Theoretical Mathematics & Applications, vol. 6, no. 2, 2016, 153-179
ISSN: 1792- 9687 (print), 1792-9709 (online)
Scienpress Ltd, 2016

Derivation of Extended Optimal Classification Rule for Multivariate Binary Variables

I. Egbo[1], S.I. Onyeagu[2] and D.D. Ekezie[3]

Abstract

The statistical classification of N individuals into G mutually exclusive groups when the actual group membership is unknown is common in the social and behavioral sciences. The result of such classification methods often has important consequences. Let π_1 and π_2 be two distinct r-variate Bernoulli populations. Given an object Φ with observation measurement vector $(X_{\Phi 1}, \ldots X_{\Phi r})$, the optimal classification rule was developed to assign Φ to either π_1 or π_2. The performance of the optimal classification rule was compared with some existing procedures. The classification procedures are Full Multinomial, Predictive, Linear discriminant function. The expected cost of Misclassification was also derived. The four classification procedures for binary variables were discussed and evaluated at each of 118 configurations of the sampling experiments. The results obtained ranked the procedures as follows: Optimal, Linear discriminant, predictive and full multinomial rule. Further analysis revealed that increase in sample size and number of variables increase classification accuracy and lower the probability of Misclassification. Also apparent error rate obtained using resubstitution method is optimally biased and the bias decreases as the sample sizes increases.

[1] Department of Mathematics, Alvan Ikoku University of Education, Owerri, Nigeria.
E-mail: Egboike@gmail.com, Egboikechukwuj@yahoo.com
[2] Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria.

[3] Department of Statistics, Imo State University, Owerri, Nigeria.

Article Info: *Received:* September 22, 2015. *Revised:* October 24, 2015.
Published online : April 5, 2016.

Mathematics Subject classification: 62H30

Keywords: Classification rules; Optimal; Full multinomial; Predictive rule; Linear discriminant function

1 Introduction

Statistical classification of individuals into observed groups is a very common practice throughout the social, behavioural and physical sciences (Arabie, Hubert, & Desote, [1]; Keogh, [4]; and Zigler [10]. In education and psychology, examples abound in which researches attempt to find statistical models that can be used to classify individuals into one of several known categories, such as those based on disability status (e.g, Lillvist, [5], Mammarella and Cornoldi, [6], career choice (Russel, [9]) and student preferences regarding mode of instruction (Clayton, Blumberg & Auld, [10]), to name but a few. In all of these cases, the group membership is directly observable rather than latent in nature. It should be noted, however, that there is growing interest in a range of techniques designed specifically for use when group membership cannot be directly observed but rather is latent and thus must be inferred using a set of observed measures. The focus of the current study was on the case where group membership is observable and on a set of methods that can be used in that case. Generally speaking, these methods for the observed group context are not applicable to the situation where group membership is latent, and vice versa. Nevertheless, both scenarios are very applicable in the behavioural and social sciences and worthy of study. In this study, using the error of misclassification, we study the performances of four classification rules assuming the underlying populations have multivariate Bernoulli distribution.

Let π_1 and π_2 be two mutually exclusive parent populations available with infinite number of individual objects. Let there be r characteristics of interest with corresponding measurement variable x_1, x_2 ... x_r where $r \geq 1$. Let the measurement vector of an individual in π_1 be $x_1 = (x_{11}, x_{12} \ldots x_{1r})$ and in $x_2 = (x_{21}, x_{22} \ldots x_{2r})$. Supposing we find an object Φ with measurement vector

$x_\Phi = (x_{\Phi 1}, x_{\Phi 2} ... x_{\Phi r})$ outside π_1 and π_2 and which must belong to either π_1 or π_2. The problem is how to assign Φ to π_1 or π_2 such that the risk or expected cost or probability of error is a minimum. The measurement vector X_r can be discrete or continuous or a mixture of discrete and continuous variables. Our interest is about $\underline{X}$ whose arguments are 0 or 1. The case of continuous measurement vector $\underline{X}$ has been studied extensively and the case of mixed variables (discrete and continuous) is yet to be studied in detail. In this situation, the researcher can commit one of the following errors. An object from π_1 may be classified into π_2. Likewise, an object from π_2 may be misclassified into π_1. If misclassification occurs, a loss would be suffered. Let C(i/j) be the cost of misclassifying an object π_i into π_j. For the two population setting, we have that C(2/1) means cost of misclassifying an object into π_2 given that it is from π_1. C(1/2) is the cost of misclassifying an object into π_1 given that it is from π_2. The relative magnitude of the loss L(j,i) =C(i/j) depends on the case in question. For instance, a failure to detect an early cancer in a patient is costlier than stating that a patient has cancer and discovering otherwise. Therefore, if we are confronted with this kind of problem of classifying an object of an unknown origin with measurement vector, how do you choose the "BEST" rule so that the expected cost associated with misclassification will be minimum? The problem of this study is to find the "best" classification rule when $\pi_i, i = 1, 2$ is a multivariate Bernoulli population. "Best" here means the rule that minimizes the expected cost of misclassification (ECM), which is the risk of misclassification. Such a rule is referred to as the optimal classification rule (OCR). We want to find the OCR when X is multivariate Bernoulli. The three classification rules compared with the OCR in this study are (i) the full multinomial rule, (ii) the linear discriminant function rule, (iii) the predictive rule.

The problem of classification in the special case of binary variables is receiving extensive coverage in statistical literature. One reason for the rebirth of interest in the area is frequent use of classification in the social and behavioural sciences where data are often of finite discrete type. In studies involving questionnaire data, demographic variables (more often than not measured by a two, three or four point scale) are utilized to discriminate between two or more groups. In such cases, it is more natural to assume underlying multinomial structures and proceed with classification procedures based on such characterizations. Could the use of the four approaches of discriminant function appropriate in minimizing error rate of misclassification?

2 The Optimal Classification Rule

Independent Random Variables:

Let π_1 and π_2 be any two multivariate Bernoulli populations. Let $c(i/j)$ be the cost of misclassifying an item with measurement $\underline{x}$ from π_j into π_i and let q_j be the prior probability on π_i, where $i=1,2$ with $q_1+q_2=1$ and probability mass Function $f_i(x)$ in π_i where $i=1,2$. Suppose that we assign an item with measurement vector x to π_1 if it is in some region $R_1 \subseteq R^r$ and to π_2 if $\underline{x}$ is in some region $R_2 \subseteq R^r$ where $R^r = R_1 \cup R_2$ and $R_1 \cap R_2 = 0$. The expected cost of misclassification is given by:

$$ECM = c(2/1)q_1 \sum_{R_2} f(x/\pi_1) + c(1/2)q_2 \sum_{R_1} f(x/\pi_2) \tag{2.1}$$

where $\sum_{R_2} f(x/\pi_1) = \text{p}$ (classifying into π_2/π_1) = p(2/1).

The optimal rule is the one that partitions R^r such that

$ECM = \sum_{R_1} f(x/\pi_2)$ = p(classifying into π_1/π_2) = p(1/2) is a minimum.

$$ECM = c(2/1)q_1 \left[1 - \sum_{R_2} f(x/\pi_1)\right] + c(1/2)q_2 \sum_{R_1} f(x/\pi_2) \tag{2.2}$$

$$= c(2/1)q_1 + \sum_{R_1} \left[c(1/2)q_2 f(x/\pi_2) - c(2/1)q_1 f(x/\pi_1)\right] \tag{2.3}$$

ECM is minimized if the second term is minimized. ECM is minimized if R_1 is chosen such that

$$c(1/2)q_2 f(x/\pi_2) - c(2/1)q_1 f(x/\pi_1) \leq 0 \tag{2.4}$$

$$c(2/1)q_1 f(x/\pi_1) \geq c(2/1)q_2 f(x/\pi_2) \tag{2.5}$$

$$R_1 = \left[x / \frac{f(x/\pi_1)}{f(x/\pi_2)} \geq \frac{c(1/2)q_2}{c(2/1)q_1} \right], \quad f(x/\pi_2), c(2/1)\, q_1 \neq 0 \tag{2.6}$$

Therefore the optimal classification rule with respect to minimization of the expected cost of misclassification (ECM) is given by classify object with measurement x_0 into π_1 if

$$\frac{f_1}{f_2} \geq \frac{q_2 c(1/2)}{q_1 c(2/1)}, \qquad f_2, q_1 c(2/1) \neq 0 \tag{2.7}$$

Otherwise classify into π_2 .

