INVARIANT TESTS OF DISCRIMINANT COEFFICIENTS IN CLASSIFICATION PROBLEMS

N. GIRI

University of Montreal, Canada.

Introduction

The classification problem arises when one makes a number of observations on an individual or an object and then wants to classify the individual (or the object) into one of several categories on the basis of these observations. Here we will be concerned with two categories and a situation where these categories are completely specified by densities p_1, p_2. Let $\underline{z}=(z_1, ..., z_p)'$ be the observation vector on the individual. We know for certainty that it comes from either p_1 or p_2. Our problem is to decide from which population did he arise? Let a_1 denote the action that $\underline{z}$ comes from p_1 and a_2 denote the action that $\underline{z}$ comes from p_2 and let c_1 be the penalty of taking action a_1 when $\underline{z}$ actually belongs to p_2 and let c_2 be the penalty of taking action a_2 when $\underline{z}$ actually belongs to p_1. The penalty of taking correct action is zero in either case.

Let $(\pi_1, 1-\pi_1)$ denotes the aprior probability distribution on the states of nature (p_1, p_2). Then the a posterior probability distribution $(\xi_1, 1-\xi_1)$ on (p_1, p_2) given $\underline{z}$ is given by

$$\xi_1=\frac{\pi_1 p_1(\underline{z})}{\pi_1 p_1(\underline{z})+(1-\pi_1)p_2(\underline{z})}. \quad ...(1{\cdot}1)$$

The posterior expected penalty for taking action a_1 is

$$\frac{(1-\pi_1)p_2(\underline{z})c_1}{\pi_1 p_1(\underline{z})+(1-\pi_1)p_2(\underline{z})}$$

and for taking action a_2 is

$$\frac{\pi_1 p_1(\underline{z})c_2}{\pi_1 p_1(\underline{z})+(1-\pi_1)p_2(\underline{z})}.$$

Thus optimum procedure in this case is to take

action a_1, if $\quad \dfrac{p_2(\underline{z})}{p_1(\underline{z})}<\dfrac{\pi_1 c_2}{(1-\pi_1)c_1}$;

action a_2, if $\quad \dfrac{p_2(\underline{z})}{p_1(\underline{z})}>\dfrac{\pi_1 c_2}{(1-\pi_1)c_1}$.

If

$$\frac{p_2(\underline{z})}{p_1(\underline{z})}=\frac{\pi_1 c_2}{(1-\pi_1)c_1}$$

Then we must randomize to decide the type of action to be taken.

If p_1, p_2 are specified by p-dimensional normal density functions with the same positive definite covariance matrix $\sum$ but different means $\underline{\alpha}=(\alpha_1, .., \alpha_p)'$, $\underline{\beta}=(\beta_1, ..., \beta_p)'$ respectively, then we take

$$\text{action } a_1 \text{ ; if } \underline{z}' \sum^{-1} (\underline{\beta}-\underline{\alpha}) > \tfrac{1}{2}\, tr\, (\underline{\beta}\underline{\beta}'-\underline{\alpha}\underline{\alpha}') \sum^{-1} + \log \frac{\pi_1 c_2}{(1-\pi_1)c_1} \text{ ;}$$

$$\text{action } a_2 \text{ ; if } \underline{z}' \sum^{-1} (\underline{\beta}-\underline{\alpha}) < \tfrac{1}{2}\, tr\, (\underline{\beta}\underline{\beta}'-\underline{\alpha}\underline{\alpha}') \sum^{-1} + \log \frac{\pi_1 c_2}{(1-\pi_1)c_1}. \qquad ...(1 \cdot 2)$$

According to Fisher (1938) the linear function $\underline{z}' \sum^{-1} (\underline{\alpha}-\underline{\beta})$ is known as "discriminant function" and the components of the vector $\sum^{-1} (\beta-\alpha)$ are called "discriminant coefficients".

