Journal of Statistical Planning and Inference 140 (2010) 539 – 550

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Principal point classification: Applications to differentiating drug and placebo responses in longitudinal studies

Thaddeus Tarpey[a,*], Eva Petkova[b]

[a] Department of Mathematics and Statistics, Wright State University, Dayton, OH, USA
[b] Child Study Center, School of Medicine, New York University, New York, NY 10016-6023, USA

ARTICLE INFO

Article history:
Received 3 January 2009
Received in revised form
14 July 2009
Accepted 29 July 2009
Available online 12 August 2009

Keywords:
Best linear unbiased predictors
Cluster analysis
Discriminant analysis
k-means algorithm
Logistic regression
Mixed effects models
Placebo effect

ABSTRACT

Principal points are cluster means for theoretical distributions. A discriminant methodology based on principal points is introduced. The principal point classification method is useful in clinical trials where the goal is to distinguish and differentiate between different treatment effects. Particularly, in psychiatric studies where placebo response rates can be very high, the principal point classification is illustrated to distinguish specific drug responders from non-specific placebo responders.

1. Introduction

In typical applications of discriminant analysis, there is training data from two or more groups and the goal is to define a discriminant function to classify new observations to the correct groups. The goal of this paper is a bit different from the classical discriminant analysis problem. Consider a clinical trial with an active drug arm and a placebo arm. The problem of interest is to determine which subjects in the active drug arm are responding primarily to a non-specific (placebo) effect rather than the specific effect of the drug. Because we know which subjects receive the active drug or the placebo, there is no ambiguity as to which treatment group the subjects belong. However, if a drug-treated subject responds, we do not know to what degree the subject responded due to the specific (drug) effect and to non-specific (placebo) effects of the treatment. If a placebo treated subject responds, then we know it must be due to the non-specific effects of the treatment. Outside of clinical experiments in everyday treatment, patients typically receive the active drug for treatment, not a placebo. Interest lies in classifying drug treated patients who respond primarily due to specific (drug) effects or to non-specific (placebo) effects, or perhaps a combination of these two effects.

The problem of distinguishing a placebo response from a drug response in psychiatric illnesses has been of high interest for clinical research and practice for many years (e.g., see Quitkin et al., 1987; Ross et al., 2002). There is a need to apply modern statistical methods to address this important problem. One approach to the specific/non-specific treatment effects problem is to assume that the population consists of distinct latent subgroups of specific responders and non-specific responders.

* Corresponding author. Tel.: +1 937 775 2861; fax: +1 937 775 2081.
E-mail address: thaddeus.tarpey@wright.edu (T. Tarpey).

0378-3758/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2009.07.030

For longitudinal studies, a growth mixture model can be postulated (Muthén and Shedden, 1999; James and Sugar, 2003; Elliot et al., 2005). However, if subjects can experience both specific and non-specific effects ranging continuously from very weak to very strong, then modeling the outcomes as a growth mixture could erroneously lead investigators to believe real sub-populations exist. In this paper, an alternative strategy is implemented by determining an optimal partition of the underlying distribution. The partitioning method is based on determining optimal cluster centers, called principal points (Flury, 1990), for theoretical distributions.

In the classical normal theory discriminant analysis, observations are assigned to the different groups based on their proximity to the group means. In longitudinal studies, the outcomes of interest are curves over time. Focusing only on the mean curve ignores the fact that groups often contain a variety of distinct curve shapes. For instance, in longitudinal clinical studies, there may exist different types of outcome profile shapes over time corresponding to different types of response to treatment. The principal point classification method developed in this paper assigns observations based on proximity to principal points which can be regarded as a generalization of the mean from one to several points. In addition, if we know the outcome distributions differ for different groups (e.g., different treatment arms in a clinical trial), the principal point classification method developed below can identify prototypical outcome profiles that can distinguish these differences.

In Section 2 we define principal points and discuss methods of estimating principal points. The principal point classification method is described in Section 3. Results of a simulation experiment comparing principal point classification with normal theory discriminant analysis are provided in Section 4. Principal points for linear mixed effect models are described in Section 5. The principal point classification method is used to distinguish two drug therapies for depression (fluoxetine and imipramine) in Section 6. Additionally, outcome profiles from specific and non-specific effects are distinguished using the principal point classification methodology in this section as well. Finally, the paper is concluded in Section 7.

