Sankhyā : The Indian Journal of Statistics
1977, Volume 39, Series B, Pt. 2, pp. 145–150.

PROBABILITY INEQUALITIES INVOLVING ESTIMATES OF PROBABILITY OF CORRECT CLASSIFICATION USING DEPENDENT SAMPLES

By SHIBDAS BANDYOPADHYAY
Indian Statistical Institute, Calcutta

SUMMARY. Periodic observations on the same unit are common in time related studies. Such observations are considered in two population classification problem where the two populations are two distinct points of time. Various classification rules are considered depending on the knowledge of the parameters in the distribution. Fisher's and Smith's estimators of the probability of correct classification are studied for each of these rules and probability inequalities involving them are established.

1. Introduction

Let ω be an experimental unit which is a random outcome from a population π. It is known that π is identical to one of two specified populations π_1 and π_2 where π_1 and π_2 denote the same population π^* at two distinct points of time t_1 and t_2 respectively. Let $X_t = X(\omega)$ be a $p \times 1$ vector of observations on the unit ω observed at time t. The problem is to classify ω, i.e., identify π with one of π_1 and π_2.

We shall assume that

$$X_t = m_t + U_t \qquad \ldots \quad (1.1)$$

where

$$U_t = vU_{t-1} + \epsilon_t, \quad t = 0, \ \pm 1, \pm 2, \ldots \qquad \ldots \quad (1.2)$$

and ϵ_t's are i.i.d. $N_p[0, \Lambda]$, $|v| < 1$. Then, the joint distribution of X_{t_1} and X_{t_2} may be written as (see Anderson, 1971)

$$\begin{pmatrix} X_{t_1} \\ X_{t_2} \end{pmatrix} \sim N_{2p}\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma & \rho\Sigma \\ \rho\Sigma & \Sigma \end{pmatrix}\right], \quad \rho^2 < 1. \qquad \ldots \quad (1.3)$$

We may note that when ρ in (1.3) is zero, our problem reduces to the standard classification problem.

When parameters in (1.3) are not completely known, information about them is obtained from a sample of N units $\omega_1, \omega_2, \ldots, \omega_N$ from π^*, each unit is observed at t_1 and t_2 with X_{t_i} as the X-observation on the unit ω_s

observed at time t_i, $(i = 1, 2)$, $(\alpha = 1, 2, \ldots, N)$. Then, $(X_{1\alpha}\ X_{2\alpha})$, $(\alpha = 1, 2, \ldots, N)$ are i.i.d. and the distribution is given by (1.3).

Let H_i denote the hypothesis that X_t is from π_i, i.e., $t = t_i$, $(i = 1, 2)$. Let $P_i(R)$ denote the probability of correct classification (PCC) of a classification rule R when H_i obtains, $(i = 1, 2)$. In this paper, we consider the rules R such that

$$P_1(R) = P_2(R) = P(R). \qquad \ldots \quad (1.4)$$

When parameters in (1.3) are known, the rule ψ^* which accepts H_1 if, and only if,

$$(\mu_2-\mu_1)'\Sigma^{-1}(2X_t-\mu_1-\mu_2) \leqslant 0 \qquad \ldots \quad (1.5)$$

is minimax Bayes rule (Anderson, 1958).

When the parameters in (1.3) are not completely known, we consider the rule ψ which accepts H_1 if, and only if,

$$(\overline{X}_2-\overline{X}_1)'B^{-1}(2X_t-\overline{X}_1-\overline{X}_2) \leqslant 0 \qquad \ldots \quad (1.6)$$

where
$$\overline{X}_i = (1/N) \sum_{\alpha=1}^{N} X_{i\alpha}, \quad (i = 1, 2)$$

and, B is a consistent and unbiased estimator of Σ. B depends on the knowledge of Σ and ρ as given below.

Case (a) : Σ is known. In this case,

$$B = \Sigma. \qquad \ldots \quad (1.7)$$

Case (b) : Σ is unknown but ρ is known. In this case,

$$B = (S_{11}-\rho S_{12}-\rho S_{21}+S_{22})/2(N-1)(1-\rho^2) \qquad \ldots \quad (1.8)$$

where
$$S_{ij} = \sum_{\alpha=1}^{N} (X_{i\alpha}-\overline{X}_i)(X_{j\alpha}-\overline{X}_j)', \; (i, j = 1, 2).$$

Case (c) : Both Σ and ρ are unknown. In this case,

$$B = (S_{11}+S_{22})/2(N-1). \qquad \ldots \quad (1.9)$$

Note that, μ_1 and μ_2 are treated to be unknown. In case they are known, the analysis is done replacing $\overline{X}_i$ by μ_i, $(i = 1, 2)$.

Let ψ_1, ψ_2 and ψ_3 denote the rule ψ given by (1.6) for B given respectively by (1.7), (1.8) and (1.9). The rules ψ_1 and ψ_2 are also likelihood ratio rules.