In actual practice the mean vectors $\underline{\alpha}, \underline{\beta}$ and the covariance matrix $\sum$ are unknown. The problem of statistical inference and estimation of these coefficients are solved on the basis of sample observations $\underline{X}^1, ..., \underline{X}^m$ from the normal population (p-dimensional) with mean $\underline{\alpha}$ and covariance matrix $\sum$ and sample observations $\underline{Y}^1, ..., \underline{Y}^n$ from the normal population (p-dimensional) with mean $\underline{\beta}$ and covariance matrix $\sum$. Sufficiency consideration leads us to restrict our attention to the set of sufficient statistic $(\overline{\underline{X}}=\sum_1^m \underline{X}^\alpha/m,\ \overline{\underline{Y}} = \sum_1^n \underline{Y}^\alpha/n,\ S=\sum_1^m (\underline{X}^\alpha-\overline{\underline{X}})(\underline{X}_\alpha-\overline{\underline{X}})'+\sum_1^n (\underline{Y}^\alpha-\overline{\underline{Y}})(\underline{Y}_\alpha-\overline{\underline{Y}})')$ where $\overline{\underline{X}}, \overline{\underline{Y}}$ are independently distributed p-dimensional normal random variables and S is distributed as a Wishart random variable. For the inference problem, invariance and sufficiency consideration always permit us to consider the statistic $(\overline{\underline{X}}-\overline{\underline{Y}}, \underline{S})$. Since $\overline{\underline{X}}-\overline{\underline{Y}}$ is distributed as p-dimensional normal random variable with mean $\underline{\alpha}-\underline{\beta}$ and covariance matrix $\left(\frac{1}{m}+\frac{1}{n}\right)\sum$, by relabeling variables we can consider the following canonical form where X is normally distributed with mean α and covariance matrix $\sum$ and S is distributed as Wishart with parameter $\sum$ and degrees of freedom $N-1(=n)$, and we want to consider inference problems concerning $\sum^{-1} \underline{\alpha}=\underline{\eta}$.

Let $\eta=(\eta_1, ..., \eta_p)'$ and let $q<p'<p$. We will consider here the following three different problems :

A. To test the null hypothesis $H_{10} : \underline{\eta}=0$ against the alternatives $H_{11} : \eta_{q+1} = ... = \eta_p = 0$ when both $\underline{\alpha}, \sum$ are unknown.

B. To test the null hypothesis $H_{20} : \eta_{q+1} = \ldots = \eta_p = 0$ against the alternatives $H_{11} : \eta \neq 0$, when both $\underline{\alpha}$, Σ are unknown.

C. To test the null hypothesis $H_{30} : \eta_{q+1} = \ldots = \eta_p = 0$ against the alternatives $H_{31} : \eta_{p'+1} = \ldots = \eta_p = 0$ when both $\underline{\alpha}$, $\sum$ are unknown.

We will find here the best invariant tests for these problems. We will show that these best invariant tests are also the likelihood ratio tests. In section 2 we will discuss the principal of invariance and the reduction of these problems in terms of maximal invariants which characterizes the invariant tests. In section 3 we will find these best invariant tests.

2. Invariance

The notation of invariance of a statistical testing problem under a transformation is essentially the same as the notion of invariance in any branch of Mathematics. It is a generally accepted principal that if a problem with an unique solution is invariant under certain transformation then the solution is invariant under that transformation. The main reason for the strong intuitive appeal of invariant decision procedure is the feeling that there should be an unique best way of analysing a collection of statistical information. Nevertheless in cases where the use of an invariant procedure conflicts violently with the desire to make a correct decision with high probability or have a small expected penalty (loss) it must be abandoned.

Lex χ be sample space. β a σ-algebra of subsets of χ and let Ω be the parametric space. Denote by P_θ the probability measure on β corresponding to θ in Ω. Let A be the action space and let L be a real valued function on $\Omega \times A \times \chi$, the loss function (penalty function). A pure decision procedure is a function "d" on χ to A and its associated risk is

$$R(\theta, d) = E_\theta L(\theta, d(x), x) \qquad \ldots(2\cdot1)$$

where E_θ denotes the mathematical expectation when X is distributed according to P_θ. In order that (2·1) should be meaningful we introduce a σ-algebra α of subsets of A and require that L be $\alpha\beta$ measurable in its last two arguments and that d be (β, α) measurable. More precisely

$$\text{for each } \theta \in \Omega \text{ and real } c \qquad \ldots(2\cdot2)$$

$$\{(a, x) : L(\theta, a, x) \leqslant c\} \in \alpha\,\beta$$

where we define $\alpha\beta$ as the smallest σ-algebra containing all cartesian product $A_1 \times B_1$ with $A_1 \in \alpha$ and $B_1 \in \beta$ and

$$\{\chi : d(x) \in A_1\} \in \beta \text{ for all } A_1 \in \alpha. \qquad \ldots(2\cdot3)$$

We also require that L be bounded from below.

