Information Geometry
https://doi.org/10.1007/s41884-024-00133-4

RESEARCH PAPER

A harmonic property of right invariant priors

Tomonari Sei[1] · Fumiyasu Komaki[1,2,3]

Received: 16 July 2023 / Revised: 5 April 2024 / Accepted: 8 April 2024

Abstract
It is shown that, on any Lie group, the density ratio of the right invariant measure to the left invariant measure is harmonic with respect to the left invariant Riemannian metric. This result is applied to the Bayesian prediction theory on group invariant statistical models. A method of constructing Bayesian prior distributions that asymptotically dominate the right invariant priors is provided.

Keywords Bayesian prediction · Fisher metric · Group invariant model · Laplacian · Superharmonic prior

1 Introduction

In Bayesian statistics, if the model has a group structure, inference based on the right invariant prior is known to have desirable properties; see [1] and references therein. The same holds true in Bayesian prediction [5, 10].

On the other hand, in the theory of Bayesian prediction, prior distributions that are superharmonic with respect to the Fisher metric have better performance than the Jeffreys prior [6].

These two facts raise the problem of the relation between right invariant priors and superharmonic priors. The Jeffreys prior corresponds to the left invariant prior in the models with group structures. In some examples such as location-scale models,

Communicated by Hiroshi Matsuzoe.

✉ Fumiyasu Komaki
komaki@mist.i.u-tokyo.ac.jp

1 Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

2 Mathematical Informatics Collaboration Unit, RIKEN Center for Brain Science, 2-1 Hirosawa, Wako City, Saitama 351-0198, Japan

3 International Research Center for Neurointelligence, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

the ratio of the right invariant prior to the left invariant prior is known to be harmonic with respect to the Fisher metric [7]. However, it was not known whether the harmonic property of the ratio of the right invariant prior to the left invariant prior holds in general. This paper proves the claim. The result is helpful for understanding the dominance property of the right invariant prior to the Jeffreys prior as shown in Lemma 2. We also provide a method of constructing prior distributions that asymptotically dominate the right invariant prior in Lemma 3, as we will demonstrate through examples.

In Sect. 2, we prove that the ratio of the right invariant measure to the left invariant measure is harmonic with respect to any left invariant metric. In Sect. 3, we apply the theorem to the Bayesian prediction problem.

2 Main result

Let G be a Lie group and e be its identity element. Choose a left invariant Riemannian metric h on G. We use the symbol h for Riemannian metrics to distinguish with elements of G usually denoted as g. In application to statistics, h is the Fisher metric of group-invariant models; see Sect. 3.

Let ν_{L} be the left invariant measure (left Haar measure) on G. Up to multiplicative constants, ν_{L} is written in terms of h as

$$\nu_{\mathrm{L}}(\mathrm{d}x) = \sqrt{|h|}\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n,$$

where x^i is a local coordinate of $x \in G$ and $|h|$ is the determinant of the metric with respect to the local coordinate system. Denote the reciprocal of the modulus of G by $\pi_{\mathrm{R/L}}$, that is,

$$\pi_{\mathrm{R/L}}(g) \int_G f(x)\nu_{\mathrm{L}}(\mathrm{d}x) = \int_G f(xg)\nu_{\mathrm{L}}(\mathrm{d}x) \tag{1}$$

(Eq. (1.2) of [1]) for any $g \in G$ and $f \in C_0(G)$, where $C_0(G)$ denotes the set of continuous functions with compact supports. The map $\pi_{\mathrm{R/L}} : G \to \mathbb{R}_{>0}$ is a group homomorphism. Define the right invariant measure ν_{R} by

$$\nu_{\mathrm{R}}(\mathrm{d}x) = \pi_{\mathrm{R/L}}(x)\nu_{\mathrm{L}}(\mathrm{d}x)$$

(see Eq. (1.4) of [1]). It is said that G is unimodular if $\pi_{\mathrm{R/L}}(x) = 1$ for all $x \in G$. We are interested in groups that are not unimodular.

Define the Laplace–Beltrami operator Δ associated with the metric h by

$$\Delta f = \frac{1}{\sqrt{|h|}}\partial_i(\sqrt{|h|}h^{ij}\partial_j f), \quad f \in C^2(G),$$

where ∂_i denotes the partial derivative with respect to the local coordinate, h^{ij} is the inverse matrix of $h_{ij} = h(\partial_i, \partial_j)$, and Einstein's summation convention is used. We call Δ the Laplacian for simplicity. The Laplacian does not depend on the choice of the local coordinate system. A function f is said to be harmonic if $\Delta f = 0$ everywhere and superharmonic if $\Delta f \leq 0$ everywhere.

