Problem of Classification

Unit - IV

Problem of Classification

Introduction:- The problem of classification arises when on the basis of measurements on an individual, we wish to classify the individual into one of several categories usually specified by prob. distⁿ of the measurements. Thus, an individual is considered as a random obsⁿ and we have to decide about the popⁿ from which it could arise.

The problem of classification may be considered as a problem of statistical decision funcⁿs. We have a no. of hyp. with respect to the distⁿ of the obsⁿs out of which we have to accept one against the another.

Ex:- Let us consider the admission of a student on the basis of his score in a test and classify the student one of the two classes namely (i) of those students who will successfully complete the training (test).
(ii) Of those who will not.

The main objective in the classification Procedure :- The main objective in the classification procedure is to minimize the risk (expected loss of misclassification), Let us consider the case of two popⁿ, suppose an individual in question is an obsⁿ from either popⁿ π_1 or π_2. Let the vector of measurements on the individual be $\underset{\sim}{x} = (x_1, x_2, \dots x_p)'$. We may regard it has a point in p-dimentional space. We divide the space into two regions say R_1 and R_2 such that, if the point $\underset{\sim}{x}$ falls in R_1, we classify the individual to the popⁿ π_1 and if it falls in R_2 then we assign the individual to the popⁿ π_2.

The misclassification may arise in two ways

(1) The individual may be classified into π_2 when it actually belong to π_1.

(2) The individual may be classifed into π_1 when it actually belong to π_2.

These two misclassification are known as Type One and Type Two misclassification.

Let the cost of Type One misclassification be $C(2|1) \geq 0$ and the cost of Type two misclassification be $c(1|2) > 0$.

These cost may be measured in any kinds of unit. The cost of misclassification can be summarised as

Statistical Disⁿ

		π_1	π_2	
popⁿ	π_1	0	$c(2\|1)$	Type I
	π_2	$c(1\|2)$	0	
		Type II		

Discrimination and Classification :- Discrimination and classification are multivariate techniques concerned with separating distinct set of obsⁿ's, and with (or objects) allocating with new obsⁿ to previously defined groups. Discriminant Analysis rather explorately in nature. As a separative procedure, it is often employed on a one time bases in ordered to investigate observed differences when casual relationship are not well understood. Classification procedure are less exploratory in the sense that they need to well defined rules, which can be use for assigning new objects. Classification ordinarily requires more problem structure than discriminant does. Thus, the imidiate goal of discrimation and classification respectivelly are as follows :

(4) Goal One :- To describe either graphically or algebrically, the differential features of obs^{n}s from several known pop^{n}s. We try to find discriminonts whose numerical values are such that the pop^n are separated as much as possible. In other words, Discriminant Analysis is a useful tool for situation where the Total samples (or pop^n) is to be divide into two or more mutually collectively exhaustive roots on the basis of the set of the preditor variables.

Ex :- A problem involving classify people into successful or unsuccessful, classifying customer into owner and non-owner of video tap recorder.

Goal Two :- Goal two which is reffer to classifaction, to sort obs^{n}s into two or more label classes. The emphasis is on driving a rule which can be use to optimally assign new objects to the labeled classes.

Notations :-

(1) Problem of Type One Misclassification = Problem of classifying the $\underset{\sim}{x}$ in π_2 when it belongs to π_1. According to the procedure risk $R = P(2|1|R)$

(2) Problem of Type Two Misclassification = Problem of classifying $\underset{\sim}{x}$ into π_1 when it belongs to π_2.

According to the procedure

$$R = P(1|2|R)$$

Now, cost of Type One misclassification is denated by $C(2|1|R)$ and cost of Type Two misclassification is denated by $C(1|2|R)$

Average Loss Due to Misclassification :-

Here we consider two cases:

Case I :- When prior prob. are known:

Let $R = (R_1, R_2)$ be the procedure of classification of an obs^n into one of the two pop^{n}s say π_1 and π_2.

Let the pdf of $\underset{\sim}{x}$, when it belongs to π_1 and π_2 by $p_1(\underset{\sim}{x})$ and $p_2(\underset{\sim}{x})$ respectively. Now, define the procedure $R = (R_1, R_2)$ as fallows.

If $\underset{\sim}{x} \in R_1$ then classify the obs^n in π_1 and if $\underset{\sim}{x} \in R_2$, the classify the obs^n in π_2.

Thus the prob. of correct classification are

$$P(1|1,R) = \int_{R_1} p_1(\underset{\sim}{x})\,d\underset{\sim}{x} \quad \text{and}$$

$$P(2|2,R) = \int_{R_2} p_2(\underset{\sim}{x})\,d\underset{\sim}{x}$$

and the prob. of misclassifications are

$$P(2|1,R) = \int_{R_2} p_1(\underset{\sim}{x})\,d\underset{\sim}{x}$$

and $$P(1|2,R) = \int_{R_1} p_2(\underset{\sim}{x})\,d\underset{\sim}{x}$$

If the prob. of drawing an obs^n from π_1 is q_1, then the prob. of drawing an obs^n from π_1 and correctly classify it is $q_1 P(1|1,R)$. Similarly, the prob of drawing an obs^n from π_1 and misclassifying it is $q_1 P(2|1,R)$.

Also, let the prob. of drawing an obs^n from π_2 be q_2, then the prob. of drawing an obs^n from π_2 and correctly classifying it is $q_2 P(2|2,R)$ and the prob. of drawing an obs^n from π_2 and misclassifying it is $q_2 P(1|2,R)$.

Thus, the avg loss due to misclassification is

$$L = C(2|1).P(2|1,R)q_1 + C(1|2).P(1|2,R)q_2$$

This is the avg loss that we wish to minimized. For given q_1 and q_2, if R_1 and R_2 are regions in which the whole space is divided are so choosen that the risk (loss)(L) is minimize, we call the procedure R as base procedure.

Case IInd :- when prior prob. are not known.

In this case the expected loss if the obs^n is from π_1 is

$$r(1,R) = C(2|1)P(2|1,R)$$

and the expected loss if the obs^n is from π_2 is

$$r(2,R) = C(1|2)P(1|2,R)$$

A procedure R is atleast as good as procedure R^* if ~~$r(1,R) \le$~~

$$r(1,R) \le r(1,R^*) \text{ and}$$
$$r(2,R) \le r(2,R^*)$$

R is better than R^*, If atleast one of these inequalities is strictly inequality (hold).

R is called admissible if there is no procedure R^*, that is better than R.

