

The kurtosis coefficient and the linear discriminant function

Daniel Peña*, Francisco J. Prieto

Universidad Carlos III de Madrid, Department of Statistics and Econometrics, c/Madrid, 126, Getafe 28903, Spain

Received July 1999; received in revised form September 1999

Abstract

In this note we analyze the relationship between the direction obtained from the minimization of the kurtosis coefficient of the projections of a mixture of multivariate normal distributions and the linear discriminant function. We show that both directions are closely related and, in particular, that given two vector random variables having symmetric distributions with unknown means and the same covariance matrix, the direction which minimizes the kurtosis coefficient of the projection is the linear discriminant function. This result provides a way to compute the discriminant function between two normal populations in the case in which means and common covariance matrix are unknown. © 2000 Elsevier Science B.V. All rights reserved

Keywords: Classification; Kurtosis coefficient

1. Introduction

Given two multivariate normal populations, $N_i(\mu_i, V)$, $i = 1, 2$, it is well known that the optimal way to classify a new element X , coming from any of these two populations with equal a priori probabilities, is to project the means (μ_1, μ_2) of the two populations onto the direction $u = V^{-1}(\mu_1 - \mu_2)$, and to classify the observation X by considering which of the two projected means, $u'\mu_1, u'\mu_2$, is closer to the projected value $u'X$. This linear discriminant function was first obtained by Fisher (1936) and has the two following properties: (1) it is implied by a decision rule that classifies the observation by using the Mahalanobis distance between the point and the two means; (2) it produces the maximum separation between the projected means with respect to the variance of the projected distribution.

The generalization of this idea when we have p normal populations leads to a best discriminant space associated with the eigenvalues of the matrix $V^{-1}B$, where B denotes the between groups matrix (see for instance, Seber, 1984).

We will show in this note that the direction obtained by minimizing the kurtosis coefficient of the projected distributions is closely related to the linear discriminant space, and it coincides with the linear discriminant

* Corresponding author.

E-mail addresses: dpena@est-econ.uc3m.es (D. Peña), fjp@est-econ.uc3m.es (F.J. Prieto).

function in the case of two equiprobable populations. This result, in addition to its mathematical interest, has two interesting implications: The first one is to emphasize that the kurtosis coefficient not only measures the thickness of the tails of the distribution but it is also a measure of the lack of bimodality in the distribution. In our case, it reaches its minimum value when the projected observations follow a bimodal univariate distribution. The second, is to show that, if we have two normal distributions with unknown different means but the same covariance matrix, by minimizing the kurtosis coefficient of the projections we can obtain the (Fisher) best linear discriminant function without knowing neither the means of the two distributions nor their common covariance matrix, which we believe is a remarkable fact.

2. The model

Consider a random variable x obtained as a mixture of p normal populations $N_i(\mu_i, V)$ with different mean vectors μ_i and the same covariance matrix V , and having mixture coefficients α_i , $\sum_i \alpha_i = 1$. We assume that the matrix V has full rank.

We are interested in those directions that minimize the kurtosis coefficient of the projected observations. Consider a projection direction u ; denote the projected observations by $z = u'x$. For each one of the individual distributions we will have $z_i = u'x_i$, where z_i follows a normal distribution $N_i(m_i, u'Vu)$, $m_i = u'\mu_i$. Note also that $E(x) = \mu = \sum_i \alpha_i \mu_i$.

The mean and variance of the projected observations will be given by

$$E(z) = \bar{m} = \sum_i \alpha_i m_i = u'\mu,$$

$$E(z - E(z))^2 = \sum_i \alpha_i E(z_i - m_i)^2 + \sum_i \alpha_i (m_i - \bar{m})^2 = u'(V + B)u,$$

where $B_i = (\mu_i - \mu)(\mu_i - \mu)'$ and $B = \sum_i \alpha_i B_i$.

As a consequence, the fourth central moment $k \equiv E(z - E(z))^4$ will be given by

$$k = \sum_i \alpha_i E(z_i - m_i + m_i - \bar{m})^4$$

$$= \sum_i \alpha_i E(z_i - m_i)^4 + 6 \sum_i \alpha_i (m_i - \bar{m})^2 E(z_i - m_i)^2 + \sum_i \alpha_i (m_i - \bar{m})^4$$

$$= 3(u'Vu)^2 + 6(u'Vu)(u'Bu) + \sum_i \alpha_i (m_i - \bar{m})^4$$

$$= 3(u'(V + B)u)^2 - 3(u'Bu)^2 + \sum_i \alpha_i (u'B_i u)^2. \tag{1}$$

In what follows, we will also use the second and fourth moments for the univariate distribution of the projected means, and its kurtosis coefficient, that we will denote as

$$\sigma_m^2 = \sum_i \alpha_i (m_i - \bar{m})^2 = u'Bu,$$

$$k_m = \sum_i \alpha_i (m_i - \bar{m})^4 = \sum_i \alpha_i (u'B_i u)^2,$$

$$\kappa_m = \frac{k_m}{\sigma_m^4} = \frac{\sum_i \alpha_i (u'B_i u)^2}{(u'Bu)^2}.$$

We are interested in the projection direction obtained as the solution for the problem

$$\begin{aligned} \min_u \quad & k(u) \\ \text{s.t.} \quad & u'(V + B)u = 1, \end{aligned} \tag{2}$$

that is, the direction minimizing the kurtosis coefficient.