Our main theorem is stated as follows.

Theorem 1 *The function $\pi_{\mathrm{R/L}}$ is harmonic.*

Proof Take a function $f \in C_0^\infty(G)$ such that $\int f(x)\nu_{\mathrm{L}}(\mathrm{d}x) \neq 0$. Equation (1) is written as

$$\pi_{\mathrm{R/L}}(g) \int f(x)\nu_{\mathrm{L}}(\mathrm{d}x) = \int (f \circ L_x)(g)\nu_{\mathrm{L}}(\mathrm{d}x),$$

where L_x is the left translation by x. Applying the Laplacian to both sides with respect to g yields

$$\begin{aligned}(\Delta\pi_{\mathrm{R/L}})(g) \int f(x)\nu_{\mathrm{L}}(\mathrm{d}x) &= \int \Delta(f \circ L_x)(g)\nu_{\mathrm{L}}(\mathrm{d}x)\\ &= \int ((\Delta f) \circ L_x)(g)\nu_{\mathrm{L}}(\mathrm{d}x)\\ &= \pi_{\mathrm{R/L}}(g) \int (\Delta f)(x)\nu_{\mathrm{L}}(\mathrm{d}x)\\ &= 0,\end{aligned}$$

where the first equality uses Lebesgue's convergence theorem, the second equality uses the isometric property of L_x and the invariance of Δ with respect to isometries (see p.246, Proposition 2.4 of [4]), the third equality uses Eq. (1) again, and the fourth equality uses an integral formula on the Laplacian (see p.245 Proposition 2.3 of [4]). This proves $\Delta\pi_{\mathrm{R/L}}(g) = 0$. □

Example 1 *(Affine transformations)* Consider the group of affine transformations

$$G = \left\{ g = \begin{pmatrix} 1 & 0 \\ \mu & \sigma \end{pmatrix} \mid \mu \in \mathbb{R},\ \sigma > 0 \right\},$$

which is used to analyze the location-scale family in statistics. Let us directly show that $\pi_{\mathrm{R/L}}$ is harmonic, as pointed out by [7]. It is widely known that the left and right invariant measures are

$$\nu_{\mathrm{L}}(\mathrm{d}g) = \frac{\mathrm{d}\mu \wedge \mathrm{d}\sigma}{\sigma^2}$$

and

$$\nu_{\mathrm{R}}(\mathrm{d}g) = \frac{\mathrm{d}\mu \wedge \mathrm{d}\sigma}{\sigma},$$

respectively (p. 63 of [3]). The density function of ν_{R} with respect to ν_{L} is

$$\pi_{\mathrm{R/L}}(g) = \sigma.$$

To derive the Laplacian, we determine a left invariant Riemannian metric. The metric tensor h_e at the identity element e is arbitrarily chosen. From the G-invariance, the metric tensor at $g \in G$ is

$$h_g = (L_{g^{-1}})^* h_e = B^\top h_e B = \frac{1}{\sigma^2} h_e,$$

where $(L_{g^{-1}})^*$ denotes the pull-back operator associated with the left translation $L_{g^{-1}}$ and B is the Jacobian matrix of $L_{g^{-1}}$. Indeed, the left translation

$$\begin{pmatrix} 1 & 0 \\ \mu_y & \sigma_y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \mu & \sigma \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ \mu_x & \sigma_x \end{pmatrix}$$

has the Jacobian matrix

$$B = \frac{\partial(\mu_y, \sigma_y)}{\partial(\mu_x, \sigma_x)} = \frac{1}{\sigma} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The Laplacian is

$$\begin{aligned} \Delta &= \frac{1}{\sqrt{|h_g|}} \partial_i (\sqrt{|h_g|} (h_g)^{ij} \partial_j) \\ &= \frac{\sigma^2}{\sqrt{|h_e|}} (\partial_\mu, \partial_\sigma) \left\{ \sigma^{-2} \sqrt{|h_e|} \sigma^2 h_e^{-1} \begin{pmatrix} \partial_\mu \\ \partial_\sigma \end{pmatrix} \right\} \\ &= \sigma^2 (\partial_\mu, \partial_\sigma) \left\{ h_e^{-1} \begin{pmatrix} \partial_\mu \\ \partial_\sigma \end{pmatrix} \right\}. \end{aligned}$$

It is immediate to see that $\pi_{\mathrm{R/L}}(g) = \sigma$ is harmonic for any choice of h_e.