Procedure of classification into One of the two pop^{n}s with known prob. dist^n :-

Case I :- when the prior prob. are known:

Let $R=(R_1, R_2)$ be the procedure of classification of an obs^n $\underset{\sim}{x}$ into one of the two pop^n π_1 and π_2. Let the pdf of $\underset{\sim}{x}$ when it belongs to π_1 and π_2 be $p_1(\underset{\sim}{x})$ and $p_2(\underset{\sim}{x})$ respectively.

(8) Let q_1 and q_2 be the prior prob. of drawing an obs^n from π_1 and π_2 respectively. Then, we select R_1 and R_2 of R such that the total loss due to misclassification is minimum (when cost of misclassification are not taken into account). For simplicity we assume that cost of type 1 misclassification is equal to type 2 misclassification i.e. $c(1|2) = c(2|1)$

Then the expected loss is

$$L = q_1 P(2|1, R) + q_2 P(1|2, R)$$

$$= q_1 \int_{R_2} p_1(\underset{\sim}{x}) d\underset{\sim}{x} + q_2 \int_{R_1} p_2(\underset{\sim}{x}) d\underset{\sim}{x} \quad \text{---(1)}$$

This is also called the prob. of misclassification. Hence, we wish to minimize the prob. of misclassification.

A procedure R which minimize eq^n (1) is defined as follow:

Assign the obs^n $\underset{\sim}{x}$ to π_1 if

$$\frac{q_1 p_1(\underset{\sim}{x})}{q_1 p_1(\underset{\sim}{x}) + q_2 p_2(\underset{\sim}{x})} \geq \frac{q_2 p_2(\underset{\sim}{x})}{q_1 p_1(\underset{\sim}{x}) + q_2 p_2(\underset{\sim}{x})}$$

otherwise, classify $\underset{\sim}{x}$ to π_2. This leads to the rule

$$\left.\begin{aligned} R_1 &: q_1 p_1(\underset{\sim}{x}) \geq q_2 p_2(\underset{\sim}{x}) \\ R_2 &: q_1 p_1(\underset{\sim}{x}) < q_2 p_2(\underset{\sim}{x}) \end{aligned}\right\} \quad \text{---(2)}$$

** If ~~q_1~~ $q_1 p_1(\underset{\sim}{x}) = q_2 p_2(\underset{\sim}{x})$, the point could be classified as either from π_1 or π_2, we have arbitrary put it into R_1.

loss = 0

** If $q_1 p_1(\underset{\sim}{x}) + q_2 p_2(\underset{\sim}{x}) = 0$ for a given $\underset{\sim}{x}$, that point also may go into either region.

Now, let us proof formally that eq^n (2) is the best procedure. Let $R^* = (R_1^*, R_2^*)$ be an alternative procedure of classification, then the expected loss due to misclassification is

$$q_1 \int_{R_2^*} p_1(\underset{\sim}{x}) d\underset{\sim}{x} + q_2 \int_{R_1^*} p_2(\underset{\sim}{x}) d\underset{\sim}{x}$$

$$= q_1 \int_{R_2^*} p_1(\underset{\sim}{x}) d\underset{\sim}{x} - q_2 \int_{R_2^*} p_2(\underset{\sim}{x}) d\underset{\sim}{x} + q_2 \int_{R_2^*} p_2(\underset{\sim}{x}) d\underset{\sim}{x} + q_2 \int_{R_1^*} p_2(\underset{\sim}{x}) d\underset{\sim}{x}$$

$$= \int_{R_2^*} \left(q_1 p_1(\underset{\sim}{x}) - q_2 p_2(\underset{\sim}{x})\right) d\underset{\sim}{x} + q_2 \int_{R^*} p_2(\underset{\sim}{x}) d\underset{\sim}{x}$$

$$= \int_{R_2^*} \left(q_1 p_1(\underset{\sim}{x}) - q_2 p_2(\underset{\sim}{x})\right) d\underset{\sim}{x} + q_2 \times 1$$

Hence, the second term is constant. Therefore expected loss will be min. if first term is min. The first term is minimize if R_2^* includes the point $\underset{\sim}{x}$ such that $q_1 p_1(\underset{\sim}{x}) < q_2 p_2(\underset{\sim}{x})$ and excludes the point $\underset{\sim}{x}$ for which $q_1 p_1(\underset{\sim}{x}) > q_2 p_2(\underset{\sim}{x})$.

If $P\left[\frac{p_1(\underset{\sim}{x})}{p_2(\underset{\sim}{x})} = \frac{q_2}{q_1} \,\middle|\, \pi_j\right] = 0 \; ; j = 1, 2$

Then the bayes procedure is unique ~~ex~~ except for set of prob. zero.

If $c(1|2) \neq 1$ and $c(2|1) \neq 1$, then $R : (R_1, R_2)$

$$L = c(2|1) q_1 P(2|1, R) + c(1|2) q_2 P(1|2, R)$$

$$= c(2|1) q_1 \int_{R_2} p_1(\underset{\sim}{x}) d\underset{\sim}{x} + c(1|2) q_2 \int_{R_1} p_2(\underset{\sim}{x}) d\underset{\sim}{x}$$

We choose R_1 and R_2 according to

$R_1 : c(2|1) q_1 p_1(\underset{\sim}{x}) \geq c(1|2) q_2 p_2(\underset{\sim}{x})$

$R_2 : c(2|1) q_1 p_1(\underset{\sim}{x}) < c(1|2) q_2 p_2(\underset{\sim}{x})$

Since cost are non-negative constant. Therefore, we have classified into π_1 if

$$R_1 : \frac{p_1(\underset{\sim}{x})}{p_2(\underset{\sim}{x})} \geq \frac{q_2 c(1|2)}{q_1 c(2|1)}$$

$$R_2 : \frac{p_1(\underset{\sim}{x})}{p_2(\underset{\sim}{x})} < \frac{q_2 c(1|2)}{q_1 c(2|1)}$$

Case 2: If prior prob. are not known.

If q_1 and q_2 are not ~~un~~known then we adopt minimax procedure for classification. i.e. A procedure which minimize the max. expected loss. $r(i, R)$; $i = 1, 2$

where, $r(1, R) = c(2|1) \cdot P(2|1, R)$ and

$r(2, R) = c(1|2) \cdot P(1|2, R)$

The procedure R in this is called minimax procedure and its regions R_1 and R_2 are known as minimax region of classification.