This direction can be characterized in terms of the gradient of $k(u)$. From (1), the gradient will take the value

$$\nabla k(u) = 12(u'(V + B)u)(V + B)u - 12(u'Bu)Bu + 4 \sum_i \alpha_i (u'B_i u) B_i u. \tag{3}$$

3. A property of the kurtosis coefficient

Lemma 1. For a vector random variable x defined as a mixture of p multivariate normal distributions $N_i(\mu_i, V)$, $i = 1, \dots, p$ with different mean vectors μ_i , the same covariance matrix V and probabilities α_i , the direction which minimizes the kurtosis coefficient is given by the eigenvector linked to the largest eigenvalue of the matrix $V^{-1}\bar{B}$, where

$$\bar{B} = \sum_i \omega_i B_i,$$

and

$$\omega_i = \alpha_i \left(\frac{3\sigma_m^2 - (m_i - \bar{m})^2}{\sigma_m^4} + \kappa_m - 3 \right).$$

Proof. The direction u that minimizes the kurtosis coefficient of the projections will be the solution of (2), and must satisfy the corresponding first-order condition $\nabla k(u) = 2\lambda(V + B)u$, which from (3) implies

$$12(u'(V + B)u)(V + B)u - 12(u'Bu)Bu + 4 \sum_i \alpha_i (u'B_i u) B_i u = 2\lambda(V + B)u. \tag{4}$$

The value of λ can be obtained after multiplying the equation by u and using the constraint, and is given by

$$\lambda = 6 - 6(u'Bu)^2 + 2 \sum_i \alpha_i (u'B_i u)^2.$$

Replacing this value, using the constraint on u in (4) and rearranging terms we have

$$\left(3u'Bu - 3(u'Bu)^2 + \sum_j \alpha_j (u'B_j u)^2 \right) Bu - \sum_i \alpha_i (u'B_i u) B_i u = \left(3(u'Bu)^2 - \sum_j \alpha_j (u'B_j u)^2 \right) Vu.$$

Premultiplying this equation by V^{-1} , and replacing the definitions of B and the moments for the distribution of the means, we obtain that the solution of (2) must satisfy

$$V^{-1} \sum_i \left[\alpha_i \left(\frac{3}{\sigma_m^2} - 3 + \kappa_m - p_i \right) \right] B_i u = (3 - \kappa_m)u \tag{5}$$

for $p_i = (u'B_i u)/\sigma_m^4 = (m_i - \bar{m})^2/\sigma_m^4$. Note that from (1), the kurtosis coefficient of the projected variables can be written as a function of the distributions of the means as

$$k = 3 - \sigma_m^4(3 - \kappa_m),$$

which implies that in order to minimize k , as σ_m^4 is always positive, we must make σ_m^4 as large as possible and κ_m as small as possible. Thus, the direction u in (5) is the eigenvector linked to the largest eigenvalue of the matrix $V^{-1}\bar{B}$, proving the result. \square

Lemma 2. For the vector random variable x in Lemma 1, if the direction which minimizes the kurtosis coefficient of the projections u satisfies for some scalar $\beta(u)$ that

$$u' B_i u = \beta(u) \quad \forall i, \tag{6}$$

then u is proportional to the largest eigenvector of the matrix $V^{-1}B$.

Proof. This is an immediate consequence of Lemma 1. From (5), if (6) holds then it must hold that $p_i = \beta/\sigma_m^4$ and

$$V^{-1} \sum_i \alpha_i \left(\frac{3}{\sigma_m^2} - 3 + \kappa_m - \frac{\beta}{\sigma_m^4} \right) B_i u = (3 - \kappa_m)u$$

and this implies that

$$V^{-1}Bu = \omega u,$$

for $\omega = \sigma_m^4(3 - \kappa_m)(3\sigma_m^2 - \beta - \sigma_m^4(3 - \kappa_m))^{-1}$, and this proves the desired result. \square

Theorem 1. For a random variable corresponding to the mixture of two normal populations $N_i(\mu_i, V)$, $i=1, 2$, having different mean vectors μ_i , the same covariance matrix V and equal probabilities $\alpha_i = \frac{1}{2}$, the direction which minimizes the kurtosis coefficient is the linear discriminant function $V^{-1}(\mu_1 - \mu_2)$.