B in (1.8) is based on the maximum likelihood estimator of Σ using the sample only and so is B in (1.9) when $p = 1$. When $p > 1$ and under case (c), the maximum likelihood estimator of Σ is obtained iteratively (Kim, 1971) and we take B given by (1.9) as an approximation.

Following Fisher (1936) and Smith (1947) two estimators of PCC will be considered, namely $\Phi(\hat{\Delta}/2)$ and $C_1(\psi)$ respectively, where

$$\hat{\Delta} = [(\bar{X}_1-\bar{X}_2)'B^{-1}(\bar{X}_1-\bar{X}_2)]^{\frac{1}{2}} \quad \ldots \quad (1.10)$$

and $C_1(\psi)$ is the proportion of correctly classified observations from π_1 when the rule ψ is used.

Let $\hat{\Delta}_i^2$, $(i = 1, 2, 3)$ be $\hat{\Delta}^2$ given by (1.10) for B given by (1.7), (1.8) and (1.9) respectively.

In this study, following inequalities are established.

$$\mathcal{E}C_1(\psi_i) > P(\psi^*) > P(\psi_i), \quad (i = 1, 2) \quad \ldots \quad (1.11)$$

and

$$\mathcal{E}C_1(\psi_1) > \mathcal{E}\Phi(\hat{\Delta}_1/2) \quad \ldots \quad (1.12)$$

for all p. Moreover, when Δ is sufficiently small

$$\mathcal{E}\Phi(\hat{\Delta}_i/2) > P(\psi^*), \quad (i = 1, 2, 3) \quad \ldots \quad (1.13)$$

holds for all p. In the standard case, i.e., when $\rho = 0$, inequalities (1.11) and (1.12) have been established by Hills (1966) for $p = 1$ and generalised by Das Gupta (1974) for all p.

2. Inequalities

Since ψ^* and ψ_i, $(i = 1, 2, 3)$ satisfy (1.4) and ψ^* is unique Bayes rule when the parameters are known, following arguments of Hills (1966, (2), page 6) we have

$$P(\psi^*) > P(\psi_i), \quad (i = 1, 2, 3). \quad \ldots \quad (2.1)$$

For the rule ψ^*,

$$P(\psi^*) = P\left[(\mu_2-\mu_1)'\Sigma^{-1}\left(X_t-\frac{\mu_1+\mu_2}{2}\right) < 0 \mid X_t \sim N_p(\mu_1, \Sigma)\right]$$

$$= \Phi(\Delta/2),$$

where $\quad \Delta = [(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)]^{-\frac{1}{2}}$.

The exact expressions for $P(\psi_i)$, $(i = 1, 2, 3)$ are quite involved when $p > 1$. For $p = 1$, they are equal to $P(\psi)$, which is easily obtained as,

$$P(\psi) = \Phi(a)+\Phi(b)-2\Phi(a)\Phi(b), \quad \dots \quad (2.2)$$

where $a = (\Delta/2)[(1-\rho)/2N]^{-\frac{1}{2}}$

and $b = -(\Delta/2)[1+(1+\rho)/2N]^{-\frac{1}{2}}$.

Next we consider Smith's estimators. For any of the three choices of B,

$$\mathcal{E}C_1(\psi) = P[(\bar{X}_2-\bar{X}_1)'B^{-1}\{(X_{11}-\bar{X}_1)+(\tfrac{1}{2})(\bar{X}_1-\bar{X}_2)\} < 0].$$

When $p = 1$,

$$\begin{aligned}\mathcal{E}C_1(\psi) &= \mathcal{E}C_1(\psi_i), \quad (i = 1, 2, 3)\\ &= \Phi(a)+\Phi(c)-2\Phi_2(a, c; \rho_1), \quad \dots \quad (2.3)\end{aligned}$$

where

$$c = -(\Delta/2)[1-(1+\rho)/2N]^{-\frac{1}{2}},$$

$$\rho = c/a$$

and
$$\Phi_2(u, v; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{u}\int_{-\infty}^{v} \exp\left[-\frac{1}{2(1-\rho^2)}(u^2+v^2-2\rho uv)\right] du dv.$$

For $p = 1$ and $V_1 = X_{11}-\frac{1}{2}(\bar{X}_1+\bar{X}_2), V_2=\bar{X}_2-\bar{X}_1$, the conditional expectation of V_1 given $V_2 = v$ is $(-v/2)$. Thus $P[V_1 > 0, V_2 > 0] < P[V_1 < 0, V_2 > 0]$. So we get

$$\begin{aligned}\mathcal{E}C_1(\psi) &= P[V_1V_2 < 0]\\ &= P[V_1 < 0, \ V_2>0]+P[V_1 > 0, \ V_2 < 0]\\ &> P[V_1 > 0, \ V_2 > 0]+P[V_1 > 0, \ V_2 < 0]\\ &= P[V_1 > 0]\\ &= \Phi(-c). \quad \dots \quad (2.4)\end{aligned}$$

Thus, from (2.1) and (2.4) it follows that (1.11) holds for $(i = 1, 2, 3)$ when $p = 1$.

