}(\theta), i, j = 1, \dots, m - 1; J_{im} = 0, J_{mm} \neq 0.$$

$$\left(\begin{array}{ccc|c} & & & 0 \\ & I(\theta) & & \dots \\ & & & 0 \\ \hline 0 & \dots & 0 & J_{mm}(\theta) \end{array} \right)$$

Then, if $\mathbf{E}I^{-1}(\theta) < \infty, \mathbf{E}J_{mm}^{-1}(\theta) < \infty$, it is true that

1. $J(\theta) > 0$
2. $\inf_{\theta^*} \mathbf{R}(\theta^*) \succeq C_1(\alpha_m) \text{Diag}(n^{-2/\alpha_m}) \mathbf{E}J^{-1}(\theta)$

Theorem 2. Let $\alpha_1 = \dots = \alpha_m = 1$,

$$J_{ij}(\theta) = 0, i \neq j; J_{ii} \neq 0.$$

$$\begin{pmatrix} J_{11}(\theta) & 0 & \dots & 0 \\ 0 & J_{22}(\theta) & 0 & \dots \\ \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & J_{mm}(\theta) \end{pmatrix}$$

Then, if $\mathbf{E}J_{ii}^{-1}(\theta) < \infty$, it is true that

1. $J(\theta) > 0$
2. $\inf_{\theta^*} \mathbf{R}(\theta^*) \succeq C_2 n^{-2} \mathbf{E}J^{-1}(\theta)$

Proofs of Theorems 1 and 2 are technically similar to the proof of the main results of [5], although the definition of Hellinger information was not explicitly provided in that paper.

3.4. Hellinger Priors

If $J > 0$, as in the conditions of Theorems 1 and 2, the vector Hellinger prior can be defined as

$$\pi_H(\theta) \propto \sqrt{\det J(\theta)}.$$

In the case of all components $\alpha_i \equiv 2$, Hellinger information reduces to the Fisher information matrix, and our approach leads to the Jeffreys prior [12].

Example 3. Truncated Weibull distribution (see Theorem 1):

$$p(x; \beta, \phi, \tau) = \beta \phi^\beta x^{\beta-1} \exp\{-\phi^\beta(x^\beta - \tau^\beta)\}, x \in [\tau, \infty)$$

with two parameters of interest: regular pseudo-scale $\phi > 0$ and nonregular threshold $\tau > 0$. Assume $\beta > 2$ fixed. See Figure 2.

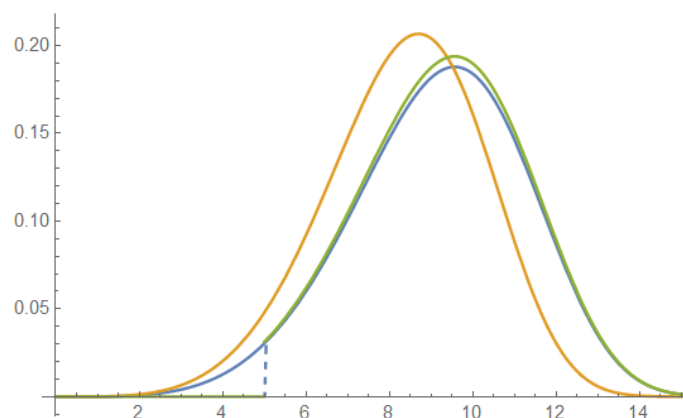


Figure 2. Truncated Weibull $p(x; 5, 0.1, 0); p(x; 5, 0.11, 0); p(x; 5, 0.1, 5)$.

Using notation $\theta = (\phi, \tau)$, obtain $\alpha_1 = 2, \alpha_2 = 1, \delta_1 = \epsilon, \delta_2 = \epsilon^2$ and

$$H(\theta, \Delta) = \begin{pmatrix} \delta_1 \phi^{-2} & o(\delta_1 \delta_2) \\ o(\delta_1 \delta_2) & \delta_2 \phi^{2\beta} \tau^{2\beta-2} \end{pmatrix}.$$

After limit transition in δ ,

$$J(\theta) = \begin{pmatrix} \phi^{-2} & 0 \\ 0 & \phi^{2\beta} \tau^{2\beta-2} \end{pmatrix}.$$

Therefore, $\pi_H(\theta) \propto \phi^{\beta-1} \tau^{\beta-1}$, which is also the reference prior for the vector parameter [21].