Without loss of generality, we assume that $q_1 = q_2 = 1/2$ and c(1/2)=c(2/1). Then the minimization of the ECM becomes the minimization of the probability of misclassification, p(mc) under these assumptions, the optimal rule reduces to classifying an item with measurement x_0 into π_1 if

$$R_{opt} : \frac{f_1(x_0/\pi_1)}{f_2(x_0/\pi_2)} \geq 1, \qquad f_2(X_0/\pi_2) \neq 0 \tag{2.8}$$

Otherwise classify the item into π_2. Since x is multivariate Bernoulli with $P_{ij}>0$, i=1,2, j=1,2…r the optimal rule is: classify an item with response pattern $\underline{x}$ into π_1 if

$$\frac{\underset{j=1}{\overset{r}{\pi}} \left[p_{1j}^{x_j} (1-p_{1j})^{1-x_j} \right]}{\underset{j=1}{\overset{r}{\pi}} \left[p_{2j}^{x_j} (1-p_{2j})^{1-x_j} \right]} > 1, \quad p_{2j}^{x_j} \neq 0 \tag{2.9}$$

Otherwise, classify the item into π_2. This rule simplifies to:

Classify an item with response pattern $\underline{x}$ into π_1 if

$$\sum x_j In\left(\frac{p_{ij}}{q_{ij}} \cdot \frac{q_{2j}}{p_{2j}} \right) > \sum_{j=1}^{r} In \frac{q_{2j}}{q_{1j}}, \qquad q_{ij}, p_{2j} \neq 0 \tag{2.10}$$

Otherwise, classify into π_2 .

If the parameters are unknown, then they are estimated by their maximum likelihood estimators given by

$$\hat{p}_{ij} = \sum_{k=1}^{n} \frac{x_{ijk}}{n_i} = \frac{n_i(x_j)}{n_i} = \overline{x}_{ij} \tag{2.11}$$

Where $n_i(x_j) = \sum_{k=1}^{n_1} x_{ijk}$ is equal to the number of observation from π_i with jth variable. The rule for unknown parameters is: classify an item with response pattern $\underline{x}$ into π_1 if

$$\sum_{j=1}^{r} In \left(\frac{\hat{p}_{1j}}{\hat{q}_{1j}} \cdot \frac{\hat{q}_{2j}}{\hat{p}_{2j}} \right) x_j > \sum_{j=1}^{r} In \frac{\hat{q}_{2j}}{\hat{q}_{1j}} \tag{2.12}$$

otherwise classify the item into π_2.

2.1 The Optimal Rule for a case of two variables

Suppose we have only two independent Bernoulli variables, x_1, x_2. Then the rule becomes: classify an item with response pattern $\underline{x}$ into π_1 if:

$$R_{B_2} : In\left[\frac{p_{11}q_{21}}{q_{11}p_{21}}\right]x_1 + In\left[\frac{p_{12}q_{22}}{q_{12}p_{22}}\right]x_2 > In\frac{q_{21}}{q_{11}} + In\frac{q_{22}}{q_{12}} \tag{2.13}$$

Otherwise, classify the item into π_2. Written in another form the rule simplifies to: classify an item with response pattern $\underline{x}$ into π_1 if:

$$R_{B_2} : w_1x_1 + w_2x_2 > c \tag{2.14}$$

Otherwise, classify the item into π_2 where

$$w_1 = In\left[\frac{p_{11}}{1-p_{11}} - \frac{1-p_{21}}{p_{21}}\right] = In\frac{p_{11}}{1-p_{11}} - In\frac{p_{21}}{1-p_{21}} \tag{2.15}$$

$$w_2 = In\frac{p_{12}}{1-p_{12}} - In\frac{p_{22}}{1-p_{22}} \tag{2.16}$$

$$c = In\left[(1-p_{21})(1-p_{22})\right] - In\left[(1-p_{11})(1-p_{12})\right] \tag{2.17}$$

To find the distribution of z we note that

$$p[x_j = x_j / \pi_i] = \begin{cases} p_{ij}^{x_j}(1-p_{ij}), & i=1,2,\ j=1,2 \\ 0, & \text{otherwise,} \end{cases} \tag{2.18}$$

Since $z = \sum_{j=1}^{2} w_j x_j = w_1x_1 + w_2x_2$ (2.19)

The range of z is

$R_2 = \{0, w_1, w_2, w_1 + w_2\}$

$$p[z = 0/\pi_i] = p(x_1 = 0, x_2 = 0/\pi_i) = q_{i1}q_{i2} \quad (2.20)$$

$$p(z = w_1/\pi_i) = p(x_1 = 1, x_2 = 0/\pi_i) = p_{i1}q_{i2} \quad (2.21)$$

$$p(z = w_1 + w_2/\pi_i] = p(x_1 = 1, x_2 = 1/\pi_i) = p_{i1}p_{12}q_{i1}q_{i2} \quad (2.22)$$

if $z = 0$

$p(z/\pi_i) = p_{i1}q_{i2}$

if $z_1 = w_1$ (2.23)

$q_{i1}p_{i2}$ if $z = w_2$ (2.24)

$p_{i1}p_{i2}$ if $z = w_1 + w_2, i = 1, 2$ (2.25)

If $w_1 < w_2$ the distribution function of z is given by 0 if z =0

$q_{i1}q_{i2}$ if $0 \le z < w_1$ (2.26)

$$p(z/\pi_i) = q_{i1}q_{i2} + p_{i1}q_{i2} \text{ if } w_1 \le z < w_2$$

$$= q_{i1}q_{i2} + p_{i1}q_{i2} + p_{i1}q_{i2} \text{ if } w_2 \le z < w_1 + w_2 \quad (2.27)$$

1 if $w_1 + w_2 \le z$

2.2 Optimal rule for a case of three variables

Suppose we have three independent variables according to Onyeagu (2003), the rule is: classify an item with response pattern $\underline{x}$ into π_1 if:

$$R_{B_3} : In\left(\frac{p_{11}q_{21}}{q_{11}p_{21}}\right)x_1 + In\left(\frac{p_{12}}{q_{12}} \cdot \frac{q_{22}}{p_{22}}\right)x_2 + In\left(\frac{p_{13}}{q_{13}} \cdot \frac{q_{23}}{p_{23}}\right)x_3 > In\left(\frac{q_{21}q_{22}q_{23}}{q_{11}q_{12}q_{13}}\right) \quad (2.28)$$

otherwise, classify the item into π_2. Written in another form the rule simplifies to: classify an item with response pattern $\underline{x}$ into π_1 if:

$$R_{B_3} : w_1x_1 + w_2x_2 + w_3x_3 > c \quad (2.29)$$

otherwise classify the item into π_2.

$$w_1 = In\left(\frac{p_{11}}{q_{11}} \cdot \frac{q_{21}}{p_{21}}\right), w_2 = In\left(\frac{p_{12}}{q_{12}} \cdot \frac{q_{22}}{p_{22}}\right)$$

$$w_3 = In\left(\frac{p_{13}}{q_{13}} \cdot \frac{q_{23}}{p_{23}}\right), c = In\left(\frac{q_{21}q_{22}q_{23}}{q_{11}q_{12}q_{13}}\right) \quad (2.30)$$