Classification into One of two known multi-variate Normal pop^n :- Let the two pop^n be $\pi_1 : N_p(\underset{\sim}{\mu}^{(1)}, \underset{\sim}{\Sigma})$ and $\pi_2 : N_p(\underset{\sim}{\mu}^{(2)}, \underset{\sim}{\Sigma})$ where the par. $\underset{\sim}{\mu}^{(1)}, \underset{\sim}{\mu}^{(2)}$ and $\underset{\sim}{\Sigma}$ are known. Then the i^{th} density of $\underset{\sim}{x} (= x_1, x_2, \ldots x_p)'$ is

$$p_i(\underset{\sim}{x}) = \frac{1}{(2\pi)^{p/2}} \cdot \frac{1}{|\underset{\sim}{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{\mu}^{(i)})' \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}^{(i)})} \; ; i = 1, 2 \quad \text{---(1)}$$

The ratio of the ~~two~~ density is

$$\frac{p_1(\underset{\sim}{x})}{p_2(\underset{\sim}{x})} = \frac{e^{-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{\mu}^{(1)})' \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}^{(1)})}}{e^{-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{\mu}^{(2)})' \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}^{(2)})}}$$

$$= \exp\left[-\frac{1}{2}\left\{(\underset{\sim}{x} - \underset{\sim}{\mu}^{(1)})' \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}^{(1)}) - (\underset{\sim}{x} - \underset{\sim}{\mu}^{(2)})' \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}^{(2)})\right\}\right]$$

$$= \exp\left[-\frac{1}{2}\left\{\cancel{\underset{\sim}{x}' \underset{\sim}{\Sigma}^{-1} \underset{\sim}{x}} - \underset{\sim}{x}' \underset{\sim}{\Sigma}^{-1} \underset{\sim}{\mu}^{(1)} - \underset{\sim}{\mu}^{(1)'} \underset{\sim}{\Sigma}^{-1} \underset{\sim}{x} + \underset{\sim}{\mu}^{(1)'} \underset{\sim}{\Sigma}^{-1} \underset{\sim}{\mu}^{(1)} - \cancel{\underset{\sim}{x}' \underset{\sim}{\Sigma}^{-1} \underset{\sim}{x}} + \underset{\sim}{x}' \underset{\sim}{\Sigma}^{-1} \underset{\sim}{\mu}^{(2)} + \underset{\sim}{\mu}^{(2)'} \underset{\sim}{\Sigma}^{-1} \underset{\sim}{x} - \underset{\sim}{\mu}^{(2)'} \underset{\sim}{\Sigma}^{-1} \underset{\sim}{\mu}^{(2)}\right\}\right]$$

$$= \exp\left[-\frac{1}{2}\left\{-2\, \underset{\sim}{x}' \underset{\sim}{\Sigma}^{-1} \underset{\sim}{\mu}^{(1)} + 2\, \underset{\sim}{x}' \underset{\sim}{\Sigma}^{-1} \underset{\sim}{\mu}^{(2)} + \underset{\sim}{\mu}^{(1)'} \underset{\sim}{\Sigma}^{-1} \underset{\sim}{\mu}^{(1)} - \underset{\sim}{\mu}^{(2)'} \underset{\sim}{\Sigma}^{-1} \underset{\sim}{\mu}^{(2)}\right\}\right]$$

(12)

$= \exp\left[x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \frac{1}{2}\{\mu^{(1)'}\Sigma^{-1}\mu^{(1)} - \mu^{(2)'}\Sigma^{-1}\mu^{(2)}\}\right]$

$= \exp\left[x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \frac{1}{2}\{\mu^{(1)'}\Sigma^{-1}\mu^{(1)} + \mu^{(1)'}\Sigma^{-1}\mu^{(2)} - \mu^{(1)'}\Sigma^{-1}\mu^{(2)} - \mu^{(2)'}\Sigma^{-1}\mu^{(2)}\}\right]$

$= \exp\left[x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \frac{1}{2}\{\mu^{(1)'}\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) + \mu^{(2)'}\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\}\right]$

$= \exp\left[x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \frac{1}{2}\{(\mu^{(1)}+\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\}\right]$

$\{\because \mu^{(1)'}\Sigma^{-1}\mu^{(2)} = \mu^{(2)'}\Sigma^{-1}\mu^{(1)}\}$

Let R_1 be the region of classification in which if x falls, we classify the obs^n into π_1, then R_1 is given by

$$\frac{p_1(x)}{p_2(x)} \geq K$$

where K is some suitable choosen constant.

$\Rightarrow \exp\left[x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \frac{1}{2}(\mu^{(1)}+\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\right] \geq K$

$\Rightarrow x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \frac{1}{2}(\mu^{(1)}+\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) \geq \log K$

In the above expression the first term is the "well known discriminent func^n." It is a linear func^n of the components of the obs^n vector. Thus, if π_i $(i=1,2)$ has the density eq^n (1), the best region of classification are given by

$R_1: \underbrace{x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})}_{\text{Discriminent f}^n} - \frac{1}{2}(\mu^{(1)}+\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) \geq \log K$

$R_2: x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \frac{1}{2}(\mu^{(1)}+\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) < \log K$

Particular Case :- If the prior prob is such that $K=1$, then best region of classification are given by

$R_1: x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \frac{1}{2}(\mu^{(1)}+\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) \geq \log 1 = 0$

$R_2: x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \frac{1}{2}(\mu^{(1)}+\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) < 0$

Distribution of 'U' $= x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \frac{1}{2}(\mu^{(1)}+\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})$:-

Case I :- when $x \sim N_p(\mu^{(1)}, \Sigma)$:

Since U is a linear f^n of the components of x. So, U will follow a univariate normal dist^n with mean and var as follow.

$$E_1(U) = \mu^{(1)'}\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \frac{1}{2}(\mu^{(1)}+\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})$$

$$= \left\{\mu^{(1)'} - \frac{1}{2}(\mu^{(1)})' - \frac{1}{2}\mu^{(2)'}\right\}\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})$$

$$= \frac{1}{2}(\mu^{(1)}-\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})$$

$$= \frac{\alpha}{2} \text{ (say)}$$

where, α is the distance b/w two p variate normal pop^n as defined by mahanalobies.