Proof. For a mixture of two normal random variables with weights $\alpha_1 = \alpha_2 = 1/2$ it holds that

$$\mu_1 - \mu = \frac{1}{2}(\mu_1 - \mu_2), \quad \mu_2 - \mu = -\frac{1}{2}(\mu_1 - \mu_2)$$

and also

$$u' B_1 u = (u'(\mu_1 - \mu))^2 = u' B_2 u.$$

This implies the satisfaction of condition (6), and from Lemma 2, the direction that minimizes the kurtosis coefficient is the largest eigenvector of $V^{-1}B$.

In our case, B has the form

$$B = \frac{1}{4}(\mu_1 - \mu_2)(\mu_1 - \mu_2)',$$

and the eigenvectors of the matrix $V^{-1}B$ are either proportional to $v = V^{-1}(\mu_1 - \mu_2)/\eta$, where $\eta^2 = (\mu_1 - \mu_2)'V^{-1}(V + B)V^{-1}(\mu_1 - \mu_2)$ is chosen to ensure the satisfaction of the constraint in (2), or are vectors w orthogonal to $\mu_1 - \mu_2$ that also satisfy the constraint $w'(V + B)w = 1$.

To analyze which solutions are minimizers, we consider the objective function of (2). Using (1) and taking into account that $B_i w = 0$,

$$k(w) = 3.$$

For v , replacing the expressions for B and B_i in terms of $\mu_1 - \mu_2$,

$$\begin{aligned} k(v) &= 3 - 3(v' B v)^2 + \frac{1}{2} \sum_i (v' B_i v)^2 = 3 - \frac{3}{16}(v'(\mu_1 - \mu_2))^4 + \frac{1}{16}(v'(\mu_1 - \mu_2))^4 \\ &= 3 - \frac{1}{8\eta^4}((\mu_1 - \mu_2)'V^{-1}(\mu_1 - \mu_2))^4 < 3. \end{aligned}$$

This shows that the linear discriminant direction $V^{-1}(\mu_1 - \mu_2)$ is the minimizer of the kurtosis coefficient. \square

Finally, we show that these results do not depend on the hypothesis of normality for the data.

Corollary 1. *The results stated in Lemmas 1, 2 and Theorem 1 also stand if the vector random variable x is defined as a mixture of multivariate ellipsoidal distributions with different means μ_i and the same covariance matrix V , having a sufficiently large (univariate) kurtosis coefficient b .*

Proof. The main change in the preceding relationships corresponds to the expression for the fourth central moment (1), that now takes the form

$$\begin{aligned} k &= b(u'Vu)^2 + 6(u'Vu)(u'Bu) + \sum_i \alpha_i(u'B_iu)^2 \\ &= b(u'(V+B)u)^2 - b(u'Bu)^2 + (6-2b)(u'Vu)(u'Bu) + \sum_i \alpha_i(u'B_iu)^2. \end{aligned}$$

for some constant $b \geq 1$, independent of u .

The results in Lemmas 1 and 2 follow from the fact that b is independent of u , as they do not depend on the particular value of b .

For Theorem 1, the only change is in the value of the objective function for the different eigenvectors of the matrix $V^{-1}B$. Using the same notation as in Theorem 1 we now have $k(w) = b$ and

$$\begin{aligned} k(v) &= b + (6-2b)(v'Vv)\frac{1}{4}(v'(\mu_1 - \mu_2))^2 + \frac{1}{16}(v'(\mu_1 - \mu_2))^4 - \frac{b}{16}(v'(\mu_1 - \mu_2))^4 \\ &= b + \frac{\delta^3}{16\eta^4}(24 - 8b + (1-b)\delta), \end{aligned}$$

where $\delta = (\mu_1 - \mu_2)'V^{-1}(\mu_1 - \mu_2)$.

This implies $k(w) > k(v)$ as long as $24 - 8b + (1-b)\delta < 0$, or equivalently as long as $b > 3 - 2\delta/(8 + \delta)$. Note that for $b = 3$ this condition always holds, and also that if b is smaller than $3 - 2\delta/(8 + \delta)$, we have a maximizer along the linear discriminant function. \square

Acknowledgements

This research has been sponsored by the Cátedra BBV of Quality and by DGES Grant PB96-0111. We are very grateful to an anonymous referee for her/his help in improving the presentation of this article.

References

- Fisher, R.A., 1936. The use of multiple measurement in taxonomic problems. *Ann. Eugen.* 7, 179–188.
Seber, G.A.F., 1984. *Multivariate Observations*. Wiley, New York.