Example 4. Circular beta distribution on a disc (see Theorem 1):

$$p(r, w; \beta_1, \beta_2, \rho) = \frac{r}{\pi \rho^2} \left(\frac{w}{2\pi}\right)^{\beta_1-1} \left(1 - \frac{w}{2\pi}\right)^{\beta_1-1}$$

with three parameters of interest: regular $\beta_1, \beta_2 > 1$ and nonregular radius $\rho > 0$. See Figure 3.

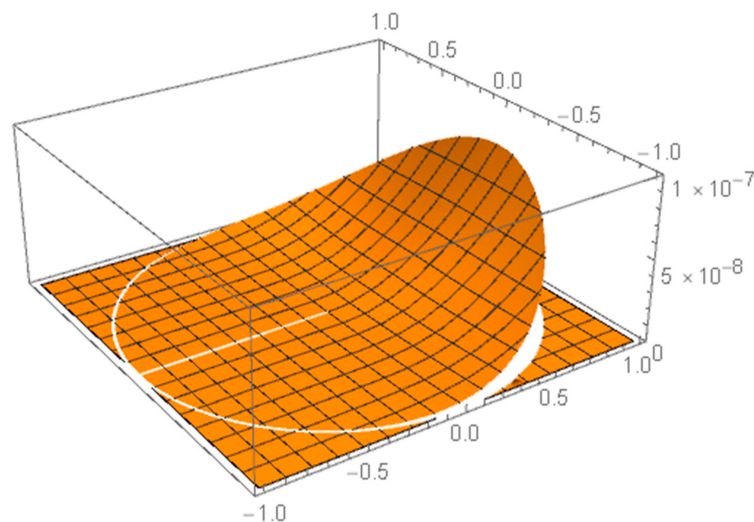


Figure 3. Truncated beta on a disc.

$$J(\theta) = \begin{pmatrix} \psi(\beta_1) - \psi(\beta_1 + \beta_2) & -\psi(\beta_1 + \beta_2) & 0 \\ -\psi(\beta_1 + \beta_2) & \psi(\beta_1) - \psi(\beta_1 + \beta_2) & 0 \\ 0 & 0 & 4\rho^{-2} \end{pmatrix},$$

where $\psi(z)$ is the polygamma function of order 1.

$$\text{Therefore, } \pi_H(\theta) \propto \rho^{-1} \sqrt{\psi(\beta_1)\psi(\beta_2) - \psi(\beta_1 + \beta_2)[\psi(\beta_1) + \psi(\beta_2)]}$$

Example 5. Uniform on a rectangle (see Theorem 2): $X \sim \text{Unif}(\theta, \theta^2), \theta \in (1, \infty)$ with two parameters of interest: regular pseudo-scale $\phi > 0$ and nonregular threshold $\tau > 0$.

Using notation $\theta = (\theta_1, \theta_2)$, obtain $\alpha_1 = \alpha_2 = 1, \delta = \delta_1 = \delta_2 = \epsilon^2$ and

$$H(\theta, \Delta) = \begin{pmatrix} \delta \theta_1^{-2} & \delta^2 (\theta_1 \theta_2)^{-1} \\ \delta^2 (\theta_1 \theta_2)^{-1} & \delta \theta_2^{-2} \end{pmatrix}.$$

After the limit transition,

$$J(\theta) = \begin{pmatrix} \theta_1^{-2} & 0 \\ 0 & \theta_2^{-2} \end{pmatrix}.$$

Therefore, $\pi_H(\theta) \propto \theta_1^{-1} \theta_2^{-1}$, also the reference prior.

Example 6. Uniform with two moving boundaries $X \sim \text{Unif}(\theta_1, \theta_1 + \theta_2), \theta_2 \in (0, \infty)$ (neither Theorem 1 nor Theorem 2 applies): using notation $\theta = (\theta_1, \theta_2)$, obtain $\alpha_1 = \alpha_2 = 1, \delta = \delta_1 = \delta_2 = \epsilon^2$ and

$$H(\theta, \Delta) = \begin{pmatrix} 2\delta\theta_2^{-2} & \delta\theta_2^{-2} \\ \delta\theta_2^{-2} & \delta\theta_2^{-2} \end{pmatrix}.$$

After limit transition,

$$J(\theta) = \begin{pmatrix} 2\theta_2^{-2} & \theta_2^{-2} \\ \theta_2^{-2} & \theta_2^{-2} \end{pmatrix}.$$

Therefore, $\pi_H(\theta) \propto \theta_2^{-2}$, also the reference prior.