$$V_1(U) = V_1\left[x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \frac{1}{2}(\mu^{(1)}+\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\right]$$

$$= V_1\left[x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\right]$$

$$= E_1\left[x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - E_1\{x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\}\right]^2$$

$$= E_1\left[x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \mu^{(1)'}\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\right]^2$$

$$= E_1\left[(x-\mu^{(1)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\right]^2$$

$$= E_1\left[\left[(x-\mu^{(1)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\right]'\left[(x-\mu^{(1)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\right]\right]$$

$$= E_1\left[(\mu^{(1)}-\mu^{(2)})'\Sigma^{-1}(x-\mu^{(1)})(x-\mu^{(1)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\right]$$

$$= (\mu^{(1)}-\mu^{(2)})'\Sigma^{-1}E_1\left[(x-\mu^{(1)})(x-\mu^{(1)})'\right]\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})$$

$$= (\mu^{(1)}-\mu^{(2)})'\Sigma^{-1}\Sigma\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})$$

$$= (\mu^{(1)}-\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})$$

$$= \alpha$$

Thus, $\boxed{U \sim N_1(\alpha/2, \alpha) \text{ when } X \sim N_p(\mu^{(1)}, \Sigma)}$

Case II :- when $X \sim N_p(\mu^{(2)}, \Sigma)$

Since U is a linear f^n of the components of X. So, U will follow a univariate normal dist^n with mean and var as follow.

$$E_2(U) = \mu^{(2)'}\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \frac{1}{2}(\mu^{(1)}+\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})$$

$$= \left\{\mu^{(2)'} - \frac{1}{2}(\mu^{(1)'}) - \frac{1}{2}\mu^{(2)'}\right\}\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})$$

$$= +\frac{1}{2}(\mu^{(2)'}-\mu^{(1)'})\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})$$

$$= -\frac{1}{2}(\mu^{(1)}-\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})$$

$$= -\frac{\alpha}{2} \text{ (say)}$$

where, α is the distance b/w two p variate normal pop^n as defined by mahanalobies.

$$V_2(U) = V_2\left[x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \frac{1}{2}(\mu^{(1)}+\mu^{(2)})'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\right]$$

$$= V_2\left[x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\right]$$

$$= E_2\left[x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - E_2\{x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\}\right]^2$$

$$= E_2\left[x'\Sigma^{-1}(\mu^{(1)}-\mu^{(2)}) - \mu^{(2)'}\Sigma^{-1}(\mu^{(1)}-\mu^{(2)})\right]^2$$

$= E_2\left[(x - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)}) \right]^2$

$= E_2\left[\{(x-\mu^{(2)})' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)})\}' \{(x-\mu^{(2)})' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)})\} \right]$

$= E_2\left[(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (x-\mu^{(2)}) (x-\mu^{(2)})' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)}) \right]$

$= (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} E_2\left[(x-\mu^{(2)}) (x-\mu^{(2)})' \right] \Sigma^{-1} (\mu^{(1)} - \mu^{(2)})$

$= (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} \Sigma \Sigma^{-1} (\mu^{(1)} - \mu^{(2)})$

$= (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)})$

$= \alpha$

Thus, $\boxed{U \sim N_2(-\alpha/2, \alpha) \text{ when } x \sim N_p(\mu^{(2)}, \Sigma)}$

An Alternative Procedure to obtain the discriminent f^n :-

Suppose pop^n π_1 and π_2 are fully specified as $\pi_1 : N_p(\mu^{(1)}, \Sigma)$ and $\pi_2 : N_p(\mu^{(2)}, \Sigma)$, then R.A. Fisher define the discreminent f^n b/w two p-variate normal pop^n's as that linear f^n of the components of x for which the ratio

$$Q = \frac{[E_1(x'l) - E_2(x'l)]^2}{V(x'l)} \text{ is max.} \quad \text{---(1)}$$

where, $E_1 = E(x)$ when $x \in \pi_1$,

$E_2 = E(x)$ when $x \in \pi_2$.

Let $y = l_1x_1 + l_2x_2 + \dots + l_px_p = l'x = x'l$ be a linear f^n of the components of x where $l_1, l_2, \dots l_p$ are constant and we shall determine these constant such that eq^n (1) is maximum.

To maximize eq^n (1), we shall use Lagrange's multiplier method. According to this method we shall maximize the Numerator of eq^n (1) holding the denominator in eq^n (1) fixed.

Consider a Numerator

$$[E_1(x'l) - E_2(x'l)]^2 = [l'\mu^{(1)} - l'\mu^{(2)}]^2$$

$$= [(\mu^{(1)} - \mu^{(2)})'l]^2$$

$$= l'(\mu^{(1)} - \mu^{(2)})(\mu^{(1)} - \mu^{(2)})'l$$

$$= l'dd'l \quad \text{where, } d = (\mu^{(1)} - \mu^{(2)})$$

Now, denominator of eq^n (1) is

$$V(x'l) = E[x'l - E(x'l)]^2$$

$$= E[\{x' - E(x')\}l]^2$$

$$= l'E[\{x' - E(x')\}\{x' - E(x')\}']l$$

$$= l'\Sigma l$$

$\therefore$ eq^n (1) reduces to

(18)

$$Q = \frac{\underset{\sim}{l}'\underset{\sim}{d}\underset{\sim}{d}'\underset{\sim}{l}}{\underset{\sim}{l}'\underset{\sim}{\Sigma}\underset{\sim}{l}} \quad \text{—— (2)}$$

If λ is a lagrange's multiplier, then we have to maximize

$$\phi = \underset{\sim}{l}'\underset{\sim}{d}\underset{\sim}{d}'\underset{\sim}{l} - \lambda(\underset{\sim}{l}'\underset{\sim}{\Sigma}\underset{\sim}{l} - c) \quad \text{—— (3)}$$

where, c is a constant.

Now, $\frac{\delta\phi}{\delta \underset{\sim}{l}} = 0$ gives

$$= 2\underset{\sim}{d}\underset{\sim}{d}'\underset{\sim}{l} - 2\lambda \underset{\sim}{\Sigma}\underset{\sim}{l} = 0$$

$$\Rightarrow \quad \underset{\sim}{d}\underset{\sim}{d}'\underset{\sim}{l} = \lambda\underset{\sim}{\Sigma}\underset{\sim}{l} \quad \text{—— (4)}$$

Since $\underset{\sim}{d}'\underset{\sim}{l}$ is a scalar quantity say ν then from eq^n (4), we have

$$\underset{\sim}{d} = \frac{\lambda}{\nu}\underset{\sim}{\Sigma}\underset{\sim}{l}$$

$$\Rightarrow \underset{\sim}{l} = \frac{\nu}{\lambda}\underset{\sim}{\Sigma}^{-1}\underset{\sim}{d}$$

$$\Rightarrow \underset{\sim}{l} = k\underset{\sim}{\Sigma}^{-1}\underset{\sim}{d} \quad \text{where, } k = \frac{\nu}{\lambda}$$

$$\Rightarrow \underset{\sim}{l} \propto \underset{\sim}{\Sigma}^{-1}\underset{\sim}{d} \qquad \underset{\sim}{d}'\underset{\sim}{l} = \underset{\sim}{d}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{d}$$

