Journal of Multivariate Analysis 150 (2016) 75–90

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Classification into Kullback–Leibler balls in exponential families

Alexander Katzur, Udo Kamps *

Institute of Statistics, RWTH Aachen University, D-52056 Aachen, Germany

ARTICLE INFO

Article history:
Received 26 November 2015
Available online 26 May 2016

AMS subject classifications:
primary 62H30
secondary 62B10

Keywords:
Bregman balls
Bregman divergence
Classification
Convex duality
Exponential families
Functions of Legendre type
(generalized) Chernoff information
Itakura–Saito balls
Itakura–Saito divergence
Kullback–Leibler divergence
(lower dimensional) Kullback–Leibler balls
Minimal enclosing balls
Multivariate normal distribution
Sequential order statistics

ABSTRACT

A classification procedure for a two-class problem is introduced and analyzed, where the classes of probability density functions within a regular exponential family are represented by left-sided Kullback–Leibler balls of natural parameter vectors. If the class membership is known for a finite number of densities, only, classes are defined by constructing minimal enclosing left-sided Kullback–Leibler balls, which are seen to uniquely exist. A connection to Chernoff information between distributions is pointed out.

1. Introduction

The research area of classification and discrimination has gained major attention since R.A. Fisher has published his ground-breaking work on the classification of species of iris (see [20]). Several well-known books on multivariate analysis such as Rencher [40] and Anderson [4] provide overviews of classical classification approaches, and several monographs have been published focusing on this topic (e.g. [31,22,34]). A variety of articles can be found when the underlying class conditional distribution is multivariate normal or multinomial (see, e.g., [17,34]). For other distributions the amount of literature is significantly smaller. Taniguchi [44] analyzes the discriminant function in exponential families of distributions. Classification based on distance measures and each class consisting of a family of underlying distributions has been considered by Kullback [29], p. 85, Matusita [32] and Menéndez et al. [35].

In the present paper, a classification approach for exponential families is suggested. The use of a divergence measure in a classification method within some exponential family was proposed by Kullback [29]. Cacoullos [12] studied classification of

* Corresponding author.
E-mail address: udo.kamps@rwth-aachen.de (U. Kamps).

http://dx.doi.org/10.1016/j.jmva.2016.05.007

observations to known normal populations by means of the Mahalanobis distance. Matusita [33] proposed the application of distance functions or affinities to classify a set of given observations with respect to two predefined distributions. Cacoullos and Koutras [13] extended Matusita's approach to spherical distributions, and Koutras [28] considered discrimination rules for elliptical normal mixtures by means of a transformed Matusita affinity. Cacoullos and Koutras [14] investigated minimum distance classification via Mahalanobis distance when the underlying distribution departs from multivariate normality; they examine Kotz-type elliptical distributions. Menéndez et al. [35] studied classes consisting of a finite number of distributions with unknown parameters, which are replaced by their respective maximum likelihood estimators; they aimed at classifying another distribution to the class having smallest distance, which is measured by an f-dissimilarity as introduced by Györfi and Nemetz [21]. For other fields of statistical inference with divergence measures, we refer to Pardo [39]. Taniguchi [44] considered classifying a multivariate observation into one of two populations being members of an exponential family; log likelihood ratio classification statistics were used with plug-in estimates.

We consider two-class classification where each class within an exponential family has an underlying set of distributions which forms a left-sided Kullback–Leibler ball (KL ball). Classification is then performed by measuring the distances of the object to be classified to the two KL balls of distributions by means of the Kullback–Leibler divergence. If in the actual KL ball some representatives are known, only, then minimum enclosing (lower-dimensional) KL balls in exponential families are considered. Several results for those minimal enclosing balls are provided and an interesting one-to-one connection between those balls and a generalized version of the Chernoff information is proven. In a simulation study, the proposed procedure is analyzed by means of misclassification probabilities. More details on classification in exponential families can be found in [27].