4. Discussion

A Hellinger information matrix can be defined in a reasonable way to serve as a substitute for the Fisher information matrix in multivariate nonregular cases. It can be used as a technically simple tool for the elicitation of non-informative priors.

Properties of the Hellinger distance (symmetry, etc.) grant certain advantages vs analogous constructions based on Kullback–Leibler divergence.

Some interesting nonregularities are not covered by the conditions of Theorems 1 and 2 (see Example 6). More general results related to a positive definite property of $J(\theta)$ would be interesting.

It is tempting to obtain Hellinger priors as the solution of a particular optimization problem (similar to the reference priors).

Funding: This research was supported by the Center for Applied Mathematics at the University of St. Thomas.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The author is grateful for helpful discussions at the online Borovkov Readings and O-Bayes-22 in Santa Cruz, CA in August 2022. The author is also thankful for many helpful comments by the reviewers.

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

References

- Li, W.; Zhao, J. Wasserstein information matrix. *arXiv* **2019**, arXiv:1910.11248.
- Li, W.; Rubio, F. J. On a prior based on the Wasserstein information matrix. *arXiv* **2022**, arXiv:2202.03217.
- Ibragimov, I.A.; Has'minskii, R.Z. *Statistical Estimation: Asymptotic Theory*; Springer: New York, NY, USA, 1981.
- Shemyakin, A. A New Approach to Construction of Objective Priors: Hellinger Information. *Appl. Econom.* **2012**, *28*, 124–137. (In Russian)
- Shemyakin, A. Rao-Cramer type multidimensional integral inequalities for parametric families with singularities. *Sib. Math. J.* **1991**, *32*, 706–715. (In Russian) [[CrossRef](#)]
- Shemyakin, A. On Information Inequalities in the Parametric Estimation. *Theory Probab. Appl.* **1992**, *37*, 89–91. [[CrossRef](#)]
- Lin, Y.; Martin, R.; Yang, M. On Optimal Designs for Nonregular Models. *Ann. Statist.* **2019**, *47*, 3335–3359. [[CrossRef](#)]
- Shemyakin, A. Hellinger Distance and Non-informative Priors. *Bayesian Anal.* **2014**, *9*, 923–938. [[CrossRef](#)]
- Bhattacharyya, A. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.* **1943**, *35*, 99–109.
- Bobrovsky, B.Z.; Mayer-Wof, E.; Zakai, M. Some classes of global Cramer-Rao bounds. *Ann. Statist.* **1987**, *15*, 1421–1438. [[CrossRef](#)]
- Brown, L.D.; Gajek, L. Information inequalities for the Bayes risk. *Ann. Statist.* **1990**, *18*, 1578–1594. [[CrossRef](#)]
- Jeffreys, H. An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.* **1946**, *186*, 453–461.
- Ghosal, S. Probability matching priors for non-regular cases. *Biometrika* **1999**, *86*, 956–964. [[CrossRef](#)]
- Berger, J.O.; Bernardo, J.M.T. On the development of reference priors (with discussion). *Bayesian Anal.* **1992**, *4*, 35–60.
- Ghosal, S.; Samanta, T. Expansion of Bayes risk for entropy loss and reference prior in nonregular cases. *Statist. Decis.* **1997**, *15*, 129–140. [[CrossRef](#)]
- Berger, J.O.; Bernardo, J.M.; Sun, D. The formal definition of reference priors. *Ann. Statist.* **2009**, *37*, 905–938. [[CrossRef](#)]
- Sun, D.; Berger, J.O. Reference priors with partial information. *Biometrika* **1998**, *85*, 55–71. [[CrossRef](#)]

18. Le, T.M. The Formal Definition of Reference Priors under a General Class of Divergence. Electronic Dissertations, University of Missouri-Columbia, Columbia, MO, USA, 2014. [[CrossRef](#)]
19. Smith, R.L. Nonregular regression. *Biometrika* **1994**, *81*, 173–183. [[CrossRef](#)]
20. Ghosal, S.; Ghosh, J.; Samanta, T. On convergence of posterior distributions. *Ann. Statist.* **1995**, *23*, 2145–2152. [[CrossRef](#)]
21. Sun, D. A note on non-informative priors for Weibull distribution. *J. Stat. Plan. Inference* **1997**, *61*, 319–338. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.