Thus, the best discriminant f^n is

$$y = \underset{\sim}{l}'\underset{\sim}{x} = \underset{\sim}{x}'\underset{\sim}{l} = \underset{\sim}{x}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{d}$$
$$= \underset{\sim}{x}'\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\mu}^{(1)} - \underset{\sim}{\mu}^{(2)})$$

(19)

Classification into One of the two Normal pop^n when the par. are estimated:-

Suppose we wish to classify an obs^n $\underset{\sim}{x}$ as coming from π_1 or π_2 where $\pi_1: N_p(\underset{\sim}{\mu}^{(1)}, \underset{\sim}{\Sigma})$ and $\pi_2: N_p(\underset{\sim}{\mu}^{(2)}, \underset{\sim}{\Sigma})$ which are completely specified.

Let $\underset{\sim}{x}_\alpha^{(i)}$; $\alpha = 1, 2, \dots N_i$, $i = 1,2$ be two ind r.s. from $N_p(\underset{\sim}{\mu}^{(i)}, \underset{\sim}{\Sigma})$; $i = 1,2$. The best estimate of $\underset{\sim}{\mu}^{(1)}$, $\underset{\sim}{\mu}^{(2)}$ and $\underset{\sim}{\Sigma}$ are

$$\hat{\underset{\sim}{\mu}}^{(1)} = \bar{\underset{\sim}{x}}^{(1)} = \sum_{\alpha=1}^{N_1}\frac{x_\alpha^{(1)}}{N_1}, \quad \hat{\underset{\sim}{\mu}}^{(2)} = \bar{\underset{\sim}{x}}^{(2)} = \sum_{\alpha=1}^{N_2}\frac{x_\alpha^{(2)}}{N_2}$$

and $$\hat{\underset{\sim}{\Sigma}} = \underset{\sim}{S} = \frac{1}{N_1+N_2-2}\sum_{i=1}^{2}\sum_{\alpha=1}^{N_i}(x_\alpha^{(i)} - \bar{x}^{(i)})(x_\alpha^{(i)} - \bar{x}^{(i)})'$$

The criteria for classifying $\underset{\sim}{x}$ is suggested as

$$\underset{\sim}{V} = \underset{\sim}{x}'\underset{\sim}{S}^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)}) - \frac{1}{2}(\bar{x}^{(1)} + \bar{x}^{(2)})'\underset{\sim}{S}^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)})$$

The first term of $\underset{\sim}{V}$ is the discriminant f^n based on two samples. $\underset{\sim}{V}$ is used for classification as

$R_1: V \geq c$, $R_2: V < c$

where, c is determine as the constant

The dist^n of V is quite complicated but it has shown by Wald in 1944 that has the dist^n of V is the same as dist^n of U when $N_1 \to \infty$, $N_2 \to \infty$. i.e.

$$V \xrightarrow[\text{}]{d\text{ (in law)}} U \quad \text{as } N_1 \to \infty,\ N_2 \to \infty$$

i.e. Under π_1, $V \xrightarrow{asym} N_1(\alpha/2, \alpha)$

π_2, $V \xrightarrow{asym} N_2(-\alpha/2, \alpha)$

where, $\alpha = (\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})$

Alternative Procedure to Obtain Discriminent f^n :-

Statement: The discriminent f^n $x' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})$ maximize

$$Q = \frac{\text{Var. b/w samples}}{\text{Var. within samples}} \quad \text{— (A)}$$

Proof:- Let us make transformation

$$y_\alpha^{(1)'} = x_\alpha^{(1)'} l \; ; \; y_\alpha^{(2)'} = x_\alpha^{(2)'} l$$

$$\therefore \; \bar{y}_\alpha^{(1)'} = \bar{x}_\alpha^{(1)'} l \; ; \; \bar{y}_\alpha^{(2)'} = \bar{x}_\alpha^{(2)'} l$$

Now, we have pooled mean

$$\bar{y} = \frac{N_1 \bar{y}^{(1)} + N_2 \bar{y}^{(2)}}{N_1 + N_2}$$

when two samples are combined together then the total variation due to two samples may be written as

$$\sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} (y_\alpha^{(i)} - \bar{y})^2 = \sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} (y_\alpha^{(i)} - \bar{y}^{(i)} + \bar{y}^{(i)} - \bar{y})^2$$

$$= \sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} (y_\alpha^{(i)} - \bar{y}^{(i)})^2 + \sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} (\bar{y}^{(i)} - \bar{y})^2 + 0$$

$$= \text{Variation within samples} + \text{Variation b/w samples.}$$

Now, consider variation b/w samples:-

$$\text{Var b/w samples} = \sum_{i=1}^{2} N_i (\bar{y}^{(i)} - \bar{y})^2$$

$$= N_1 (\bar{y}^{(1)} - \bar{y})^2 + N_2 (\bar{y}^{(2)} - \bar{y})^2$$

$$= N_1 \left(\bar{y}^{(1)} - \left(\frac{N_1 \bar{y}^{(1)} + N_2 \bar{y}^{(2)}}{N_1 + N_2} \right) \right)^2 + N_2 \left(\bar{y}^{(2)} - \left(\frac{N_1 \bar{y}^{(1)} + N_2 \bar{y}^{(2)}}{N_1 + N_2} \right) \right)^2$$

$$= N_1 \left(\frac{N_2 (\bar{y}^{(1)} - \bar{y}^{(2)})}{N_1 + N_2} \right)^2 + N_2 \left(\frac{N_1 (\bar{y}^{(2)} - \bar{y}^{(1)})}{N_1 + N_2} \right)^2$$

$$= \frac{N_1 N_2}{(N_1 + N_2)^2} (\bar{y}^{(1)} - \bar{y}^{(2)})^2 (N_2 + N_1)$$

$$= \frac{N_1 N_2}{(N_1 + N_2)} (\bar{y}^{(1)} - \bar{y}^{(2)})^2$$

$\therefore$ the mean variation of (variance) b/w samples is variation b/w samples ~~devided~~ devided by its df and is equal to

$$\text{Mean Variation of variance b/w samples} = \frac{1}{2-1} \frac{N_1 N_2}{N_1 + N_2} (\bar{y}^{(1)} - \bar{y}^{(2)})^2$$

