Statistics and Probability Letters 138 (2018) 111–115

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Scoring rules for statistical models on spheres

Yuya Takasu [a], Keisuke Yano [a,*], Fumiyasu Komaki [a,b]

[a] Department of Mathematical Informatics Graduate School of Information Science and Technology The University of Tokyo, Tokyo, Japan
[b] RIKEN Brain Science Institute, Saitama, Japan

ARTICLE INFO

Article history:
Received 19 April 2017
Received in revised form 12 February 2018
Accepted 20 February 2018
Available online 7 March 2018

Keywords:
Directional statistics
Divergence
Normalizing constant

ABSTRACT

We introduce a novel class of strictly proper scoring rules for statistical models on spheres that does not require the calculation of normalizing constants. The class contains the Hyvärinen scoring rule investigated by Mardia et al.(2016).

1. Introduction

Consider parameter estimation in a parametric model $\mathcal{P}$ on a p-dimensional unit sphere $\mathbb{S}^p := \{z = (z^1, \ldots, z^{p+1})^\top \in \mathbb{R}^{p+1} : (z^1)^2 + \cdots + (z^{p+1})^2 = 1\}$. The maximum likelihood estimation in a parametric model on a sphere is often difficult. Suppose that $\mathcal{P}$ is given by $\{f(\cdot;\theta) = f_1(\cdot;\theta)/c(\theta) : \theta \in \Theta\}$ for $\Theta \subset \mathbb{R}^d$ and $d \in \mathbb{N}$, where $f_1(\cdot;\theta)$ is a non-negative function on $\mathbb{S}^p$ for each $\theta \in \Theta$. The normalizing constant $c(\theta)$ of $f_1(\cdot;\theta)$ often does not have an explicit form. One of the most important examples is the Fisher–Bingham model (Mardia and Jupp, 2000). To conduct maximum likelihood estimation for the Fisher–Bingham model, the saddle-point approximation (Kume and Wood, 2005) and the holonomic gradient method (Nakayama et al., 2011) have been proposed.

Instead of the maximum likelihood estimation, we consider parameter estimation based on proper scoring rules. A scoring rule S is a loss function $S(z, f) : \mathbb{S}^p \times \mathcal{P} \to \mathbb{R} \cup \{\infty\}$ that measures the quality of a density f as an estimate of the density of a random variable Z on $\mathbb{S}^p$ in the case that the realized value of Z is z. It is said to be proper if the expected score $\int S(z, g)f(z)\mathrm{d}\mu(z)$ is minimized at $g = f$ for any $f \in \mathcal{P}$. It is said to be strictly proper if the minimizer is unique. On the basis of samples $Z_1, Z_2, \ldots, Z_n$ and a proper scoring rule S, we estimate θ by the minimum score estimator $\widehat{\theta}(Z_1, \ldots, Z_n) \in \mathrm{argmin}_{\theta\in\Theta} \sum_{i=1}^n S(Z_i, f(\cdot;\theta))$. Note that the estimator based on a strictly proper scoring rule is typically consistent; see Dawid (2007).

Hyvärinen (2005) proposed the Hyvärinen scoring rule, a strictly proper scoring rule that does not require the normalizing constant. Ehm and Gneiting (2012) and Parry et al. (2012) characterized proper scoring rules that do not require the normalizing constant. Their contributions mainly go through the case in which samples take their values on Euclidean spaces. Recently, Mardia et al. (2016) constructed an extension of the Hyvärinen scoring rule to compact oriented Riemannian manifolds.

In this paper, for parametric estimation on spheres, we propose an easy-to-construct class of strictly proper scoring rules that does not require the normalizing constant. Our class contains the Hyvärinen scoring rule (Mardia et al., 2016) and preserves three major statistical properties that the Hyvärinen scoring rule has; The first is being strictly proper. Since our

* Corresponding author.
E-mail addresses: yuya_takasu@mist.i.u-tokyo.ac.jp (Y. Takasu), yano@mist.i.u-tokyo.ac.jp (K. Yano), komaki@mist.i.u-tokyo.ac.jp (F. Komaki).

https://doi.org/10.1016/j.spl.2018.02.054

scoring rules are strictly proper, the minimum score estimators are typically consistent; see Remark 1 and Dawid (2007). The second is avoiding the calculation of the normalizing constant. The third is orthogonal invariance. Our scoring rules are invariant under orthogonal transformation of samples.

Relationship to the prior work (Mardia et al., 2016; Ehm and Gneiting, 2012; Parry et al., 2012) should be mentioned. As mentioned above, our class contains the Hyvärinen scoring rule (Mardia et al., 2016). Detailed discussions are presented in Section 4. Numerical experiments in Section 5 report that there is a room in our class for improvement with respect to the Hyvärinen scoring rule. Note that our numerical experiments are limited and the generalization takes more experiments. We also give a comment on a difference between this paper and Ehm and Gneiting (2012) and Parry et al. (2012). The main difference between this paper and Ehm and Gneiting (2012) and Parry et al. (2012) is that we provide an easy-to-construct class of proper scoring rules instead of characterizing all proper scoring rules that do not require the normalizing constant.

2. Preparation

We summarize additional terminologies related to scoring rules.