We are interested here in testing the hypothesis $H : \theta \in \Omega_H C\Omega$ against the alternatives $H_1 : \theta \in \Omega_{H_1} C\Omega$. In this case the action space A contains only two elements, namely, accept the hypothesis H or reject the hypothesis H (accept H_1). A commonly used loss function in such cases is the normalized $0-1$ loss function, the loss (penalty) of accepting the correct hypothesis being 0 and the associated risks are the two types of error of the testing hypothesis problem. The decision function $d(x)$ is replaced by $\phi(x)$ which is, in Lehmann's terminology (1959), the probability of rejecting the hypothesis. For nonrandomized tests $\phi(x)$ is either 1 or 0. Let g be $1-1$ transformation on χ onto χ, (β, β) measurable in both direction and $\hat{g}$ a $1-1$ funtion on A onto A, (α, α) measurable in both direction. We say that the statistical testing problem we have formulated is invariant under (g, g) if the following conditions hold :

1. There is a function $\bar{g}$ on Ω onto Ω such that for each θ if X is distributed according to P_θ, the gx is distributed according to $P_{\bar{g}\theta}$. This can be expressed by saying that for all $B \in \beta$ the probability that

$$P_{\bar{g}\theta}(B) = P_\theta(g^{-1}B) \text{ (g^{-1} is the inverse of g).} \qquad \ldots(2{\cdot}4)$$

If P is $1-1$, which we shall always assume, it is easy to see that g is $1-1$ onto and is uniqnely determined.

2. $L(\bar{g}\theta, \hat{g}a, gx) = L(\theta, a, x)$ for all $\theta \in \Omega$, $a \in A$ and $x \in \chi$.

3. Under these circumstances the test function $\phi(x)$ will be said to be invariant under g if for all x

$$\phi(gx) = \phi(x). \qquad \ldots(2{\cdot}5)$$

In terms of decision procedure this can be stated as

$$d(gx) = \hat{g}d(x). \qquad \ldots(2{\cdot}6)$$

This says that if we use the test procedure ϕ then we get the same conclusion whether we use x or gx. Roughly speaking two people using essentially the same test procedure but different coordinate system will get the same result. In this context it must be understood that the solution is expressed in terms of the numerical coordinates alone without direct reference to the coordinate system used.

For any two transformations g_1, g_2 on χ onto χ satisfying the conditions $1-3$ above it is now clear that the transformation g_1g_2 and the inverse transformation g_1^{-1}, defined by $(g_1g_2)(x) = g_1(g_2(x))$ and $g_1g_1^{-1}(x) = x$ for all x satisfies $\overline{g_1g_2} = \overline{g_1}\ \overline{g_2}$, $\overline{g_1^{-1}} = \overline{g_1}^{-1}$ and the conditions $1-3$ above. Thus, given a set S of transformation, satisfying conditions $1-3$ we can always extend it to a group, each of whose members satisfies the conditions $1-3$. Thus in finding invariant tests we will always refer it with respect to a group **G** rather than the set S. Furthermore the induced transformation on Ω corresponding to G on χ also form a group $\overline{G}$.

It is wellknown that some simplification is introduced in the testing problem by characterizing the statistical tests as a function of the sufficient statistics. In the case of invariant tests it is also convenient to characterize the totality of invariant tests as a function of statistic which is popularly known as maximal invariant. A function $T(x)$ defined on χ is called a maximal invariant with respect to a group of transformations G on χ onto χ if (i) $T(x)=T(gx)$ for all $x \in \chi$ and $g \in G$ and (ii) if $T(x)=T(y)$ for $x, y \in \chi$ then there exists a $g \in G$ such that $y=gx$. Now, any invariant test $\phi(x)$ is a function of $T(x)$ follows from the fact that if $\phi(x)=h(T(x))$ for all x ($h\equiv$function) then