3 Application to Bayesian prediction

3.1 Bayesian prediction problem

We briefly recall the Bayesian prediction problem and its relation with geometric quantities such as the Fisher metric and Laplacian.

A statistical model, or simply called a model, is a set of probability measures on a given measurable space $(\mathcal{X}, \mathcal{F})$ indexed by a parameter θ as

$$\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}.$$

We assume that the model is identifiable, that is, $\theta_1 \neq \theta_2$ implies $P_{\theta_1} \neq P_{\theta_2}$. The parameter space Θ is assumed to be an orientable d-dimensional C^∞-manifold. Let P_θ be absolutely continuous with respect to a base measure $v(\mathrm{d}x)$ and its density function

$p(x|\theta)$ be positive everywhere and differentiable with respect to θ. The Fisher metric on Θ is defined by

$$h_{ij}(\theta) = \int \{\partial_i \log p(x|\theta)\}\{\partial_j \log p(x|\theta)\} P_\theta(\mathrm{d}x),$$

where ∂_i is the partial derivative with respect to local coordinates of θ. We assume that the Fisher metric is of C^∞ class and positive definite everywhere. The Fisher metric does not depend on the choice of the base measure $v(\mathrm{d}x)$.

A Borel measure on Θ is called a Bayesian prior distribution or just a prior. The volume element

$$J(\mathrm{d}\theta) = \sqrt{|h|}\mathrm{d}\theta^1 \wedge \cdots \wedge \mathrm{d}\theta^d$$

induced from the Fisher metric is called the Jeffreys prior. The Jeffreys prior does not depend on the choice of the local coordinate system. We focus on priors $\pi(\theta)J(\mathrm{d}\theta)$ that are absolutely continuous with respect to the Jeffreys prior. We call $\pi(\theta)$ the prior density. Since $J(\mathrm{d}\theta)$ does not depend on the local coordinate system, $\pi(\theta)$ is a scalar function. The functions π are assumed to be positive-valued and of C^2 class. We consider not only proper priors but also improper priors.

A statistical prediction problem is to estimate the distribution of future observation $y \in \mathcal{X}$ based on an independent sample $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$ from P_θ. The Bayesian predictive density

$$p_\pi(y|x^n) = \int p(y|\theta)\pi(\theta|x^n)J(\mathrm{d}\theta), \tag{2}$$

based on the posterior density

$$\pi(\theta|x^n) = \frac{\prod_{i=1}^n p(x_i|\theta)\pi(\theta)}{\int \prod_{i=1}^n p(x_i|\theta)\pi(\theta)J(\mathrm{d}\theta)}$$

is of interest.

The Bayesian prediction problem is to find a prior density function that has smaller prediction risk. We adopt the following risk function.

Definition 1 *(Asymptotic risk; Eq. (13) of* [7]*)* The asymptotic risk function of the prior density $\pi \in C^2(\Theta)$ is defined by

$$r(\pi) = r(\pi, \theta) = \frac{1}{\sqrt{\pi(\theta)}}\Delta\sqrt{\pi(\theta)},$$

where Δ denotes the Laplacian on Θ with respect to the Fisher metric. A prior density π_1 is said to dominate π_2 asymptotically if $r(\pi_1, \theta) \leq r(\pi_2, \theta)$ for all θ and $r(\pi_1, \theta) < r(\pi_2, \theta)$ for some θ.

The asymptotic risk is the leading term of the asymptotic expansion of the Kullback–Leibler risk of the Bayesian predictive density (2) as $n \to \infty$. See Eq. (4) of [6] and Eq. (13) of [7] for details. It is straightforward to see

$$r(\pi) = \frac{1}{2\pi}\Delta\pi - \frac{1}{4}h^{ij}(\partial_i \log \pi)(\partial_j \log \pi). \tag{3}$$