Let $z = w_1x_1 + w_2x_2 + w_3x_3$ (2.31)

Then the range of z is

$$R_{B_3} = \{0, w_1, w_2, w_3, w_1 + w_2, w_1 + w_3, w_2 + w_3, w_1 + w_2 + w_3\}$$

$$p(z = 0/\pi_i) = p(x_1 = 0, x_2 = 0, x_3 = 0/\pi_i) = q_{i1}q_{i2}q_{i3} \quad (2.32)$$

$$p(z = w_1/\pi_i) = p(x_1 = 0, x_2 = 1, x_3 = 0/\pi_i) = q_{i1}p_{i2}q_{i3} \quad (2.33)$$

$$p(z = w_2/\pi_i) = p(x_1 = 0, x_2 = 1, x_3 = 1/\pi_i) = q_{i1}q_{i2}p_{i3} \quad (2.34)$$

$$p(z = w_3/\pi_i) = p(x_1 = 0, x_2 = 0, x_3 = 1/\pi_i) = q_{i1}q_{i2}p_{i3} \quad (2.35)$$

$$p(z = w_1 + w_2/\pi_i) = p(x_1 = 1, x_2 = 1, x_3 = 0/\pi_i) = p_{i1}p_{i2}q_{i3} \quad (2.36)$$

$$p(z = w_1 + w_3/\pi_i) = p(x_1 = 1, x_2 = 0, x_3 = 1/\pi_i)$$

$$= p_{i1}q_{i2}p_{i3}$$

(2.37)

$$p(z = w_2 + w_3/\pi_i) = p(x_1 = 0, x_2 = 1, x_3 = 1/\pi_i)$$

$$= q_{i1}p_{i2}p_{i3} \quad (2.38)$$

$$p(z = w_1 + w_2/\pi_i) = p(x_1 = 1, x_2 = 1, x_3 = 1/\pi_i) = p_{i1}p_{i2}p_{i3} \quad (2.39)$$

The probability mass function of z

$$\begin{aligned} &q_{i1}q_{i2}q_{i3} \text{ if } z = 0 \\ &p_{i1}q_{i2}q_{i3} \text{ if } z = w_1 \\ &q_{i1}p_{i2}q_{i3} \text{ if } z = w_2 \\ &q_{i1}q_{i2}p_{i3} \text{ if } z = w_3 \end{aligned} \quad (2.40)$$

$$\begin{aligned} p(z = z/\pi_i) &= p_{i1}p_{i2}q_{i3} \text{ if } z = w_1 + w_2 \\ &= p_{i1}q_{i2}p_{i3} \text{ if } z = w_1 + w_3 \\ &= q_{i1}p_{i2}p_{i3} \text{ if } z = w_2 + w_3 \\ &= p_{i1}p_{i2}p_{i3} \text{ if } z = w_1 + w_2 + w_3 \end{aligned} \quad (2.41)$$

If $w_1 < w_2 < w_3$ the distribution function of z is 0 if $z < 0$

$$q_{i1}q_{i2}q_{i3} \text{ if } 0 \le z < w_1$$

$$q_{i1}q_{i2}q_{i3} + p_{i1}q_{i2}q_{i3} \text{ if } w_1 \le z < w_2 \qquad (2.42)$$

$$q_{i1}q_{i2}q_{i3} + p_{i1}q_{i2}q_{i3} + q_{i1}p_{i2}q_{i3} \text{ if } w_2 \le z < w_3$$

$$q_{i1}q_{i2}q_{i3} + p_{i1}q_{i2}q_{i3} + q_{i1}p_{i2}q_{i3} + q_{i1}q_{i2}p_{i3} \text{ if } w_3 \le z < w_1 + w_2$$

$$p(z/\pi_i) = q_{i1}q_{i2}q_{i3} + p_{i1}q_{i2}q_{i3} + q_{i1}p_{i2}q_{i3} + q_{i1}q_{i2}p_{i3} + p_{i1}p_{i2}q_{i3}$$

$$\text{if } w_1 + w_2 \le z < w_1 + w_3 \qquad (2.43)$$

$$= q_{i1}q_{i2}q_{i3} + p_{i1}q_{i2}q_{i3} + q_{i1}p_{i2}q_{i3} + q_{i1}q_{i2}p_{i3} + p_{i1}p_{i2}q_{i3} + p_{i1}q_{i2}p_{i3}$$

$$\text{if } w_1 + w_3 \le z < w_2 + w_3$$

$$q_{i1}q_{i2}q_{i3} + p_{i1}q_{i2}q_{i3} + q_{i1}p_{i2}q_{i3} + q_{i1}q_{i2}p_{i3} + p_{i1}p_{i2}q_{i3} + p_{i1}q_{i2}p_{i3} + q_{i1}p_{i2}p_{i3}$$

$$\text{if } w_2 + w_1 \le z < w_1 + w_2 + w_3$$

$$1 \text{ if } w_1 + w_2 + w_3 \le z \qquad (2.44)$$

2.3 Optimal rules for a case of four variables

Suppose we have four independent Bernoulli variables, the rule is classify an item with response pattern $\underline{x}$ into π_1 if

$$R_{B_4} : In\left(\frac{p_{11}}{q_{11}} \cdot \frac{q_{21}}{p_{21}}\right)x_1 + In\left(\frac{p_{12}}{q_{12}} \cdot \frac{q_{22}}{p_{22}}\right)x_2 + In\left(\frac{p_{13}}{q_{13}} \cdot \frac{q_{23}}{p_{23}}\right)x_3$$
$$+ In\left(\frac{p_{14}}{q_{14}} \cdot \frac{q_{24}}{p_{24}}\right)x_4 > In\frac{q_{21}}{q_{11}} + In\frac{q_{22}}{q_{12}} + In\frac{q_{23}}{q_{13}} + In\frac{q_{24}}{q_{14}} \qquad (2.45)$$

otherwise, classify the item into π_2. Written in another form, the rule simplifies to:

classify an item with response pattern $\underline{x}$ into π_1 if:

$$R_{B4} = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 > c_1 + c_2 + c_3 + c_4$$

otherwise, classify the item into π_2. For the case of four variables, let

$$w_1 = In\left(\frac{p_{11}}{q_{11}} \cdot \frac{q_{21}}{p_{21}}\right), w_2 = In\left(\frac{p_{12}}{q_{12}} \cdot \frac{q_{22}}{p_{22}}\right), w_3 = In\left(\frac{p_{13}}{q_{13}} \cdot \frac{q_{23}}{p_{23}}\right),$$
$$w_4 = In\left(\frac{p_{14}}{q_{14}} \cdot \frac{q_{24}}{p_{24}}\right) \quad (2.46)$$

$$c_1 = In\frac{q_{21}}{q_{11}}, c_2 = In\frac{q_{22}}{q_{12}}, c_3 = In\frac{q_{23}}{q_{13}}, c_4 = In\frac{q_{24}}{q_{14}}$$

Then $z = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 = \sum_{j=1}^{4} w_j x_j$ (2.47)