$$\phi(gx)=h(T(gx))=h(T(x))=\phi(x)$$

and conversely if ϕ is invariant and if $T(x)=T(y)$ then $y=gx$ for some $g \in G'$ and therefore $\phi(x)=\phi(y)$. Sufficiency provides a simplification to a statistical problem by reducing the dimension of the sample space to the dimension of the space of sufficient statistic but the process does not change the parametric space. On the other hand invariance by reducing the sample space to the dimension of the space of the maximal invariant shrinks also the parametric space. This follows from the fact that the distribution of the maximal invariant in the sample space with respect to the group G depends on the parameters only through the maximal invariant in Ω with respect to the induced group $\overline{G}$. This is seen to be the case by observing the following: Let $T(x)$ be a maximal invariant in χ with respect to $\mathbf{G}$ and $\nu(\theta)$ be a maximal invariant in Ω with respect to $\overline{G}$ and let $\nu(\theta_1)=\nu(\theta_2)$ for $\theta_1, \theta_2 \in \Omega$. Then $\theta_1=\overline{g}\,\theta_2$ for some $\overline{g} \in \overline{G}$. Now for any $B \in \beta$

$$\begin{aligned} P_{\theta_1}(T(x) \in B) &= P_{\theta_1}(T(gx) \in B) \\ &= P_{\bar{g}\theta_2}(T(gx) \in B) \\ &= P_{\theta_2}(T(x) \in B). \end{aligned}$$

3. Invariant Tests

Let $\underline{X}^{\alpha}=(X_{\alpha 1}, \ldots, X_{\alpha p})', \alpha=1, \ldots, N$ be a set of N observations from a normal population with mean $\underline{\alpha}$ and covariance matrix $\sum$ and let $\overline{\underline{X}}=\sum_1^N \underline{X}^{\alpha}/N$ and $S=\sum_1^N(\underline{X}^{\alpha}-\overline{\underline{X}})(\underline{X}^{\alpha}-\overline{\underline{X}})'$. We will assume throughout that $N>p$ so that S is positive definite with probability one. Write for any p-vector $\underline{b}=(b_1, \ldots, b_p)'$ $\underline{b}_{(1)}=(b_1, \ldots, b_q)'$, $\underline{b}_{(2)}=(b_1, \ldots, b_{p'})'$ and for any $p\times p$ matrix $\mathbf{C}$, $\mathbf{C}_{11}$ is the upper left-hand cornered $q\times q$ sub-matrix of $\mathbf{C}$ and $\mathbf{C}_{(22)}$ is the upper left-hand cornered $p'\times p'$ submatrix of $\mathbf{C}$.

Problem A. The problem of testing H_{10} against H_{11} remains invariant under the group G of $q\times q$ nonsingular matrices.

$$g=\begin{pmatrix} g_{11}, & 0 \\ g_{21}, & g_{22} \end{pmatrix}$$

operating as $\left(\underline{X}, \underline{\alpha}, \sum\right) \rightarrow (gX, g\alpha, g \Sigma g')$, where g_{11} is the $q \times q$ submatrix of g. A set of maximal invariant in the sample space with respect to **G** is (R_1, R_2) where

$$R_1 = N \underline{\bar{X}}'_{(1)} (S_{11} + N\underline{\bar{X}}_{(1)} \underline{\bar{X}}'_{(1)})^{-1} \underline{\bar{X}}_{(1)},$$

$$R_1 + R_2 = N \underline{\bar{X}}' (S + N\underline{\bar{X}}\,\underline{\bar{X}}')^{-1} \underline{\bar{X}}. \qquad \ldots(3\cdot1)$$

A corresponding maximal invariant in the parametric space Ω is (δ_1, δ_2) where

$$\delta_1 = N\underline{\alpha}'_{(1)} \left(\sum\nolimits_{11}\right)^{-1} \underline{\alpha}_{(1)},$$

$$\delta_1 + \delta_2 = N \underline{\alpha}' \sum\nolimits^{-1} \underline{\alpha}. \qquad \ldots(3\cdot2)$$