This is proved as

$$\begin{aligned}
\frac{1}{\sqrt{\pi}}\Delta\sqrt{\pi} &= \frac{1}{\sqrt{\pi}\sqrt{|h|}}\partial_i(h^{ij}\sqrt{|h|}\partial_j\sqrt{\pi}) \\
&= \frac{1}{\sqrt{\pi}\sqrt{|h|}}\partial_i\left(h^{ij}\sqrt{|h|}\frac{\partial_j\pi}{2\sqrt{\pi}}\right) \\
&= \frac{1}{\sqrt{\pi}\sqrt{|h|}}\left(\frac{\partial_i(h^{ij}\sqrt{|h|}\partial_j\pi)}{2\sqrt{\pi}} - (h^{ij}\sqrt{|h|}\partial_j\pi)\frac{\partial_i\pi}{4\pi^{3/2}}\right) \\
&= \frac{1}{2\pi}\Delta\pi - \frac{1}{4}h^{ij}(\partial_i \log \pi)(\partial_j \log \pi).
\end{aligned}$$

Our problem is to find a prior density π that has smaller asymptotic risk. The asymptotic risk of the Jeffreys prior density is 0 from the definition. Non-constant superharmonic prior densities asymptotically dominate the Jeffreys prior density since (3) holds.

3.2 Group invariant models

We consider the Bayesian prediction problem over group invariant models. Refer to [1, 2, 13] for comprehensive textbooks on the invariant models.

For simplicity, we suppose that the sample space $\mathcal{X}$ is also a C^∞ manifold. Let a Lie group G act on $\mathcal{X}$ smoothly from the left. For a probability measure P on $\mathcal{X}$ and $g \in G$, the push-forward measure $g_* P$ is defined by $g_* P(B) = P(g^{-1}B)$ for Borel sets B. The group G acts on the set of all probability measures by the push-forward operation.

Definition 2 *(Group invariant model; Definition 3.1 of* [1]*)* A statistical model $\mathcal{P}$ is said to be G-invariant if for each $P \in \mathcal{P}$, $g_* P \in \mathcal{P}$ for all $g \in G$.

If a G-invariant statistical model is parameterized as $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$, the left action of G on Θ is well defined by $P_{g\theta} = g_* P_\theta$ under identifiability. We assume that G transitively acts on Θ.

Let $v(\mathrm{d}x)$ be the base measure of $\mathcal{P}$ as in the preceding subsection. We say that tensors on Θ are G-invariant if they are preserved under the group action.

Lemma 1 *Let $\mathcal{P}$ be a G-invariant model. Then, the Fisher metric h is G-invariant. In particular, the Jeffreys prior is a left G-invariant measure on Θ.*

See "Appendix" for the proof. Lemma 1 is used to prove Lemma 2.

We say that G acts freely on Θ if $g\theta = \theta$ for some $\theta \in \Theta$ implies $g = e$. If the action is free, the parameter space $\Theta = \{g\theta_0 \mid g \in G\}$ is identified with G, where $\theta_0 \in \Theta$ is a fixed element. Under the identification, the left invariant measure ν_{L} on G is equal to the Jeffreys prior and the right invariant measure ν_{R} is a prior on Θ, which we call the right invariant prior. It is known that the right invariant prior provides the best invariant predictive distribution [5, 10], which means that the right invariant prior attains the minimum of the Kullback–Leibler risk in the class of invariant predictive distributions. In particular, the right invariant prior dominates the Jeffreys prior if G is not unimodular. This fact is reflected in the following lemma. We prove the lemma in "Appendix" without using the fact.

Lemma 2 *Suppose that G is not unimodular and acts freely on Θ. Then, the asymptotic risk of the right invariant prior density $\pi_{\mathrm{R/L}}(\theta)$ is a negative constant.*

Even if the action of G is not free, the theorem holds for any Lie subgroup G_1 of G that acts freely and transitively on Θ. In that case, we can identify Θ with G_1 and construct harmonic prior densities from the right invariant measures on G_1. Furthermore, since all the conjugate subgroups gG_1g^{-1} ($g \in G$) act freely as well, various harmonic prior densities are obtained. The prior densities have the same asymptotic risk because G_1 and gG_1g^{-1} are isomorphic. We can reduce the asymptotic risk by aggregating the prior densities as follows.

Lemma 3 *Let π_1 and π_2 be smooth positive functions on Θ. Define the generalized mean $\bar{\pi}_\beta$ by*

$$\bar{\pi}_\beta = \left(\frac{\pi_1^\beta + \pi_2^\beta}{2}\right)^{1/\beta}$$

for $\beta \neq 0$ and $\bar{\pi}_0 = (\pi_1\pi_2)^{1/2}$ for $\beta = 0$. If $\beta < 1/2$, then

$$r(\bar{\pi}_\beta) \leq \frac{\pi_1^\beta r(\pi_1) + \pi_2^\beta r(\pi_2)}{\pi_1^\beta + \pi_2^\beta}.$$

The equality holds for all $\theta \in \Theta$ if and only if π_1/π_2 is constant.