To prove (1.11) for $i = 1$ and $p > 1$, define v_0, v_1, and v_2 by

$$v_1 = \Sigma^{-\frac{1}{2}}(X_{11}-\bar{X}_1), \quad v_2 = \Sigma^{-\frac{1}{2}}(\bar{X}_2-\bar{X}_1)$$

and
$$v_0 = v_1'v_2(v_2'v_2)^{-\frac{1}{2}}.$$

Then

$$\begin{aligned}\mathcal{E}C_1(\psi_1) &= P[v_0 > -\tfrac{1}{2}(v_2'v_2)^{-\frac{1}{2}}]\\ &= \mathcal{E}\Phi[(1-1/N)^{-\frac{1}{2}}\hat{\Delta}_1/2] \quad \dots \quad (2.5)\end{aligned}$$

since conditional distribution of v_0 given v_2 is $N[0, (1-1/N)]$. Thus reducing the problem to univariate one by considering a linear compound of the original variables and following Das Gupta (1974, (4.12), (4.13), page 759) we get (1.11) for $i = 1$ and $p > 1$.

For $i = 2$ and $p > 1$, variates are easily transformed so that

$$Z_1 \equiv (1-1/N)^{-\frac{1}{2}}(X_{11}-\bar{X}_1)$$

and

$$(S_{11}-\rho S_{12}-\rho S_{21}+S_{22})/(1-\rho^2) = \sum_{i=1}^{2N-2} Z_i Z_i',$$

where Z_i's are i.i.d. $N(0, \Sigma)$ and distributed independently of $\bar{X}_1$ and $\bar{X}_2$. Then following Das Gupta (1974, (4.18), (4.19), page 761) we get,

$$\mathcal{E} C_1(\hat{\psi}_2) = \mathcal{E} F[(\hat{\Delta}_2/2)(1-1/N)^{-\frac{1}{2}}],$$

where F is the c.d.f. corresponding to the density given by

$$f(u) \propto (1-u^2)^{\frac{2N-3}{2}-1}, \quad |u| < 1.$$

Note that F is free from p. Using (2.4) and the above result and once again following arguments in Das Gupta (1974, (4.21), page 761) we get (1.11) for $i = 2$ for $p > 1$.

Finally we consider Fisher's estimators.

For $p = 1$

$$\mathcal{E}\Phi(\hat{\Delta}_1/2) = P\left[(\bar{X}_2-\bar{X}_1)\left(v+\frac{\bar{X}_1-\bar{X}_2}{2}\right) < 0\right],$$

where $v \sim N(0, \Sigma)$, independently of $\bar{X}_1$ and $\bar{X}_2$. Then

$$\mathcal{E}\Phi(\hat{\Delta}_1/2) = \Phi(a)+\Phi(d)-\Phi_2(a, d; \rho_2) \qquad \ldots \quad (2.6)$$

where

$$d = -(\Delta/2)[1+(1-\rho)/2N]^{-\frac{1}{2}}$$

$$\rho_2 = d/a.$$

From (2.6) and (2.2) we have

$$\mathcal{E}\Phi(\hat{\Delta}_1/2) > P(\psi)$$

if $\rho = 0$ since $b = d$ and $\rho_2 < 0$. It also follows from (2.5) that (1.12) holds for all p. Note that, except for a constant factor $\hat{\Delta}_i^2$, $(i = 1, 2, 3)$ is a chisquare, F or a mixture of F's for B given by (1.7), (1.8) and (1.9) respectively. Hence

$$\lim_{\Delta \to 0} \mathscr{E}\Phi(\hat{\Delta}_i/2) > \tfrac{1}{2} = \Phi(0), \quad (i = 1, 2, 3).$$

Hence (1.13) holds for all p, for sufficiently small Δ.

Acknowledgement. I thank the referee for substantial reorganisation of this paper and Professor S. Das Gupta for his advice.

REFERENCES

ANDERSON, T. W. (1958): *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

———, (1971): *The Statistical Analysis of Time Series*, Wiley, New York.

DAS GUPTA, S. (1974): Probability inequalities and error in classification. *Ann. Statist.*, **4**, 751-762.

FISHER, R. A. (1936): Use of multiple measurements in Taxonomic problem. *Ann. Eugen.*, **7**, 179-188.

HILLS, M. (1966): Allocation rules and the error rates. *J. Roy. Statist. Soc.*, Ser. B., **28**, 1-31.

KIM, D. Y. (1971): Statistical inference on constants of proportionality between covariance matrices. Ph.D. Thesis, Department of Statistics, Stanford University.

SMITH, C. A. B. (1947): Some examples on discrimination. *Ann. Eugen.*, **13**, 272-282.

Paper received : December, 1975.

Revised : June, 1977.