Obviously $R_i \geqslant 0$, $\delta_i \geqslant 0$ for $i=1, 2$. From Giri (1964) the joint probability density function of (R_1, R_2) is given by

$$f(r_1, r_2) = \exp\left(-\tfrac{1}{2}(\delta_1+\delta_2) - \tfrac{1}{2}\delta_2 r_1\right)$$

$$\sum_{j=0}^{\infty} \frac{\left(r_1 \frac{1}{2}\delta_1\right)^j \Gamma\left(\frac{N}{2}+j\right)\Gamma\left(\frac{q}{2}\right)}{j!\,\Gamma\left(\frac{N}{2}\right)\Gamma\left(\frac{q}{2}+j\right)}$$

$$\cdot \sum_{j=0}^{\infty} \frac{\left(r_2 \frac{1}{2}\delta_2\right)^j \Gamma\left(\frac{N-q}{2}+j\right)\Gamma\left(\frac{p-q}{2}\right)}{j!\,\Gamma\left(\frac{N-q}{2}\right)\Gamma\left(\frac{p-q}{2}+j\right)}$$

$$\cdot \frac{\Gamma\left(\frac{N}{2}\right) r_1^{\frac{q}{2}-1} \; r_2^{\frac{1}{2}(p-q)-1} \; (1-r_1-r_2)^{\frac{1}{2}(N-p)-1}}{\Gamma\left(\frac{q}{2}\right)\Gamma\left(\frac{p-q}{2}\right)\Gamma\left(\frac{N-p}{2}\right)} \qquad \ldots(3\cdot3)$$

It is easy to see that (see for example Giri (1965)) $H_{10}: \delta_1=\delta_2=0$, $H_{11}: \delta_2=0$. Hence the ratio of the density of (R_1, R_2) under H_{11} to their density under H_{10} is given by

$$\frac{dp_{H_{11}}(r_1, r_2)}{dp_{H_{10}}(r_1, r_2)} = \exp\left(-\frac{1}{2}\delta_1\right) \sum_{j=0}^{\infty} \frac{\left(r_1 \frac{1}{2}\delta_1\right)^j \Gamma\left(\frac{N}{2}+j\right)\Gamma\left(\frac{q}{2}\right)}{j!\,\Gamma\left(\frac{q}{2}+j\right)\Gamma\left(\frac{N}{2}\right)} \qquad \ldots(3\cdot4)$$

Hence the test which rejects H_{10} if $R_1 \geqslant$ constant (depending on the size of the test) is unformly most powerful invariant. From (3·3) it follows that under H_{10}, R_1

is distributed as beta with parameter $\left(\frac{q}{2}, \frac{N-q}{2}\right)$. From Giri (1964) it can be shown that it is also the likelihood ratio test for this problem.

Problem B. The problem of testing H_{20} against H_{21} remains invariant under the group of transformations G of problem A. Under H_{20} : $\delta_2=0$ and under H_{21} : $\delta_1>0, \delta_2>0$. Giri (1964) has shown that the likelihood ratio test of this problem is to reject H_{20} if

$$Z=\frac{1-R_1-R_2}{1-R_1} \leqslant \text{constant}$$

depending on the size of the test.

From (3·3) the joint probability density function of Z and R_1 under H_{20} is

$$\exp\left(-\tfrac{1}{2}\delta_1\right) \sum_{j=0}^{\infty} \frac{(\frac{1}{2}\sigma_1 r_1)^j\, r_1^{\frac{q}{2}+1} (1-q_1)^{\frac{N-q}{2}-1} Z^{\frac{N-p}{2}-1} (1-Z)^{\frac{p-q}{2}-1}}{j!\, B\left(\frac{N-q}{2}, \frac{q}{2}+j\right) B\left(\frac{N-p}{2}, \frac{p-q}{2}\right)} \qquad \ldots(3{\cdot}5)$$

Hence under H_{20}, Z is beta distributed with parameters $\frac{N-p}{2}$, $\frac{p-q}{2}$ and is independent of R_1. Further from (3·3) R_1 is sufficient for δ_1. Giri (1964) has shown that the distribution of R_1 is boundedly complete. Hence any invariant test $\phi(r_1, r_2)$ of level α for testing H_{20} against H_{21} has Neyman structure with respect to R_1 (Lehmann (1959) p. 134), *i.e.*

$$E_{H_{20}}(\phi(R_1, R_2) \mid R_1)=\alpha. \qquad \ldots(3{\cdot}6)$$

Moreover from (3·3)

$$\frac{dP_{H_{11}}(R_2 \mid R_1)}{dP_{H_{10}}(R_2 \mid R_1)}$$

$$=\exp\left(-\tfrac{1}{2}\delta_2(1-R_1)\right) \sum_{j=0}^{\infty} \frac{\left(R_2 \frac{1}{2}\delta_2\right)^j \Gamma\left(\frac{N-q}{2}+j\right)\Gamma\left(\frac{p-q}{2}\right)}{j!\,\Gamma\left(\frac{p-q}{2}+j\right)\Gamma\left(\frac{N-q}{2}\right)} \qquad \ldots(3{\cdot}7)$$