See "Appendix" for the proof. The case $\beta = 0$ is proved in [12].

We provide two applications of the lemma. In the applications, we first find a closed subgroup G_1 that freely and transitively acts on Θ. Then, take $g \in G$ and put $G_2 = gG_1g^{-1}$. Under the identifications $G_1 \simeq \Theta$, $g_1 \mapsto g_1\theta_0$, and $G_2 \simeq \Theta$, $g_2 \mapsto g_2g\theta_0$ as G-spaces, the following equality holds for any $\theta \in \Theta$:

$$\pi_2(\theta) = \pi_1(g^{-1}\theta),$$

where π_1 and π_2 are the densities of the right invariant priors of G_1 and G_2, respectively. Indeed, the left and right invariant measures on G_2 are the push-forward of those on G_1 by $g_1 \mapsto gg_1g^{-1}$. Then the density $\pi_1(g_1\theta_0)$ is equal to $\pi_2((gg_1g^{-1})g\theta_0) = \pi_2(gg_1\theta_0)$, which proves $\pi_2(\theta) = \pi_1(g^{-1}\theta)$ for $\theta = gg_1\theta_0$.

Example 2 *(Cauchy location-scale family* [11]*)* Consider the Cauchy density function

$$p(x|\mu,\sigma) = \frac{1}{\pi\sigma(1+(x-\mu)^2/\sigma^2)}, \quad x \in \mathbb{R},$$

with respect to the Lebesgue measure, where μ and σ are called the location and scale parameters, respectively. The parameter space is $\Theta = \{(\mu,\sigma) \mid \mu \in \mathbb{R}, \sigma > 0\}$. The density function is written in terms of complex numbers as

$$p(x|\mu,\sigma) = \frac{\sigma}{\pi|x-(\mu+i\sigma)|^2}, \quad i = \sqrt{-1}.$$

The general linear group $G = \mathrm{GL}^+(2,\mathbb{R})$ with positive determinant acts on this model through the linear fractional transformation

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot x = \frac{ax+b}{cx+d}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G, \quad x \in \mathcal{X} = \mathbb{R}.$$

The action of G on the parameter space is

$$\begin{aligned} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot (\mu+i\sigma) &= \frac{a(\mu+i\sigma)+b}{c(\mu+i\sigma)+d} \\ &= \frac{ac\sigma^2+(a\mu+b)(c\mu+d)}{(c\sigma)^2+(c\mu+d)^2} + i\frac{(ad-bc)\sigma}{(c\sigma)^2+(c\mu+d)^2} \end{aligned}$$

for $(\mu,\sigma) \in \Theta$. See [11] for details. Although the action of G on Θ is not free, a subgroup

$$G_1 = \left\{ \begin{pmatrix} \sigma & \mu \\ 0 & 1 \end{pmatrix} \mid \mu \in \mathbb{R}, \sigma > 0 \right\}$$

acts freely. We can identify G_1 with Θ. As in Example 1, the left and right invariant measures of G_1 are $\sigma^{-2}\mathrm{d}\mu \wedge \mathrm{d}\sigma$ and $\sigma^{-1}\mathrm{d}\mu \wedge \mathrm{d}\sigma$, respectively. The density of the right invariant prior on G_1 is

$$\pi_1(\mu,\sigma) = \sigma.$$

From Theorem 1 and Lemma 2, the asymptotic risk of π_1 is negative constant. Now consider a conjugate group

$$G_2 = gG_1g^{-1}, \quad g = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in G,$$

which also acts freely on Θ. The density of the right invariant prior on G_2 is

$$\pi_2(\mu,\sigma) = \pi_1(g^{-1}(\mu,\sigma)) = \frac{\sigma}{\sigma^2+\mu^2}.$$

This prior density is discussed in [7].

Finally, by taking the geometric mean of π_1 and π_2, we obtain a prior density

$$\sqrt{\pi_1\pi_2} = \frac{\sigma}{\sqrt{\sigma^2+\mu^2}} = \frac{1}{\sqrt{1+(\mu/\sigma)^2}},$$

which shrinks the signal-noise ratio μ/σ to the origin. Lemma 3 implies that the asymptotic risk of $(\pi_1\pi_2)^{1/2}$ is smaller than those of the right invariant priors π_1 and π_2.