Now, consider variation within samples

$$= \sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} (y_\alpha^{(i)} - \bar{y}^{(i)})^2$$

$$= \sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} (l' x_\alpha^{(i)} - l' \bar{x}^{(i)})^2$$

$$= l' \sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} (x_\alpha^{(i)} - \bar{x}^{(i)})(x_\alpha^{(i)} - \bar{x}^{(i)})' l$$

$$= l' (N_1 + N_2 - 2) S \, l$$

$\therefore$ The mean variation (of variance) within sample is equal to variation within sample devided by its df is given by

$$\text{Mean Variation (of variance) within sample} = \frac{1}{(N_1 + N_2 - 2)} l' (N_1 + N_2 - 2) S \, l$$

$$= l' S \, l$$

So, from eq^n (A), we get

$$Q = \frac{\frac{N_1 N_2}{N_1+N_2}(\bar{x}^{(1)} - \bar{x}^{(2)})^2}{l' S l}$$

$$\left\{ \begin{array}{l} (\bar{x}^{(1)} - \bar{x}^{(2)})^2 \\ (l'\bar{x}^{(1)} - l'\bar{x}^{(2)}) \\ = l'(\bar{x}^{(1)} - \bar{x}^{(2)})(\bar{x}^{(1)} - \bar{x}^{(2)})' l \\ = l' d d' l \end{array} \right.$$

$$Q = \frac{N_1 N_2}{N_1+N_2} \frac{l' d d' l}{l' S l} \quad \text{where, } d = (\bar{x}^{(1)} - \bar{x}^{(2)})$$

Thus, to maximize Q we maximize $l' d d' l$ w.r. to l keeping the denominator fixed. Using Lagrange's multiplier, we get

$$\phi = l' d d' l - \lambda (l' S l - c)$$

$$\frac{\delta \phi}{\delta l} = 0 \Rightarrow 2 d d' l - 2\lambda S l = 0$$

$$\Rightarrow d = \frac{\lambda S l}{d' l}$$

$$d = \frac{\lambda S l}{\mu} \quad \text{where, } \mu = d' l$$

Since $d' l$ is a scalar quantity say ν. then

$$d = \frac{\lambda}{\nu} S l$$

$$\Rightarrow d = K S l$$

$$\Rightarrow d \propto S l$$

$$\Rightarrow l \propto S^{-1} d$$

$$\text{or } \quad x' l \propto x' S^{-1} d$$

$$\propto x' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})$$

The best discriminant f^n is $x' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})$

Proved

Relationship b/w Discriminant Analysis and Regression Analysis:-

Consider the dummy var y such that

$$y_\alpha^{(1)} = \frac{N_2}{N_1+N_2} \; ; \alpha = 1, 2, \ldots N_1$$

$$y_\alpha^{(2)} = \frac{-N_1}{N_1+N_2} \; ; \alpha = 1, 2, \ldots, N_2$$

Consider the regression y on x

$$y_\alpha^{(i)} = b'(x_\alpha^{(i)} - \bar{x}) \; ; i = 1, 2$$

where, b is a vector which minimize the residual sum of square and

$$Q = \sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} (y_\alpha^{(i)} - b'(x_\alpha^{(i)} - \bar{x}))^2$$

Here, $\bar{x} = \frac{N_1 \bar{x}^{(1)} + N_2 \bar{x}^{(2)}}{N_1+N_2}$ is the pooled mean. Now, to minimize Q with respect to b

$$\frac{\delta Q}{\delta b} = 0$$

$\Rightarrow$ p normal equidistance

$$-2 \sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} (x_\alpha^{(i)} - \bar{x})(y_\alpha^{(i)} - b'(x_\alpha^{(i)} - \bar{x})) = 0$$

$$\Rightarrow \sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} y_\alpha^{(i)} (x_\alpha^{(i)} - \bar{x}) = \sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} (x_\alpha^{(i)} - \bar{x})^2 b$$

$$= \sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} (x_\alpha^{(i)} - \bar{x})(x_\alpha^{(i)} - \bar{x})' b \quad \text{--- (1)}$$

Note that

$$\sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} y_\alpha^{(i)} (x_\alpha^{(i)} - \bar{x}) = \sum_{\alpha=1}^{N_1} \frac{N_2}{N_1+N_2} (x_\alpha^{(1)} - \bar{x}) - \sum_{\alpha=1}^{N_2} \frac{N_1}{N_1+N_2} (x_\alpha^{(2)} - \bar{x})$$

$$= \frac{N_1 N_2}{N_1+N_2}(\bar{x}^{(1)} - \bar{x}) - \frac{N_1 N_2}{N_1+N_2}(\bar{x}^{(2)} - \bar{x})$$

$$= \frac{N_1 N_2}{N_1+N_2}(\bar{x}^{(1)} - \bar{x}^{(2)})$$

$$(x_\alpha^{(i)} - \bar{x})^2$$
$$(x_\alpha^{(i)} - \bar{x}^{(i)} + \bar{x}^{(i)} - \bar{x})^2$$

and $\sum_{i=1}^{2}\sum_{\alpha=1}^{N_i}(x_\alpha^{(i)} - \bar{x})(x_\alpha^{(i)} - \bar{x})'$

$$= \sum_{i=1}^{2}\sum_{\alpha=1}^{N_i}(x_\alpha^{(i)} - \bar{x}^{(i)})(x_\alpha^{(i)} - \bar{x}^{(i)})' + N_i\sum_{i=1}^{2}(\bar{x}^{(i)} - \bar{x})(\bar{x}^{(i)} - \bar{x})'$$

$$= (N_1+N_2-2)S + C$$

where, $C = N_i \sum_{i=1}^{2}(\bar{x}^{(i)} - \bar{x})(\bar{x}^{(i)} - \bar{x})'$

$$= N_1(\bar{x}^{(1)} - \bar{x})(\bar{x}^{(1)} - \bar{x})' + N_2(\bar{x}^{(2)} - \bar{x})(\bar{x}^{(2)} - \bar{x})'$$

$$= N_1\left(\bar{x}^{(1)} - \frac{N_1\bar{x}^{(1)} + N_2\bar{x}^{(2)}}{N_1+N_2}\right)^2 + N_2\left(\bar{x}^{(2)} - \frac{N_1\bar{x}^{(1)} + N_2\bar{x}^{(2)}}{N_1+N_2}\right)^2$$