The range of values of z is given by R_z

$$R_z = \{0, w_1, w_2, w_3, w_4, w_1 + w_2, w_1 + w_3, w_1 + w_4, w_2 + w_4, w_3 + w_4, w_1 + w_2 + w_3, \\ w_1 + w_2 + w_4, w_1 + w_3 + w_4, w_2 + w_3 + w_4, w_1 + w_2 + w_3 + w_4\} \quad (2.48)$$

$$p(z = 0 / \pi_i) = P(x_1 = 0, x_2 = 0, x_3 = 0, x_4 = 0 / \pi_i) = q_{i1}q_{i2}q_{i3}q_{i4} \quad (2.49)$$

$$p(z = w_1 / \pi_i) = P(x_1 = 1, x_2 = 0, x_3 = 0, x_4 = 0 / \pi_i)$$
$$= p_{i1}q_{i2}q_{i3}q_{i4} \quad (2.50)$$

$$p(z = w_2 / \pi_i) = P(x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 0 / \pi_i)$$
$$= q_{i1}p_{i2}q_{i3}q_{i4} \quad (2.51)$$

$$p(z = w_3 / \pi_i) = P(x_1 = 0, x_2 = 0, x_3 = 1, x_4 = 0 / \pi_i)$$
$$= q_{i1}q_{i2}p_{i3}p_{i4} \quad (2.52)$$

$$p(z = w_4 / \pi_i) = P(x_1 = 0, x_2 = 0, x_3 = 0, x_4 = 1 / \pi_i)$$
$$= q_{i1}q_{i2}q_{i3}p_{i4} \quad (2.53)$$

$$p(z = w_1 + w_2 / \pi_i) = P(x_1 = 1, x_2 = 1, x_3 = 0, x_4 = 0 / \pi_i)$$
$$= p_{i1}p_{i2}q_{i3}q_{i4} \quad (2.54)$$

$$p(z = w_1 + w_3 / \pi_i) = P(x_1 = 1, x_2 = 0, x_3 = 1, x_4 = 0 / \pi_i)$$
$$= p_{i1}q_{i2}p_{i3}q_{i4} \quad (2.55)$$

$$p(z = w_1 + w_4 / \pi_i) = P(x_1 = 1, x_2 = 0, x_3 = 0, x_4 = 1 / \pi_i)$$
$$= p_{i1}q_{i2}q_{i3}p_{i4} \quad (2.56)$$

$$p(z = w_2 + w_3 / \pi_i) = P(x_1 = 0, x_2 = 1, x_3 = 1, x_4 = 0 / \pi_i)$$

$$= q_{i1} p_{i2} p_{i3} q_{i4} \quad (2.57)$$

$$p(z = w_2 + w_4 / \pi_i) = P(x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 1 / \pi_i)$$

$$= q_{i1} p_{i2} q_{i3} p_{i4}$$

(2.58)

$$p(z = w_3 + w_4 / \pi_i) = P(x_1 = 0, x_2 = 0, x_3 = 1, x_4 = 1 / \pi_i)$$

$$= q_{i1} q_{i2} p_{i3} p_{i4} \quad (2.59)$$

$$p(z = w_1 + w_2 + w_3 / \pi_i) = P(x_1 = 1, x_2 = 1, x_3 = 1, x_4 = 0 / \pi_i)$$

$$= p_{i1} p_{i2} p_{i3} q_{i4} \quad (2.60)$$

$$p(z = w_1 + w_2 + w_4 / \pi_i) = P(x_1 = 1, x_2 = 1, x_3 = 0, x_4 = 1 / \pi_i)$$

$$= p_{i1} p_{i2} q_{i3} p_{i4} \quad (2.61)$$

$$p(z = w_1 + w_3 + w_4 / \pi_i) = P(x_1 = 1, x_2 = 0, x_3 = 1, x_4 = 1 / \pi_i)$$

$$= p_{i1} p_{i2} q_{i3} p_{i4} \quad (2.62)$$

$$p(z = w_2 + w_3 + w_4 / \pi_i) = P(x_1 = 0, x_2 = 1, x_3 = 1, x_4 = 1 / \pi_i)$$

$$= q_{i1} p_{i2} p_{i3} p_{i4} \quad (2.63)$$

The probability mass function of z is given by $p(z = z / \pi_i)$

$$q_{i1} q_{i2} q_{i3} q_{i4} \text{ if } z = 0$$

$$p_{i1} q_{i2} q_{i3} q_{i4} \text{ if } z = w_1$$

$$q_{i1} p_{i2} q_{i3} q_{i4} \text{ if } z = w_2$$

$$q_{i1} q_{i2} p_{i3} q_{i4} \text{ if } z = w_3 \quad (2.64)$$

$$q_{i1} q_{i2} q_{i3} p_{i4} \text{ if } z = w_4$$

$$p_{i1} p_{i2} q_{i3} q_{i4} \text{ if } z = w_1 + w_2$$

$$p_{i1} q_{i2} p_{i3} q_{i4} \text{ if } z = w_1 + w_3$$

$$p_{i1} q_{i2} q_{i3} p_{i4} \text{ if } z = w_1 + w_4$$

$$p(z = z / \pi_i)_{i=1,2} = q_{i1} p_{i2} p_{i3} q_{i4} \text{ if } z = w_2 + w_3$$

$$
\begin{aligned}
& q_{i1}p_{i2}q_{i3}p_{i4} \text{ if } z = w_2 + w_4 \\
& q_{i1}q_{i2}p_{i3}p_{i4} \text{ if } z = w_3 + w_4 \\
& p_{i1}p_{i2}p_{i3}q_{i4} \text{ if } z = w_1 + w_2 + w_3 \\
& p_{i1}p_{i2}q_{i3}p_{i4} \text{ if } z = w_1 + w_2 + w_4 \\
& p_{i1}q_{i2}p_{i3}p_{i4} \text{ if } z = w_1 + w_3 + w_4 \\
& q_{i1}p_{i2}p_{i3}p_{i4} \text{ if } z = w_2 + w_3 + w_4 \\
& p_{i1}p_{i2}p_{i3}p_{i4} \text{ if } z = w_1 + w_2 + w_3 + w_4
\end{aligned} \tag{2.65}
$$

2.4 Evaluating the probability of misclassification for the optimal rule R_{opt}

The optimal classification rule R_{opt} for $\underline{x} = (x_1, x_2 \ldots x_r)$ which is distributed multivariate Bernoulli is: classify an item with response pattern $\underline{x}$ into π_1 if

$$R_{opt} : \sum_{j=1}^{r} x_j In\left(\frac{p_{1j}}{q_{1j}} \cdot \frac{q_{2j}}{p_{2j}}\right) > \sum_{j=1}^{r} In\frac{q_{2j}}{q_{1j}} \tag{2.66}$$

Otherwise classify into π_2

We can obtain the probability of misclassification for two cases

Case I Known parameters

(a) General case where $p_1 = (p_{i1}, p_{i2} \ldots p_{ir})$

(b) Special case where $p_i = (p_i, p_i \ldots p_i)$ with the assumption $p_1 < p_2$

(c) Special case (b) with additional assumption that $p_1 = \theta p_2, 0 < \theta < 1$

For case (1a) the optimal classification rule R_{opt} for $\underline{x} = (x_1, x_2 \ldots x_r)$ which is distributed multivariate Bernoulli is:

Classify an item with response pattern *x* if

$$R_{opt} : \sum x_j In\left(\frac{p_{1j}}{q_{1j}} \cdot \frac{q_{2j}}{p_{2j}}\right) > \sum_{j=1}^{r} In\frac{q_{2j}}{q_{1j}} \tag{2.67}$$