For location-scale families other than the Cauchy family, the general linear group does not act because the family is not closed under the reciprocal $1/X$ of the random variable X. However, the dominance relationship on the asymptotic risk remains true because the asymptotic risk depends only on the Riemannian structure. See also [7] for this point.

Example 3 *(two-dimensional Wishart model* [8, 12]*)* Suppose that a random variable X has the two-dimensional Wishart distribution $W_2(n, \Sigma)$ with the degree of freedom n and the covariance parameter

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

The model is G-invariant with respect to the general linear group $G = \mathrm{GL}(2, \mathbb{R})$, where the group action is defined by $(g, X) \mapsto gXg^\top$ and $(g, \Sigma) \mapsto g\Sigma g^\top$. The sample space $\mathcal{X}$ and the parameter space Θ are the set of positive definite symmetric matrices. The subgroup

$$G_1 = \left\{ \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \mid a, c > 0, b \in \mathbb{R} \right\}$$

of G has a one-to-one correspondence with Θ through the Cholesky decomposition $\Sigma = gg^\top$ with $g \in G_1$. The left and right invariant measures of G_1 are

$$\nu_{\mathrm{L}} = \frac{1}{ac^2}\mathrm{d}a \wedge \mathrm{d}b \wedge \mathrm{d}c = \frac{1}{4|\Sigma|^{3/2}}\mathrm{d}\sigma_{11} \wedge \mathrm{d}\sigma_{12} \wedge \mathrm{d}\sigma_{22}$$

and

$$\nu_{\mathrm{R}} = \frac{1}{a^2c}\mathrm{d}a \wedge \mathrm{d}b \wedge \mathrm{d}c = \frac{1}{4\sigma_{11}|\Sigma|}\mathrm{d}\sigma_{11} \wedge \mathrm{d}\sigma_{12} \wedge \mathrm{d}\sigma_{22},$$

respectively. The density of the right invariant prior on G_1 is

$$\pi_1(\Sigma) = \frac{|\Sigma|^{1/2}}{\sigma_{11}}$$

in the Σ-coordinate. A conjugate group

$$G_2 = gG_1g^\top, \quad g = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

also acts freely on Θ. The density of the right invariant prior on G_2 is

$$\pi_2(\Sigma) = \pi_1(g\Sigma g^\top) = \frac{|\Sigma|^{1/2}}{\sigma_{22}}.$$

The harmonic mean of π_1 and π_2 is

$$\left(\frac{\pi_1^{-1} + \pi_2^{-1}}{2}\right)^{-1} = \frac{2|\Sigma|^{1/2}}{\mathrm{tr}(\Sigma)},$$

which is orthogonally invariant and shrinks the ratio of the two eigenvalues towards one. Lemma 3 implies that the prior asymptotically dominates the right invariant priors π_1 and π_2. The dominance relationship holds even in finite-sample cases as shown by [8].

Similarly, the geometric mean of π_1 and π_2 is

$$\sqrt{\pi_1\pi_2} = \frac{|\Sigma|^{1/2}}{(\sigma_{11}\sigma_{22})^{1/2}},$$

which is scale invariant and shrinks the correlation coefficient towards the origin. Again, Lemma 3 tells us that the prior asymptotically dominates the right invariant priors π_1 and π_2. This relation holds even in finite-sample cases as shown by [12].

The two examples show how Theorem 1 is useful in Bayesian inference.

We finally mention the predictive metric defined by [9], which appears in the asymptotic risk when the observed and predicted variables have different statistical models. The predictive metric is G-invariant whenever the statistical models for observed and predicted variables are G-invariant. The method of obtaining harmonic prior distributions is applicable to this case.