$$= N_1\left(\frac{N_2(\bar{x}^{(1)} - \bar{x}^{(2)})}{N_1+N_2}\right)^2 + N_2\left(\frac{N_1(\bar{x}^{(2)} - \bar{x}^{(1)})}{N_1+N_2}\right)^2$$

$$= \frac{N_1 N_2}{(N_1+N_2)^2}(\bar{x}^{(1)} - \bar{x}^{(2)})^2(N_2+N_1)$$

$$= \frac{N_1 N_2}{(N_1+N_2)}(\bar{x}^{(1)} - \bar{x}^{(2)})^2$$

Now putting these values in eq^n (1), we get

$$\cancel{\frac{N_1 N_2}{N_1+N_2}(\bar{x}^{(1)} - \bar{x}^{(2)})} = \cancel{\frac{N_1 N_2}{N_1+N_2}(\bar{x}^{(1)} - \bar{x}^{(2)})^2 b}$$

$$\frac{N_1 N_2}{N_1+N_2}(\bar{x}^{(1)} - \bar{x}^{(2)}) = ((N_1+N_2-2)S + C)b$$

$$= (N_1+N_2-2)Sb + \frac{N_1 N_2}{N_1+N_2}(\bar{x}^{(1)} - \bar{x}^{(2)})^2 b$$

$$\Rightarrow \quad Sb = \frac{1}{N_1+N_2-2}\left[\frac{N_1 N_2}{N_1+N_2}\left\{(\bar{x}^{(1)} - \bar{x}^{(2)}) - (\bar{x}^{(1)} - \bar{x}^{(2)})^2 b\right\}\right]$$

$$= \frac{N_1 N_2}{(N_1+N_2)(N_1+N_2-2)}(\bar{x}^{(1)} - \bar{x}^{(2)})\left[1 - (\bar{x}^{(1)} - \bar{x}^{(2)})' b\right] \quad \text{--- (2)}$$

Note that, $\frac{N_1 N_2}{(N_1+N_2)(N_1+N_2-2)}(\bar{x}^{(1)} - \bar{x}^{(2)})\left[1 - (\bar{x}^{(1)} - \bar{x}^{(2)})' b\right]$ is a scalar quantity, and common multiplier of these p eq^n's, then the soln of b of eq^n (2) is proportional to

$$Sb \propto (\bar{x}^{(1)} - \bar{x}^{(2)})$$
$$b \propto S^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)})$$
$$x'b \propto x'S^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)})$$

Thus, the best discriminant f^n is given by $x'S^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)})$.

The Likelihood Criteria for deciding a problem of testing a composite Null Hypothesis :-

The problem of classification can be considered as a problem of testing a composite null hypothesis.

H_0: $\underset{\sim}{x}, \underset{\sim}{x}_\alpha^{(1)} \sim N_p(\underset{\sim}{\mu}^{(1)}, \underset{\sim}{\Sigma})$, $\alpha = 1, 2, \ldots N_1$

$\underset{\sim}{x}_\alpha^{(2)} \sim N_p(\underset{\sim}{\mu}^{(2)}, \underset{\sim}{\Sigma})$, $\alpha = 1, 2, \ldots N_2$ V/S

H_1: $\underset{\sim}{x}_\alpha^{(1)} \sim N_p(\underset{\sim}{\mu}^{(1)}, \underset{\sim}{\Sigma})$

$\underset{\sim}{x}, \underset{\sim}{x}_\alpha^{(2)} \sim N_p(\underset{\sim}{\mu}^{(2)}, \underset{\sim}{\Sigma})$

Here, $\underset{\sim}{\mu}^{(1)}$, $\underset{\sim}{\mu}^{(2)}$ and $\underset{\sim}{\Sigma}$ are unspecified. To solve the problem by likelihood f^n we take

$$\lambda = \frac{L_0}{\max(L_0, L_1)}$$

$$\lambda = \begin{cases} 1 & \text{if } L_0 \geq L_1 \\ \frac{L_0}{L_1} & \text{if } L_0 < L_1 \end{cases} \quad \text{—(1)}$$

The point for which $\lambda = 1$, leads to the acceptance of H_0. Hence, under H_0

$$\hat{\underset{\sim}{\mu}}_{H_0}^{(1)} = \frac{N_1 \bar{\underset{\sim}{x}}^{(1)} + \underset{\sim}{x}}{N_1 + 1} \text{ and}$$

$$\hat{\underset{\sim}{\mu}}_{H_0}^{(2)} = \bar{\underset{\sim}{x}}^{(2)} \text{ and}$$

$$\hat{\underset{\sim}{\Sigma}}_{H_0} = \frac{1}{N_1 + N_2 + 1}\left[\sum_{i=1}^{2}\sum_{\alpha=1}^{N_i}(\underset{\sim}{x}_\alpha^{(i)} - \hat{\underset{\sim}{\mu}}_{H_0}^{(i)})(\underset{\sim}{x}_\alpha^{(i)} - \hat{\underset{\sim}{\mu}}_{H_0}^{(i)})' + (\underset{\sim}{x} - \hat{\underset{\sim}{\mu}}_{H_0}^{(1)})(\underset{\sim}{x} - \hat{\underset{\sim}{\mu}}_{H_0}^{(1)})'\right] \quad \text{—(2)}$$

Now, since we have

$$\sum_{\alpha=1}^{N_1}(\underset{\sim}{x}_\alpha^{(1)} - \hat{\underset{\sim}{\mu}}_{H_0}^{(1)})(\underset{\sim}{x}_\alpha^{(1)} - \hat{\underset{\sim}{\mu}}_{H_0}^{(1)})' + (\underset{\sim}{x} - \hat{\underset{\sim}{\mu}}_{H_0}^{(1)})(\underset{\sim}{x} - \hat{\underset{\sim}{\mu}}_{H_0}^{(1)})'$$

$$= \sum_{\alpha=1}^{N_1}(\underset{\sim}{x}_\alpha^{(1)} - \bar{\underset{\sim}{x}}^{(1)})(\underset{\sim}{x}_\alpha^{(1)} - \bar{\underset{\sim}{x}}^{(1)})' + N_1(\bar{\underset{\sim}{x}}^{(1)} - \hat{\underset{\sim}{\mu}}_{H_0}^{(1)})(\bar{\underset{\sim}{x}}^{(1)} - \hat{\underset{\sim}{\mu}}_{H_0}^{(1)})' + (\underset{\sim}{x} - \hat{\underset{\sim}{\mu}}_{H_0}^{(1)})(\underset{\sim}{x} - \hat{\underset{\sim}{\mu}}_{H_0}^{(1)})'$$