2. Exponential families and divergence measures

We consider an r-*parametric exponential family* $\mathfrak{P} = \{P_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \Theta\}, \Theta \subseteq \mathbb{R}^r$, of distributions, where all member distributions have a probability density function (pdf) of the form

$$f_{\boldsymbol{\alpha}}(\boldsymbol{x}) = f(\boldsymbol{x}; \boldsymbol{\alpha}) = \exp\left\{\boldsymbol{\alpha}^\top \boldsymbol{T}(\boldsymbol{x}) - \kappa(\boldsymbol{\alpha})\right\} h(\boldsymbol{x}) \mathbf{1}_{\mathfrak{X}}(\boldsymbol{x}),$$

with respect to a σ-finite measure ν. Herein, $\mathfrak{X} \subseteq \mathbb{R}^p$ is the support of $\mathfrak{P}$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_r)^\top$ is called the *natural parameter*, $\boldsymbol{T}(\boldsymbol{x}) = (T_1(\boldsymbol{x}), \dots, T_r(\boldsymbol{x}))^\top$ is called the *natural statistic* and $\kappa(\boldsymbol{\alpha})$ and $h(\boldsymbol{x})$ are real valued functions. Furthermore, $h(\boldsymbol{x})$ and the indicator function $\mathbf{1}_{\mathfrak{X}}$ are measurable and independent of $\boldsymbol{\alpha}$. κ is sometimes referred to as the *cumulant function* or the *log-partition function*.

In the present paper it is assumed that the exponential family under consideration has a minimal representation in the sense that the statistics $T_1, \dots, T_r$ are $\mathfrak{P}$-affine independent, i.e., for $\boldsymbol{c} \in \mathbb{R}^r$ and $c_0 \in \mathbb{R}$ the implication

$$\boldsymbol{c}^\top \boldsymbol{T}(\boldsymbol{x}) = c_0 \quad \mathfrak{P}\text{-a.e.} \quad \Longrightarrow \quad \boldsymbol{c} = \mathbf{0}, \quad c_0 = 0$$

holds true. Furthermore, the natural parameter space Θ is supposed to be open. In the following, a family fulfilling those two conditions is called *regular exponential family*. In regular exponential families κ is a strictly convex function on Θ, $\boldsymbol{\pi} := \nabla\kappa$ is bijective from Θ to $\boldsymbol{\pi}(\Theta)$, $\nabla^2\kappa(\boldsymbol{\alpha})$ is positive definite for each $\boldsymbol{\alpha} \in \Theta$, and the parameters are identifiable for the distributions (see, e.g., [9]). For convenience, e.g., when applying maximum likelihood estimation and dealing with conjugate functions and for its use in Section 7, we will restrict ourselves to regular exponential families.

As examples, we consider multivariate normal distributions with a known covariance matrix, as well as sequential order statistics, which extend the model of common order statistics to a general, parametrized class of models of ordered random variables. For these parameters, maximum likelihood estimators based on independent samples can be explicitly stated (see, e.g., [26,16,8,45]).

Examples 1. (i) For a positive definite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{r\times r}$, the family of r-dimensional normal distributions $\mathfrak{P}^{(\mathcal{N})} := \{P_{\mathcal{N}_r(\boldsymbol{\alpha},\boldsymbol{\Sigma})} : \boldsymbol{\alpha} \in \mathbb{R}^r\}$ forms a regular r-parametric exponential family where the natural parameter is $\boldsymbol{\alpha}$, $\Theta = \mathbb{R}^r$, $\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{\Sigma}^{-1}\boldsymbol{x}$, $\mathfrak{X} = \mathbb{R}^r$ and $\kappa(\boldsymbol{\alpha}) = (1/2)\boldsymbol{\alpha}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}$.

(ii) Sequential order statistics (SOSs) with known absolutely continuous baseline cumulative distribution function (cdf) F with $\text{supp}(F) = \mathbb{R}_+ := (0, \infty)$, form a regular r-parametric exponential family $\mathfrak{P}^{(SOS)} := \{P^{(SOS)}_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathbb{R}^r_+\}$ (see [26,8]). SOSs are designed to model the failure times of some $(n-r+1)$-out-of-n systems, where failures affect the lifetime distributions of remaining components. These impacts are described by the model parameters $\alpha_1, \dots, \alpha_r > 0$. The natural parameter space of SOSs is given by $\Theta = \mathbb{R}^r_+$, $\mathfrak{X} = \{\boldsymbol{x} \in \mathbb{R}^r_+ : F^{-1}(0+) < x_1 < \cdots < x_r < F^{-1}1\}$, the natural statistic $\boldsymbol{T}(\boldsymbol{x}) = (T_1(\boldsymbol{x}), \dots, T_r(\boldsymbol{x}))^\top$ is defined by

$$T_1(\boldsymbol{x}) = n\ln(1 - F(x_1)),$$

$$T_j(\boldsymbol{x}) = (n - j + 1)\ln\left(\frac{1 - F(x_j)}{1 - F(x_{j-1})}\right), \quad 2 \le j \le r,$$

and

$$h(\boldsymbol{x}) = \frac{n!}{(n-r)!}\prod_{j=1}^{r}\frac{f(x_j)}{1 - F(x_j)},$$