A Proof of lemmas

A.1 Proof of Lemma 1

Recall that the Fisher metric is a (0, 2) symmetric tensor

$$h(\theta) = \int \{\mathrm{d}\log p(x|\theta)\}^2 p(x|\theta) v(\mathrm{d}x),$$

where $\mathrm{d}\log p(x|\theta) = \partial_i \log p(x|\theta)\mathrm{d}\theta^i$ is the exterior derivative with respect to $\theta \in \Theta$. We prove $g^*h = h$. The pull-back g^*h is

$$\begin{aligned} g^*h &= g^* \int \{\mathrm{d}\log p(x|\theta)\}^2 p(x|\theta)v(\mathrm{d}x) \\ &= \int \{g^*\mathrm{d}\log p(x|\theta)\}^2 p(x|g\theta)v(\mathrm{d}x). \\ &= \int \{\mathrm{d}\log p(x|g\theta)\}^2 p(x|g\theta)v(\mathrm{d}x). \end{aligned}$$

By assumption, the statistical model satisfies $g_* P_\theta = P_{g\theta}$, which is equivalent to $g_*(p(x|\theta)v(\mathrm{d}x)) = p(x|g\theta)v(\mathrm{d}x)$ and therefore

$$p(g^{-1}x|\theta)(g_*v)(\mathrm{d}x) = p(x|g\theta)v(\mathrm{d}x).$$

In particular, g_*v and v are absolutely continuous with respect to each other because $p(x|\theta)$ is assumed to be positive everywhere. We have

$$p(x|g\theta) = p(g^{-1}x|\theta)\frac{\mathrm{d}(g_*v)}{\mathrm{d}v}(x).$$

Since $\mathrm{d}(g_*v)/\mathrm{d}v$ does not depend on θ, we obtain

$$\mathrm{d}\log p(x|g\theta) = \mathrm{d}\log p(g^{-1}x|\theta).$$

Therefore

$$\begin{aligned} g^*h &= \int \{\mathrm{d}\log p(x|g\theta)\}^2 p(x|g\theta)v(\mathrm{d}x) \\ &= \int \{\mathrm{d}\log p(g^{-1}x|\theta)\}^2 p(g^{-1}x|\theta)(g_*v)(\mathrm{d}x) \\ &= \int \{\mathrm{d}\log p(x|\theta)\}^2 p(x|\theta)v(\mathrm{d}x) \\ &= h. \end{aligned}$$

This proves the G-invariance of h.

A.2 Proof of Lemma 2

From Eq. (3) and Theorem 1, the asymptotic risk of $\pi_{\mathrm{R/L}}$ is

$$r(\pi_{\mathrm{R/L}}) = -\frac{1}{4}h^{ij}(\partial_i \log \pi_{\mathrm{R/L}})(\partial_j \log \pi_{\mathrm{R/L}}).$$

The G-invariance of the asymptotic risk follows from the facts that h is G-invariant and $\pi_{\mathrm{R/L}}$ is group homomorphic. Therefore, $r(\pi_{\mathrm{R/L}})$ is a constant function on Θ since G acts transitively by assumption. If G is not unimodular, the asymptotic risk is negative because $\partial_i \log \pi_{\mathrm{R/L}}(\theta) \neq 0$ at some $\theta \in \Theta$.

A.3 Proof of Lemma 3

Consider K smooth positive functions $\pi_1, \dots, \pi_K$ for $K \geq 2$. The lemma is a special case $K = 2$. Denote the generalized mean by

$$\bar{\pi} = \left(K^{-1} \sum_{k=1}^{K} \pi_k^{\beta} \right)^{1/\beta}.$$

We prove $r(\bar{\pi}) \leq \sum_{k=1}^{K} \lambda_k r(\pi_k)$, where

$$\lambda_k = \frac{\pi_k^{\beta}}{\sum_{l=1}^{K} \pi_l^{\beta}}, \quad 1 \leq k \leq K.$$

Define

$$\mu_i = \sum_{k=1}^{K} \lambda_k \frac{\partial_i \sqrt{\pi_k}}{\sqrt{\pi_k}}, \quad 1 \leq i \leq d.$$

It is straightforward to see

$$\partial_i \sqrt{\bar{\pi}} = \sqrt{\bar{\pi}} \mu_i$$

and

$$\partial_i \left(\frac{\sqrt{\bar{\pi}} \lambda_k}{\sqrt{\pi_k}} \right) = \frac{\sqrt{\bar{\pi}} \lambda_k}{\sqrt{\pi_k}} (2\beta - 1) \left(\frac{\partial_i \sqrt{\pi_k}}{\sqrt{\pi_k}} - \mu_i \right).$$