$$= \sum_{\alpha=1}^{N_1}(\underset{\sim}{x}_\alpha^{(1)} - \bar{\underset{\sim}{x}}^{(1)})(\underset{\sim}{x}_\alpha^{(1)} - \bar{\underset{\sim}{x}}^{(1)})' + N_1\left(\bar{\underset{\sim}{x}}^{(1)} - \left(\frac{N_1\bar{\underset{\sim}{x}}^{(1)} + \underset{\sim}{x}}{N_1+1}\right)\right)^2 + \left(\underset{\sim}{x} - \left(\frac{N_1\bar{\underset{\sim}{x}}^{(1)} + \underset{\sim}{x}}{N_1+1}\right)\right)^2$$

$$= \sum_{\alpha=1}^{N_1}(\underset{\sim}{x}_\alpha^{(1)} - \bar{\underset{\sim}{x}}^{(1)})(\underset{\sim}{x}_\alpha^{(1)} - \bar{\underset{\sim}{x}}^{(1)})' + N_1\frac{(\bar{\underset{\sim}{x}}^{(1)} - \underset{\sim}{x})^2}{(N_1+1)^2} + \frac{N_1^2(\underset{\sim}{x} - \bar{\underset{\sim}{x}}^{(1)})^2}{(N_1+1)^2}$$

$$= \sum_{\alpha=1}^{N_1}(\underset{\sim}{x}_\alpha^{(1)} - \bar{\underset{\sim}{x}}^{(1)})(\underset{\sim}{x}_\alpha^{(1)} - \bar{\underset{\sim}{x}}^{(1)})' + \frac{N_1}{(N_1+1)^2}(\bar{\underset{\sim}{x}}^{(1)} - \underset{\sim}{x})^2(1+N_1)$$

$$= \sum_{\alpha=1}^{N_1}(\underset{\sim}{x}_\alpha^{(1)} - \bar{\underset{\sim}{x}}^{(1)})(\underset{\sim}{x}_\alpha^{(1)} - \bar{\underset{\sim}{x}}^{(1)})' + \frac{N_1}{(N_1+1)}(\bar{\underset{\sim}{x}}^{(1)} - \underset{\sim}{x})^2 \quad \text{—(3)}$$

Now,

$$\sum_{\alpha=1}^{N_2}(\underset{\sim}{x}_\alpha^{(2)} - \hat{\underset{\sim}{\mu}}_{H_0}^{(2)})^2 = \sum_{\alpha=1}^{N_2}(\underset{\sim}{x}_\alpha^{(2)} - \bar{\underset{\sim}{x}}^{(2)})^2$$

$$= \sum_{\alpha=1}^{N_2}(\underset{\sim}{x}_\alpha^{(2)} - \bar{\underset{\sim}{x}}^{(2)})(\underset{\sim}{x}_\alpha^{(2)} - \bar{\underset{\sim}{x}}^{(2)})' \quad \text{—(4)}$$

From eq^{n}s (2), (3) and (4), we get

$$\hat{\underset{\sim}{\Sigma}}_{H_0} = \frac{1}{N_1 + N_2 + 1}\left[\sum_{i=1}^{2}\sum_{\alpha=1}^{N_i}(\underset{\sim}{x}_\alpha^{(i)} - \bar{\underset{\sim}{x}}^{(i)})(\underset{\sim}{x}_\alpha^{(i)} - \bar{\underset{\sim}{x}}^{(i)})' + \frac{N_1}{N_1+1}(\bar{\underset{\sim}{x}}^{(1)} - \underset{\sim}{x})^2\right]$$

$$\hat{\underset{\sim}{\Sigma}}_{H_0} = \frac{1}{N_1 + N_2 + 1}\left[\underset{\sim}{C} + \frac{N_1}{N_1+1}(\bar{\underset{\sim}{x}}^{(1)} - \underset{\sim}{x})(\bar{\underset{\sim}{x}}^{(1)} - \underset{\sim}{x})'\right]$$

where, $\underset{\sim}{C} = \sum_{i=1}^{2}\sum_{\alpha=1}^{N_i}(\underset{\sim}{x}_\alpha^{(i)} - \bar{\underset{\sim}{x}}^{(i)})(\underset{\sim}{x}_\alpha^{(i)} - \bar{\underset{\sim}{x}}^{(i)})'$ —(5)

Under H_1, the MLE's of par is given by

(28) $\hat{\mu}^{(1)}_{H_1} = \bar{x}^{(1)}, \quad \hat{\mu}^{(2)}_{H_1} = \dfrac{N_2 \bar{x}^{(2)} + x}{N_2+1}$

$$\hat{\Sigma}_{H_1} = \frac{1}{N_1+N_2+1}\left[C + \frac{N_2}{N_2+1}(\bar{x}^{(2)} - x)(\bar{x}^{(2)} - x)' \right]$$

Thus, from eq^n (1), the likelihood-ratio (6) is given by

$$\lambda = \frac{\cancel{\lambda_2 L_0}}{\cancel{\lambda_1 L_1}} = \left| \frac{\hat{\Sigma}_{H_1}}{\hat{\Sigma}_{H_0}} \right|^{\frac{N_1+N_2+1}{2}}$$

$$\Rightarrow \lambda^{(2/N_1+N_2+1)} = \left| \frac{\hat{\Sigma}_{H_1}}{\hat{\Sigma}_{H_0}} \right|$$

$$= \frac{\left| C + \frac{N_2}{N_2+1}(\bar{x}^{(2)} - x)(\bar{x}^{(2)} - x)' \right|}{\left| C + \frac{N_1}{N_1+1}(\bar{x}^{(1)} - x)(\bar{x}^{(1)} - x)' \right|}$$

$$= \frac{\left| 1 + \frac{N_2}{N_2+1}(\bar{x}^{(2)} - x) C^{-1} (\bar{x}^{(2)} - x)' \right|}{\left| 1 + \frac{N_1}{N_1+1}(\bar{x}^{(1)} - x) C^{-1} (\bar{x}^{(1)} - x)' \right|}$$

Thus, the classification procedure according to L-R test is given by

$$R_1 : \lambda^{2/N_1+N_2+1} \geq \lambda_0$$

$$R_2 : \lambda^{2/N_1+N_2+1} < \lambda_0$$

where, $\lambda_0 : P[\lambda \leq \lambda_0 | H_0] = \alpha$