First, we define two-locality and homogeneity of scoring rules. These properties of scoring rules ensure that the scoring rules do not require the normalizing constant. For the reference of the definitions, see Ehm and Gneiting (2012) and Parry et al. (2012).

Definition 1 (*Two-locality*). A two-local scoring rule S is a scoring rule represented by $S(z,f) = s(z, f(z), \nabla\widetilde{f}(z), \nabla^2\widetilde{f}(z))$ using $s : \mathbb{S}^p \times \mathbb{R}_+ \times \mathbb{R}^{p+1} \times \mathbb{R}^{(p+1)\times(p+1)} \to \mathbb{R} \cup \{\infty\}$ for all $z \in \mathbb{S}^p$ and all $f \in \mathcal{P}$, where $\widetilde{f}$ is a domain extension of f to a function on $\mathbb{R}^{p+1} \setminus \{0\}$ such that $\widetilde{f}(z) := f(z/\sqrt{z^\top z})$, ∇ is the gradient operator on $\mathbb{R}^{p+1}$, and $\nabla^2\widetilde{f} = \nabla(\nabla\widetilde{f})^\top$.

Definition 2 (*Homogeneity*). A two-local scoring rule is said to be homogeneous if $s(z, f(z), \nabla\widetilde{f}(z), \nabla^2\widetilde{f}(z)) = s(z, \lambda f(z), \lambda\nabla\widetilde{f}(z), \lambda\nabla^2\widetilde{f}(z))$ for an arbitrary positive constant λ.

By definition, the estimation using a homogeneous scoring rule does not require the normalizing constant.

Second, we define orthogonal invariance of scoring rules and models.

Definition 3 (*Orthogonal Invariance*). A scoring rule S on $\mathbb{S}^p$ is said to be orthogonally-invariant if $S(Vz, f \circ V^\top) = S(z,f)$ for any orthogonal matrix V, where $f \circ V^\top$ is the induced measure of f from V. A parametric model $\mathcal{P}$ is said to be orthogonally-invariant if $f \circ V^\top \in \mathcal{P}$ for each $f \in \mathcal{P}$ and for each orthogonal transformation V.

3. Proposed scoring rules

In this section, for parametric estimation on $\mathbb{S}^p$, we introduce a useful class of strictly proper scoring rules that does not require the normalizing constant. In what follows, we work with the following assumption on the model $\mathcal{P}$.

Assumption 1. Each element in $\mathcal{P}$ has a strictly positive and twice continuously differentiable density f with respect to the uniform measure μ.

Let Φ be a function from $\mathbb{S}^p \times \mathbb{R}^{p+1}$ to $\mathbb{R}$ that satisfies the following assumption. Recall that for a function f on $\mathbb{S}^p$, $\widetilde{f}$ is the domain extension of f such that $\widetilde{f}(z) = f(z/\sqrt{z^\top z})$.

Assumption 2. For each $z \in \mathbb{S}^p$, $\Phi(z,\cdot)$ is strictly concave and twice continuously differentiable, and the derivative of $\widetilde{\Phi}(z,u)$ with respect to the second argument is differentiable with respect to the first argument for all $(z,u) \in (\mathbb{R}^{p+1} \setminus \{0\}) \times \mathbb{R}^{p+1}$.

For a function Φ satisfying Assumption 2, let $S_\Phi : \mathbb{S}^p \times \mathcal{P} \to \mathbb{R}$ be defined by

$$\begin{aligned} S_\Phi(z,f) =& \Phi(z, \nabla\log\widetilde{f}(z)) - (\nabla\log\widetilde{f}(z))^\top(\nabla_2\Phi)(z, \nabla\log\widetilde{f}(z)) \\ &- \mathrm{tr}[\nabla\{\nabla_2\widetilde{\Phi}(z, \nabla\log\widetilde{f}(z))\}^\top] \\ &- \mathrm{tr}((\nabla_2^2\Phi)(z, \nabla\log\widetilde{f}(z))\nabla^2\log\widetilde{f}(z)) + pz^\top(\nabla_2\Phi)(z, \nabla\log\widetilde{f}(z)), \end{aligned} \quad (1)$$

where $(\nabla_2\Phi)(z,u) := (\partial\Phi(z,u)/\partial u^1, \ldots, \partial\Phi(z,u)/\partial u^{p+1})^\top$ and tr is the trace of a $(p+1)\times(p+1)$ matrix.

The following theorem tells us that S_Φ is a strictly proper scoring rule that does not require the normalizing constant and that is orthogonally-invariant, which is our main result. We denote by $\|\cdot\|$ the standard norm in $\mathbb{R}^{p+1}$.

Theorem 1. *Assume that $\mathcal{P}$ satisfies Assumption 1 and that Φ satisfies Assumption 2. Then, the scoring rule S_Φ defined by (1) is strictly proper, two-local, and homogeneous. Moreover, if $\mathcal{P}$ is orthogonally-invariant, and if Φ has the form $\Phi(z,u) = r(\|u\|^2)$ with a twice continuously differentiable function $r : \mathbb{R}_+ \to \mathbb{R}$, S_Φ is orthogonally-invariant.*