2. Principal points

A classic statistical problem is to determine an optimal partition of a continuous distribution (Cox, 1957; Connor, 1972; Dalenius, 1950; Dalenius and Gurney, 1951). In signal processing, this problem is referred to vector quantization (e.g., Graf and Luschgy, 2000). In an optimal stratification of a distribution into k strata, the means of the k strata are called the k principal points of the distribution (Flury, 1990).

Let $\boldsymbol{X}$ denote a continuous random vector and consider k points $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$ to be used to represent the distribution of $\boldsymbol{X}$. We can define a k-point approximation $\boldsymbol{Y}$ to $\boldsymbol{X}$ as

$$\boldsymbol{Y} = \boldsymbol{\xi}_j \quad \text{if } \|\boldsymbol{X} - \boldsymbol{\xi}_j\| < \|\boldsymbol{X} - \boldsymbol{\xi}_h\| \quad \text{for } h \neq j.$$

$\boldsymbol{Y}$ is a *self-consistent* approximation to $\boldsymbol{X}$ if $E[\boldsymbol{X}|\boldsymbol{Y}] = \boldsymbol{Y}$ a.s. (Tarpey and Flury, 1996) in which case the points $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$ are called k self-consistent points of $\boldsymbol{X}$ (Flury, 1993). Distributions, particularly multivariate distributions, may have more than one set of k self-consistent points (Tarpey, 1998). If $E\|\boldsymbol{X} - \boldsymbol{Y}\|^2 \leq E\|\boldsymbol{X} - \boldsymbol{Y}^*\|^2$ for any other k-point approximation $\boldsymbol{Y}^*$ to $\boldsymbol{X}$, then the points $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$ are called k-principal points of $\boldsymbol{X}$. Flury (1993) showed that a set of k principal points of a distribution must be self-consistent points.

For $k = 1$, the single principal point corresponds to the mean of the distribution. For $k > 1$, the k principal points provide a k-point generalization of the mean from one to several points. For a $\mathrm{N}(\mu, \sigma^2)$ distribution, the $k = 2$ principal points are $\mu \pm \sqrt{2/\pi}\sigma$ (Flury, 1990); for $k > 2$ the principal points must be found numerically. Univariate distributions with log-concave densities have a unique set of k self-consistent points for each k and this unique self-consistent approximation must correspond to the k principal points (Truskin, 1982; Kieffer, 1982; Tarpey, 1994).

2.1. Estimation of principal points

Given a set of observations from a distribution, non-parametric estimators of k principal points can be obtained by using the cluster means from running the k-means algorithm (e.g., MacQueen, 1967; Hartigan, 1975; Hartigan and Wong, 1979). Under general circumstances, the cluster means from the k-means algorithm are strongly consistent estimators of the principal points (Pollard, 1981) and asymptotically normally distributed (Pollard, 1982). More efficient methods of estimating principal points can be obtained by utilizing distributional assumptions (e.g., Flury, 1993; Tarpey, 1997). For instance, maximum likelihood estimators of $k = 2$ principal points of a univariate normal distribution are $\bar{x} \pm \sqrt{2/\pi}s$.

For larger values of k and for multidimensional distributions, analytical formulas for principal points do not exist. Tarpey (2007b) describes a *parametric k-means algorithm* that provides a computationally intensive but easy to utilize method of determining maximum likelihood estimators of k principal points, similar to the algorithm of Linde et al. (1980) from the vector quantization literature using a known distribution. Suppose the data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ comes from a distribution with density $f(\boldsymbol{x}; \boldsymbol{\theta})$. For the parametric k-means algorithm, first obtain a maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. Next, run the k-means algorithm on a very large data set simulated from $f(\boldsymbol{x}; \hat{\boldsymbol{\theta}})$. Tarpey (2007b) showed that the cluster means obtained by running the k-means algorithm on the large simulated data set are (approximately) maximum likelihood estimators of the k principal points of $f(\cdot; \boldsymbol{\theta})$ distribution.