Otherwise classify into π_2

Case 1b: Special case where $p_i = p(p_i,...p_i)$ with the assumption that $p_1 < p_2$, the optimal classification rule R_{opt} for the r-variate Bernoulli models becomes: classify an item with response pattern $\underline{x}$ into π_1 if otherwise classify into π_2. The probability of misclassification using the special case of R_{opt} is

$$R_{opt} : \sum_{j=1}^{r} x_j \leq \frac{rIn\left(\frac{q_2}{q_1}\right)}{In\left(\frac{p_1}{q_1} \cdot \frac{q_2}{p_2}\right)} \tag{2.68}$$

$$p(2/1) = p\left[\sum_{j=1}^{r} x_j > \frac{rIn\frac{q_2}{q_1}}{In\left(\frac{p_1}{p_2} \cdot \frac{q_2}{q_1}\right)} \middle| \pi_1 \right] = 1 - B_{(r,p_1)}\left(\frac{rIn\frac{q_2}{q_1}}{In\left(\frac{p_1}{p_2} \cdot \frac{q_2}{q_1}\right)}\right) \tag{2.69}$$

$$B_{r,p}(x) = \sum_{y=0}^{r} \binom{x}{y} p^y (1-p)^{r-y} \tag{2.70}$$

$$p(1/2) = p\left[\sum_{j=1}^{r} x_j < \frac{rIn\frac{q_2}{q_1}}{In\left(\frac{p_1}{p_2} \cdot \frac{q_2}{q_1}\right)} \middle| \pi_2 \right] = B_{(r,p_2)}\left(\frac{rIn\frac{q_2}{q_1}}{In\left(\frac{p_1}{p_2} \cdot \frac{q_2}{q_1}\right)}\right) \tag{2.71}$$

$$p(mc) = \frac{1}{2}\left[1 + B_{(r,p_2)}\left(\frac{rIn\frac{q_2}{q_1}}{In\left(\frac{p_1}{p_2} \cdot \frac{q_2}{q_1}\right)}\right) - B_{(r,p_2)}\left(\frac{rIn\frac{q_2}{q_1}}{In\left(\frac{p_1}{p_2} \cdot \frac{q_2}{q_1}\right)}\right)\right] \tag{2.72}$$

Case 1c: Special case (1b) with additional assumption that $p_1 = \theta p_2$ and $q_1 = 1 - p_1 = 1 - \theta p_2$ and $q_2 = 1 - p_2$. The optimal classification rule R_{opt} for $\underline{x} = (x_1, x_2 .. x_r)$ distributed multivariate Bernoulli is: classify the item with response pattern x into π_1 if

$$R_{opt} : \sum_{j=1}^{r} x_j > \left[\frac{rIn\left(\frac{1-p_2}{1-\theta p_2}\right)}{In\theta\left(\frac{1-p_2}{1-\theta P_2}\right)}\right] \tag{2.73}$$

and to π_2 otherwise.

The probability of misclassification using the special case of R_{opt} when $p_1 = \theta p_2$ is

$$p(2/1) = 1 - B_{(r,\theta p_2)} \frac{rIn\left(\frac{1-p_2}{1-\theta p_2}\right)}{In\theta\left(\frac{1-p_2}{1-\theta p_2}\right)} \tag{2.74}$$

$$p(1/2) = B_{(r,p_2)} \frac{rIn\left(\frac{1-p_2}{1-\theta p_2}\right)}{In\theta\left(\frac{1-p_2}{1-\theta p_2}\right)}$$

$$p(mc) = \frac{1}{2}\left[1 + B_{(r,p_2)}\left(\frac{rIn\left(\frac{1-p_2}{1-\theta p_2}\right)}{In\theta\left(\frac{1-p_2}{1-\theta p_2}\right)}\right)\right] - B_{r,\theta p_2}\left[\frac{rIn\left(\frac{1-p_2}{1-\theta p_2}\right)}{In\theta\left(\frac{1-p_2}{1-\theta p_2}\right)}\right] \tag{2.75}$$

For the fixed values of r and different values of p_1 and p_2

Case 2:Unknown parameters

(a) General case $p_i = (p_{i1}, p_{i2} \ldots p_{ik})$

In order to estimate p_1 and p_2 we take training samples of size n_1 and n_2 from π_1 and π_2 respectively. In π_1 we have the sample

$$\begin{aligned} x_{11} &= (x_{111}, x_{121}, x_{131}, \ldots x_{1k1}, \ldots x_{1r1}) \\ x_{12} &= (x_{112}, x_{122}, x_{132}, \ldots x_{1k2}, \ldots x_{1r2}) \\ &. \\ &. \\ &. \\ x_{1n1} &= (x_{11n1}, x_{12n1}, x_{13n1}, \ldots x_{1kn1}, \ldots x_{1rn1}) \end{aligned} \tag{2.76}$$

The maximum likelihood estimate of p_1 is

$$\hat{p}_{1k} = \sum_{j=1}^{n_1} \frac{x_{1kj}}{n_1} \tag{2.77}$$

Similarly the maximum likelihood of estimate of p_2 is

$$\hat{p}_{2k} = \sum_{j=1}^{n_2} \frac{x_{2kj}}{n_2} \tag{2.78}$$

We plug in this estimate into the rule for the general case in 1(a) to have the following classification rule: classify an item with response pattern *x* into π_1 if

$$R_{Br} : \sum_{j=1}^{r} x_j In\left(\frac{\hat{p}_{ij}}{\hat{q}_{ij}} \cdot \frac{\hat{q}_{2j}}{\hat{p}_{2j}}\right) > rIn\frac{\hat{q}_{2j}}{\hat{q}_{ij}} \tag{2.79}$$

otherwise classify into π_2

(b) Special case of 1b where $p_i = (p_i, p_i ... p_i)$ with the assumption that $p_{1i} < p_{2i}$

In this special case

$$\hat{p}_1 = \sum_{j=1}^{n_1} \frac{x_{11j}}{n_1} = \sum_{j=1}^{n_1} \frac{x_{12j}}{n_1} ... \sum_{j=1}^{n_1} \frac{x_{1kj}}{n_1} = ... = \sum_{j=1}^{n_1} \frac{x_{1rj}}{n_1} \quad (2.80)$$

$\sum_{j=1}^{r} x_{ij}$ is distributed $B(r, p_i)$

$\sum_{k=1}^{n_1} \sum_{j=1}^{r} x_{1jk}$ is distributed $B(rn_1, p_1)$

The maximum likelihood estimate of p_1 is

$$\hat{p}_1 = \frac{\sum_{k=1}^{r} \sum_{j=1}^{n_1} x_{1jk}}{rn_1} \quad (2.81)$$

Likewise, the maximum likelihood estimate of p_2 is

$$\hat{p}_2 = \frac{\sum_{k=1}^{r} \sum_{j=1}^{n_2} x_{2jk}}{rn_2} \quad (2.82)$$

We plug in these two estimates into the equation for the special case (1b) to have the following classification rule: classify the item with response pattern x into π_1 if

$$\sum_{j=1}^{r} x_j \leq \frac{r In\left(\frac{\hat{q}_2}{\hat{q}_1}\right)}{In\left(\frac{\hat{p}_1}{\hat{q}_1} \cdot \frac{\hat{q}_2}{\hat{p}_2}\right)} \quad (2.83)$$

Otherwise classify into π_2

The probability of misclassification is given by

$$\hat{p}(mc) = \frac{1}{2}\left[1 + B_{(r,p_2)}\left(\frac{r In \frac{\hat{q}_2}{q_1}}{In\left(\frac{\hat{p}_1}{\hat{p}_2} \cdot \frac{\hat{q}_2}{\hat{q}_1}\right)}\right) - B_{(r,\hat{p}_1)} \frac{r In\left(\frac{\hat{q}_2}{\hat{q}_1}\right)}{In\left(\frac{\hat{p}_1}{\hat{p}_2} \cdot \frac{\hat{q}_2}{\hat{q}_1}\right)}\right] \quad (2.84)$$

$$\hat{p}(mc) = \frac{1}{2}\left[1 + B(r, \hat{p}_2, \lambda) - B(r, \hat{p}_1, \lambda)\right]$$

Where

$$\lambda = \frac{rIn\left(\frac{\hat{q}_2}{\hat{q}_1}\right)}{In\left(\frac{\hat{p}_1}{\hat{p}_2}\cdot\frac{\hat{q}_2}{\hat{q}_1}\right)}$$

$$B(k,\alpha,x) = \sum_{y=0}^{k}\binom{k}{y}\alpha^{y}(1-\alpha)^{k-y} \tag{2.85}$$

(c) Special case of 2b with $p_1 = \theta p_2, p_1 < p_2, 0 < \theta < 1$ we take training samples of size n_2 from π_2 and estimate p_2 by

$$\hat{p}_2 = \sum_{k=1}^{n}\sum_{j=1}^{r}\frac{x_{2jk}}{rn_2} \tag{2.86}$$

For a fixed value of $\theta, \hat{p}_1 = \theta\hat{p}_2$

The classification rule is: classify the item with response pattern x into π_1 if

$$R_{B_r} : \sum_{j=1}^{r} x_j > \frac{rIn\left[\frac{1-\hat{p}_2}{1-\theta\hat{p}_2}\right]}{In\theta\left[\frac{1-\hat{p}_2}{1-\theta\hat{p}_2}\right]} \tag{2.87}$$

otherwise classify into π_2.