Now it is evident that the distribution of $R_2=(1-R_1)(1-Z)$ on each surface $R_1=r_1$ is independent of δ_1. Hence (see Lehmann (1959)) the likelihood ratio test for testing H_{20} against H_{21} is uniformly most powerful invariant similar.

Problem C. The problem of testing H_{30} against H_{31} remains invariant under the group G_1 of $p\times p$ nonsingular matrices

$$g_1=\begin{pmatrix} g_{11}, 0, 0 \\ g_{21}, g_{22}, 0 \\ g_{31}, g_{32}, g_{33} \end{pmatrix}$$

operating as $(\underline{X}, \underline{\alpha}, \sum) \longrightarrow (g\underline{X}, g\underline{\alpha}, g\sum g')$ where g_{11}, g_{22}, g_{33} are $q\times q$, $(p'-q)\times(p'-q)$ and $(p-p')\times(p-p')$ submatrices of g respectively. Giri (1965) has shown that a maximal invariant in the sample space is (R_1, R_2, R_3), $R_i \geqslant 0$ where

$$R_1 = N\underline{\overline{X}}'_{(1)}(S_{11}+N\underline{\overline{X}}'_{(1)}\underline{\overline{X}}_{(1)})^{-1}\underline{\overline{X}}_{(1)};$$

$$R_1+R_2 = N\underline{\overline{X}}'_{[2]}(S_{[22]}+N\underline{\overline{X}}_{[2]}\underline{\overline{X}}'_{[2]})^{-1}\underline{\overline{X}}_{[2]};$$

$$R_1+R_2+R_3 = N\underline{\overline{X}}'(S+N\underline{\overline{X}}\,\underline{\overline{X}}')^{-1}\underline{\overline{X}}.$$

A corresponding maximal invariant in Ω with respect to the induced group $\overline{G_1}$ is $(\delta_1, \delta_2, \delta_3)$, $\delta_i \geqslant 0$ for all i where

$$\delta_1 = N\underline{\alpha}'_{(1)}\sum\nolimits_{11}^{-1}\underline{\alpha}_{(1)};$$

$$\delta_1+\delta_2 = N\underline{\alpha}'_{[2]}\sum\nolimits_{[22]}^{-1}\underline{\alpha}_{[2]};$$

$$\delta_1+\delta_2+\delta_3 = N\underline{\alpha}'\sum\nolimits^{-1}\underline{\alpha}.$$

Under $H_{30}: \delta_3=0, \delta_2=0, \delta_1>0$ and under $H_{31}: \delta_3=0, \delta_2>0, \delta_1>0$.

It has been shown (Giri (1965)) that the likelihood ratio test for this problem is to reject H_{30} if

$$Z = \frac{1-R_1-R_2}{1-R_1} \leqslant C$$

where C is a constant, chosen to yield a test of size α and under H_{30}, Z is distributed as beta with parameters $\frac{N-p'}{2}, \frac{p'-q}{2}$ and is independent of R_1. Furthermore Giri (1965) has shown that the likelihood ratio test for this problem is uniformly most powerful invariant similar for testing H_{30} against H_{31}. The details in this case are omitted and the reader is referred to Giri (1965) for these.

4. Summary

Certain invariant tests for discriminant coefficients in classification problems are discussed in this paper.

REFERENCES

Fisher, R.A. (1938) : The statistical utilization of multiple measurements, Annals of Eugenics Vol. 8 376-386.

Giri, N. (1964) : On the likelihood ratio test of a normal multivariate testing problem. Annals of Mathematical Statistics, 35, 181-189.

Giri, N. (1965) : On the likelihood ratio test of a normal multivariate testing problem II. Annals of Mathematical Statistics, 36, 1061-1065.

Lehmann, E.L. (1959) : Testing Statistical hypothesis, Wiley, New York.