By using the formulas, we obtain

$$\begin{aligned}
r(\bar{\pi}) &= \frac{1}{\sqrt{\bar{\pi} |h|}} \partial_i \left(\sqrt{|h|} h^{ij} \partial_j \sqrt{\bar{\pi}} \right) \\
&= \frac{1}{\sqrt{\bar{\pi} |h|}} \partial_i \left(\sqrt{|h|} h^{ij} \sqrt{\bar{\pi}} \sum_k \lambda_k \frac{\partial_j \sqrt{\pi_k}}{\sqrt{\pi_k}} \right) \\
&= \sum_k \frac{\lambda_k}{\sqrt{\pi_k}} \frac{1}{\sqrt{|h|}} \partial_i \left(\sqrt{|h|} h^{ij} \partial_j \sqrt{\pi_k} \right) + \sum_k h^{ij} \frac{1}{\sqrt{\bar{\pi}}} \partial_i \left(\frac{\sqrt{\bar{\pi}} \lambda_k}{\sqrt{\pi_k}} \right) \partial_j \sqrt{\pi_k} \\
&= \sum_k \lambda_k r(\pi_k) + \sum_k h^{ij} (2\beta - 1) \lambda_k \left(\frac{\partial_i \sqrt{\pi_k}}{\sqrt{\pi_k}} - \mu_i \right) \frac{\partial_j \sqrt{\pi_k}}{\sqrt{\pi_k}} \\
&= \sum_k \lambda_k r(\pi_k) + (2\beta - 1) \sum_k h^{ij} \lambda_k \left(\frac{\partial_i \sqrt{\pi_k}}{\sqrt{\pi_k}} - \mu_i \right) \left(\frac{\partial_j \sqrt{\pi_k}}{\sqrt{\pi_k}} - \mu_j \right).
\end{aligned}$$

The last term is non-positive since $\beta < 1/2$. This proves the desired inequality $r(\bar{\pi}) \leq \sum_k \lambda_k r(\pi_k)$. The equality holds if and only if $(\partial_i \sqrt{\pi_k})/\sqrt{\pi_k} = \mu_i$ for all i and k, or equivalently, π_k/π_l are constants for all k, l.

Acknowledgements The authors are grateful to two anonymous referees for their careful reading and insightful suggestions.

Funding Open Access funding provided by The University of Tokyo. This work was supported by JSPS KAKENHI Grant Number 21K11781, 22H00510, and AMED Grant Numbers JP23dm0207001 and JP23dm0307009.

Data availability The manuscript has no associated data.

Declarations

Conflict of interest T. Sei and F. Komaki are current members of the Editorial Board of Information Geometry. On behalf of all authors, the corresponding author states that there is no other Conflict of interest.

References

1. Eaton, M.L.: Group Invariance Applications in Statistics. IMS, Hayward, California (1989)
2. Eaton, M.L.: Multivariate Statistics: A Vector Space Approach. IMS, Beachwood, Ohio (2007)
3. Fraser, D.A.S.: The Structure of Inference. John Wiley & Sons, New York (1968)
4. Helgason, S.: Groups and Geometric Analysis. American Mathematical Society, Orlando, Florida (1984)
5. Komaki, F.: Bayesian predictive distribution with right invariant priors. Calc. Stat. Assoc. Bull. **52**, 171–179 (2002)
6. Komaki, F.: Shrinkage priors for Bayesian prediction. Ann. Stat. **34**(2), 808–819 (2006)
7. Komaki, F.: Bayesian prediction based on a class of shrinkage priors for location-scale models. Ann. Inst. Stat. Math. **59**, 135–146 (2007)
8. Komaki, F.: Bayesian predictive densities based on superharmonic priors for the 2-dimensional Wishart model. J. Multivar. Anal. **100**(10), 2137–2154 (2009)
9. Komaki, F.: Asymptotic properties of Bayesian predictive densities when the distributions of data and target variables are different. Bayesian Anal. **10**(1), 31–51 (2015)
10. Liang, F., Barron, A.: Exact minimax strategies for predictive density estimation, data compression, and model selection. IEEE Trans. Inf. Theory **50**, 2708–2726 (2004)
11. McCullagh, P.: Möbius transformation and Cauchy parameter estimation. Ann. Stat. **24**(2), 787–808 (1996)
12. Sei, T., Komaki, F.: A correlation-shrinkage prior for Bayesian prediction of the two-dimensional Wishart model. Biometrika **109**(4), 1173–1180 (2022)
13. Wijsman, R.A.: Invariant Measures on Groups and Their Use in Statistics. IMS, Hayward, California (1990)