The probability of misclassification is given by

$$\hat{p}(mc) = \frac{1}{2}\left[1 + B_{(r,p_2)}\frac{rIn\left[\frac{(1-\hat{p}_2)}{(1-\theta\hat{p}_2)}\right]}{In\theta\left[\frac{(1-\hat{p}_2)}{(1-\theta\hat{p}_2)}\right]}\right] - B_{r,\theta p_2}\left[\frac{rIn\left[\frac{(1-\hat{p}_2)}{(1-\theta\hat{p}_2)}\right]}{In\theta\left[\frac{(1-\hat{p}_2)}{(1-\theta\hat{p}_2)}\right]}\right] \tag{2.88}$$

$$\hat{p}(mc) = \frac{1}{2}\left[1 + B(r, p_2, \lambda) - B(r, \theta\hat{p}_2, \lambda)\right]$$

$$\lambda = \frac{rIn\left(\frac{1-\hat{p}_2)}{1-\theta\hat{p}_2}\right)}{In\theta\left(\frac{1-\hat{p}_2}{1-\theta\hat{p}_2}\right)} \tag{2.89}$$

If $p_1 = (p_1, p_1, p_1)$ and $p_2 = (p_2, p_2, p_2)$

$$f_1(x_1, x_2, x_3) = \pi_{i=1}^{3} p_{1i}^{x_1} q_{1i}^{1-x_1} = \pi_{i=1}^{3} p_1^{x_1} q_1^{1-x_1} = p_1^t q_1^{3-t} = \delta_1^t q_1^3$$

where $\delta_1 = \frac{p_1}{q_1}$ and $t = \sum_{i=1}^{3} x_i$ (2.90)

Similarly,

$$f_2(x_1,x_2,x_3)=\delta_2^t q_2^3 \text{ where } \delta_2=\frac{p_2}{q_2}$$

$$\frac{f_1(x_1,x_2,x_3)}{f_2(x_1,x_2,x_3)}=\left(\frac{\delta_2}{\delta_1}\right)^t\left(\frac{q_2}{q_1}\right)^3=\delta^t q^3 \tag{2.91}$$

where $\delta=\frac{\delta_2}{\delta_1}, q=\frac{q_2}{q_1}$

The classification rule is: classify the item with response pattern x into π_1 if $\delta^t q^3<1$ otherwise classify into π_2 .

For our case $p_1<p_2$ so we have $q_1>q_2$ and both imply that

$$\frac{p_2}{p_1}>1 \text{ and } \frac{q_1}{q_2}>1. \tag{2.92}$$

Therefore $\delta=\frac{\delta_2}{\delta_1}=\frac{p_2}{q_2}\cdot\frac{q_1}{p_1}=\frac{p_2}{p_1}\cdot\frac{q_1}{q_2}>1$ which implies that $In\,\delta>0$

Therefore $\delta^t q^3<1$ if and only if

$$t<\frac{-3In\,q}{In\delta}$$

The rule is classify the item with response pattern x into π_1

$$t<\frac{-3In\,q}{In\,\delta} \text{ or } \sum_{i=1}^{3}x_i<\frac{-3In\,q}{In\,\delta}$$

otherwise classify into π_2 .

Let $y=\sum_{i=1}^{3}x_i$ then $R_y=[0,1,2,3]$

$$P(Y=y/\pi_1)=\begin{cases} q_i^3 if`y=0 \\ 3p_iq_i^2 if`y=1 \\ 3p_i^2q_i if`y=2 \\ p_i^3 if`y=3 \\ 0`otherwise \end{cases} \tag{2.93}$$

For $p_1=(.3,.3,.3)$

$$P(Y=y/\pi_1)=\begin{cases} q_1^3 = 0.343 \text{ if } y=0 \\ 3p_1q_1^2 = 0.441 \text{ if } y=1 \\ 3p_1^2q_1 = 0.189 \text{ if } y=2 \\ p_1^3 = 0.027 \text{ if } y=3 \\ 0 = 0 \text{ otherwise} \end{cases} \tag{2.94}$$

For $p_2 = (.4,.4,.4)$

$$P(Y=y/\pi_2)=\begin{cases} q_2^3 = 0.216 \text{ if } y=0 \\ 3p_2q_2^2 = 0.432 \text{ if } y=1 \\ 3p_2^2q_2 = 0.288 \text{ if } y=2 \\ p_2^3 = 0.064 \text{ if } y=3 \\ 0 = 0 \text{ otherwise} \end{cases} \tag{2.95}$$

$$\delta = \frac{p_2}{p_1}\cdot\frac{q_1}{q_2} = \frac{0.4}{0.3}\,x\,\frac{0.7}{0.6} = \frac{28}{18} = \frac{14}{9}$$

$$In\delta = 0.441832$$

$$q = q_2/q_1 = .6/.7 = \frac{6}{7}$$

Substituting we have $\dfrac{-3In\,q}{In\delta} = 1.04666$

The classification rule is: classify the item with response pattern *x* into π_1 if

$$y = \sum_{i=1}^{3} x_i = 0,1$$

Otherwise classify into π_2

The probability of misclassification for this rule is:

$$P(mc) = \frac{1}{2}\left[p[Y=0,1/\pi_2] + p[Y=2,3/\pi_1\right]$$

$$= \frac{1}{2}(0.216 + 0.432 + 0.189 + 0.027)$$

$$= \frac{1}{2}(0.864) = 0.4320$$

The general classification rule is:

classify into π_1 if $y = 0,1,..n+r$ (2.96)

classify into π_2 if $y = n+r+1,...3n$

The probability of misclassification is given by

$$P(mc) = \frac{1}{2}\left[p[Y \geq n+r/\pi_2] + p[Y \leq n+r+1/\pi_1]\right] \quad (2.97)$$

Where Y is Binomial $(3n, p_i)$ $i = 1,2$

$$P(mc) = \frac{1}{2}\left[\sum_{y=0}^{n+r}\binom{3n}{y}p_2^y q_2^{3n-y} + \sum_{y=n+r+1}^{3}n\binom{3n}{y}p_1^y q_1^{3n-y}\right]$$

$$= \frac{1}{2} + \frac{1}{2}\sum_{y=0}^{n+r}\binom{3n}{y}\left(p_2^y q_2^{3n-y} - p_1^y q_1^{3n-y}\right)$$

$$= \frac{1}{2} + \frac{1}{2}\sum_{y=0}^{n+r}\binom{3n}{y}\left[\left(\frac{p_2}{q_2}\right)^y q_2^{3n} - \left(\frac{p_1}{q_1}\right)^y q_1^{3n}\right]$$

If $p_1 = (p_1 p_1 p_1 p_1)$ and $p_2 = (p_2 p_2 p_2 p_2)$ the probability of misclassification is

$$P(mc) = \frac{1}{2} + \frac{1}{2}\sum_{y=0}^{n+r}\binom{4n}{y}\left[\left(\frac{p_2}{q_2}\right)^y q_2^{4n} - \left(\frac{p_1}{q_1}\right)^y q_1^{4n}\right] \quad (2.98)$$

For r number of variables

$$p(mc) = \frac{1}{2} + \frac{1}{2}\sum_{y=0}^{n+r}\binom{rn}{y}\left[\left(\frac{p_2}{q_2}\right)^y q_2^{rn} - \left(\frac{p_1}{q_1}\right)^y q_1^{rn}\right]$$

If $p_1 = (.3,.3,.3)$ and $p_2 = (.4,.4,.4)$

$$p(mc) = \frac{1}{2}\left[\sum_{y=0}^{n+r}\binom{3n}{y}p_2^y q_2^{3n-y} + \sum_{y=n+r+1}^{3n}\binom{3n}{y}p_1^y q_1^{3n-y}\right]$$

$$= \frac{1}{2}\left[\sum_{y=0}^{1}\binom{3}{y}p_2^y q_2^{3-y} + \sum_{y=2}^{3}\binom{3}{y}p_1^y q_1^{3-y}\right] \quad (2.99)$$

$$\frac{1}{2}\left[\binom{3}{0}(0.4)^0(0.6)^3 + \binom{3}{1}(0.4)^1(0.6)^2 + \binom{3}{2}(0.3)^2(0.7)^1 + \binom{3}{3}(0.3)^3(0.7)^0\right]$$

0.432

The following table consists of 27 population pairs with their classification rules and optimum probabilities of misclassification computed using the formula above

Table 1: Population pairs with their classification rules

S/N	π_1 p_1	π_2 p_2	Classification Rule	p(mc)
1	.3,.3,.3	.4,.4,.4	classify into π_1 if y=0,1 Classify into π_2 if y=2,3	0.432
2	.3,.3,.3	.5,.5,.5	"	0.358
3	.3,.3,.3	.6,.6,.6	"	0.284
4	.3,.3,.3	.7,.7,.7	"	0.216
5	.4,.4,.4	.6,.6,.6	"	0.352
6	.4,.4,.4	.7,.7,.7	"	0.284
7	.5,.5,.5	.7,.7,.7	"	0.358
8	.6,.6,.6	.7,.7,.7	"	0.432
9	.3,.3,.3,.3	.4,.4,.4,.4	"	0.4117
10	.3,.3,.3,.3	.5,.5,.5,.5	"	0.3304
11	.3,.3,.3,.3	.6,.6,.6,.6	"	0.2637
12	.3,.3,.3,.3	.7,.7,.7,.7	classify into π_1 if y=0,1,2 classify into π_2 if y=3,4	0.216
13	.4,.4,.4,.4	.5,.5,.5,.5	classify into π_1 if y=0,1 classify into π_2 if y=2,3,4	0.4186
14	.4,.4,.4,.4	.6,.6,.6,.6	classify into π_1 if y=0,1,otherwise into π_2 if y=3,4	0.3520
15	.4,.4,.4,.4	.7,.7,.7,.7	"	0.2637
16	.5,.5,.5,.5	.6,.6,.6,.6	"	0.4187
17	.5,.5,.5,.5	.7,.7,.7,.7	"	0.3304
18	.6,.6,.6,.6	.7,.7,.7,.7	"	0.41175
19	.3,.3,.3,.3,.3	.4, . 4,.4,.4,.4	classify into π_1 if y=0,1 otherwise into π_2 if y=2,3,4,5	0.40437
20	.3,.3,.3,.3,.3	.5,.5,.5,.5,.5	"	0.32964
21	.3,.3,.3,.3,.3	.6,.6,.6,.6,.6	classify into π_1 if y=0,1,2 otherwise into π_2 if y=3,4,5	0.24026
22	.3,.3,.3,.3,.3	.7,.7,.7,.7,.7	"	0.16308
23	.4,.4,.4,.4,.4	.5,.5,.5,.5,.5	"	0.408725
24	.4,.4,.4,.4,.4	.6,.6,.6,.6,.6	"	0.31744
25	.4,.4,.4,.4,.4	.7,.7,.7,.7,.7	"	0.24026
26	.5,.5,.5,.5,.5	.6,.6,.6,.6,.6	"	0.40872
27	.5,.5,.5,.5,.5	.7,.7,.7,.7,.7	classify into π_1 if y=0,1,2,3 otherwise into π_2 if y= 4,5	0.32964
28	.6,.6,.6,.6,.6	.7,.7,.7,.7,.7	"	0.40437

3. Other Classification Procedures

3.1 The Full Multinomial Rule

Suppose we have a d-dimensional random vector $x^1 = (x_1, \ldots x_d)$ where each $x_j, j = 1, \ldots d$ assumes one of the two distinct values: 0 or 1. The sample space then has a multinomial distribution consisting of the 2^d possible states.

Given two disjoint populations, π_1 and π_2 with priori probabilities p_1 and p_2, the density is

$$f(x) = p_1 f_1(x) + p_2 f_2(x) \tag{3.1}$$

The two group problem attempts to find an optimal classification rule that assigns a new observation x to π_1 if

$$f_1(x) / f_2(x) > p_2 / p_1 \tag{3.2}$$

When x has only two states, it will be a binomial random variable with $n_i(x)$ observation from π_i and expected value $np_i f_i(x), i = 1,2.$ Estimates for prior probabilities can be obtained by $\hat{p}_i = \frac{n_i}{n}$, where $n = n_1 + n_2$ represents the total number of sample observations. The full multinomial model estimates the class-conditional densities by

$$f_i(x) = \frac{n_i(x)}{n}, \; i = 1,2. \tag{3.3}$$

where $n_i(x)$ is the number of individuals in a sample of size n_i from the population having response pattern X . The classification rule is: classify an item with response pattern X into π_i if

$$q_1 \frac{n_1(x)}{n_1} > q_2 \frac{n_2(x)}{n_2} \tag{3.4}$$

and to π_2 if $q_1 \frac{n_1(x)}{n_1} < q_2 \frac{n_2(x)}{n_2}$ (3.5)

and with probability $\frac{1}{2}$ if $q_1 \frac{n_1(x)}{n_1} = q_2 \frac{n_2(x)}{n_2}$ (3.6)

The full multinomial rule is simple to apply and the computation of apparent error does not require rigorous computational formula. However, Pires and bronco (2004) noted as pointed out by Dillon and Goldstein (1978) that one of the undesirable properties of the full multinomial Rule is the way it treats zero frequencies. If $n_1(x) = 0$ and $n_2(x) \neq 0$, a new observation with vector X will be allocated to π_2, irrespective of the sample sizes n_1 and n_2

3.2 The Linear Discriminant Function (LDF).

The linear discriminant function for discrete variables is given by

$$\hat{L}(x) = \sum_j \sum_k (\hat{p}_{2j} - \hat{p}_{1j}) s^{kj} x_k - \tfrac{1}{2} \sum \sum (\hat{p}_{2j} - \hat{p}_{1j}) s^{kj} (\hat{p}_{2k} + \hat{p}_{1k}) \qquad (3.7)$$

where s^{kj} are the elements of the inverse of the pooled sample covariance matrix, $\hat{p}_{1j}$ and $\hat{p}_{2j}$ are the elements of the sample means in π_1 and π_2 respectively. The classification rule obtained using this estimation is: classify an item with response pattern X into v if

$$\sum_j \sum_k (\hat{p}_{2j} - \hat{p}_{1j}) s^{kj} X_k - \tfrac{1}{2} \sum_j \sum_k (\hat{p}_{2j} - \hat{p}_{1j}) s^{kj} (\hat{p}_{2k} + \hat{p}_{1k}) > 0 \qquad (3.8)$$

and to π_2 or otherwise.

3.3 The Predictive Rule (P-Rule)

If the non-informative conjugate prior distribution for the parameter P_i of the multinomial model is chosen, that is the Dirichlet distribution with parameter $\alpha=1$, then the posterior distribution will be a Dirichlet distribution with parameter z_i+1, where $z_i = (n_{i1}, \dots n_{is})^T$. (Note that $\sum_{j=1}^{s} n_{ij} = n_i$). The Dirichlet distribution with parameter α has a density function given by

$$f(p) = \frac{\Gamma(\alpha_1 + \dots + \alpha_s)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_s)} \pi_{i=1}^{s} p_i^{\alpha_i - 1}, 0 < P_i < 1, \sum_{i=1}^{s} P_i = 1, \alpha_i > 0 \qquad (3.9)$$

Therefore the predictive density is simply

$$h_i(x / z_i) = \int_p p_{ij} \frac{\Gamma(s + n_i)}{\Gamma(1 + n_{i1}) \dots \Gamma(1 + n_{is})} \pi_{k=1}^{s} p_{ik}^{nik} \, dp \qquad (3.10)$$

$$= \frac{n_{ij} + 1}{n_i + s} = \frac{n_i(x) + 1}{n_i + s} \qquad (0 < p_{ij} < 1, \sum_{j=1}^{s} p_{ij} = 1 \qquad (3.11)$$

which leads to the predictive rule (or the P-rule)

Classify in π_1 if: $\dfrac{n_1(x) + 1}{n_1 + s} > \dfrac{n_2(x) + 1}{n_2 + s}$ (3.12)

Classify in π_2 if: $\dfrac{n_1(x) + 1}{n_1 + s} < \dfrac{n_2(x) + 1}{n_2 + s}$ (3.13)

Classify randomly if: $\frac{n_1(x)+1}{n_1+s} = \frac{n_2(x)+1}{n_2+s}$ (3.14)

Once again for $n_1 = n_2$, this rule is equivalent to the M-rule. The P-rule also avoids the zero frequency problems. For instance $n_1(x) = 0$ and $n_2(x) < (n_2+s)/(n_1+s)-4$ leads to classification in π_1.

4. The Simulation Experiments and Results

The four classification procedures are evaluated at each of the 118 configurations of n, r and d. The 118 configurations of n, r and d are all possible combinations of n = 40, 60, 80, 100, 200, 300, 400, 600, 700, 800, 900, 1000, r = 3, 4, 5 and d = 0.1, 0.2, 0.3, and 0.4. A simulation experiment which generates the data and evaluates the procedures is now described.

(i) A training data set of size n is generated via R-program where $n_1 = n/2$ observations are sampled from π_1 which has multivariate Bernoulli distribution with input parameter p_1 and $n_2 = n/2$ observations sampled from π_1, which is multivariate Bernoulli with input parameter $p_2, j = 1..r$. These samples are used to construct the rule for each procedure and estimate the probability of misclassification for each procedure is obtained by the plug-in rule or the confusion matrix in the sense of the full multinomial.

(ii) The likelihood ratios are used to define classification rules. The plug-in estimates of error rates are determined for each of the classification rules.

(iii) Step (i) and (ii) are repeated 1000 times and the mean plug-in error and variances for the 1000 trials are recorded. The method of estimation used here is called the resubstitution method.

The following table contains a display of one of the results obtained

Table 4.1(a) Apparent error rates for classification rules under different parameter values, sample sizes and Replications

$P_1 = (.3, .3, .3, .3, .3)$ $P_2 = (.7, .7, .7, .7, .7)$

Sample sizes	Optimal	Full M.	PR	LD
40	0.157125	0.110074	0.110787	0.204512
60	0.161900	0.127855	0.127958	0.207491
100	0.163290	0.143526	0.143680	0.209940
140	0.162967	0.149837	0.150407	0.209826
200	0.162565	0.156384	0.155280	0.211542
300	0.162783	0.159788	0.159641	0.211480
400	0.404243	0.384500	0.381672	0.414226
600	0.163018	0.161992	0.162603	0.213520
700	0.163075	0.162454	0.162878	0.213358
800	0.163463	0.163084	0.163318	0.213873
900	0.163354	0.163508	0.163218	0.214135
1000	0.163273	0.162916	0.163162	0.214277

p(mc) = 0.16308

Table 4.1(b) Actual Error rate for the classification rules under different parameter values, sample sizes and replications.

P_1 = (.3, .3, .3, .3, .3) P_2 = (.7, .7, .7, .7, .7) $\left|p(mc) - \hat{p}(mc)\right|$

Sample size	Optimal	Full M.	PR	LD
40	0.040271	0.052706	0.037112	0.041686
60	0.032751	0.042691	0.031487	0.033007
100	0.027786	0.037015	0.026152	0.027125
140	0.022462	0.031623	0.022112	0.024082
200	0.017981	0.026657	0.018218	0.019071
300	0.0150903	0.020882	0.015743	0.015671
400	0.012793	0.018476	0.013194	0.014210
600	0.010874	0.014643	0.011278	0.011926
700	0.009666	0.013574	0.009999	0.010861
800	0.009308	0.012778	0.009379	0.009582
900	0.008725	0.012243	0.008765	0.0090252
1000	0.010713	0.022517	0.012981	0.010732

Tables 4.1(a) and (b) present the mean apparent error rates and standard deviation (actual error rates) for classification rules under different parameter values. The mean apparent error rates increases with the increase in sample sizes and standard deviation decreases with the increase in sample sizes. From the analysis, optimal

is ranked first, followed by linear discriminant analysis, predictive rule and full multinomial come last.

Classification Rule	Performance
Optimal (OP)	1
Linear Discriminant Analysis (LDA)	2
Predictive Rule (PR)	3
Full Multinomial (FM)	4

5 Conclusion

We obtained two major results from this study. Firstly, using the simulation experiments we ranked the procedures as follows: Extended Optimal, Linear Discriminant Function, Predictive and Full Multinomial. The best method was the extended optimal procedure. Secondly, we concluded that it is better to increase the number of variables because accuracy increases with increasing number of variables. Moreover, our study showed that the extended optimal are more flexible in such a way to allow the analyst to incorporate some priori information in the models. Nevertheless, this does not exclude the use of other statistical techniques once the required hypotheses are satisfied.

Acknowledgements I am grateful to Assoc. Prof. E.C. Nwogu for helpful comments and Ndukauba Roseline for typesetting the paper.

References

[1] P. Arabie, I.J. Hubert and G. de Sote, *Clustering and classification*, River Edge NJ, World Scientific, 1996.

[2] K. Clayton, F. Blumberg and D.P. Auld, The relationship between motivation, learning strategies and choice of environment or including an online component, *British Journal of Educational technology*, 41, (2010), 349-364.

[3] W.R. Dillon and M. Goldstain, On the performance of some multinomial classification rules, *Journal of American Statistical Association*, **73**(362), (1978), 305-320.

[4] B.K. Keogh, Revisiting classification and identification, *Learning Disability Quarterly,* 28, (2005), 100-102.

[5] A. Lillvist, Observation of social competence of children in need of special support based on traditional disability categories versus a functional approach, *Early Child Development and Care,* 18, (2010), 1129-1142.

[6] I.C. Mammarella and C. Cornoldi, Spatial working memory and arithmetic deficits in children with nonverbal learning difficulties, *Journal of Learning Disabilities,* 43, (2010), 455-468.

[7] S.I. Onyeagu, Derivation of an optimal classification rule for discrete variables, *Journal of Nigerian Statistical Association*, 4, (2003), 79-80.

[8] A.M. Pires and J.A. Branco, Comparison of multinomial classification rules, *Journal of the American Statistics Association*, 73, (2004), 305-313.

[9] J.A. Russell, A discriminant analysis of the factors associated with the career plans of string music educators, *Journal of Research in Music Education, 5*6, (2008), 204-219.

[10] E. Zigler and L. Phillips, Psychiatric diagnosis: A critique, *Journal of Abnormal and Social Psychology,* 63, (1961), 607-618.