Journal of the Korean Data & Information Science Society
2009, **20**(4), 757–764

한국데이터정보과학회지

Bootstrap confidence intervals for classification error rate in circular models when a block of observations is missing

Hie-Choon Chung[1] · Chien-Pai Han[2]

[1]Department of e-business, Gwangju University
[2]Department of Mathematics, University of Texas at Arlington

Received 9 May 2009, revised 17 July 2009, accepted 20 July 2009

Abstract

In discriminant analysis, we consider a special pattern which contains a block of missing observations. We assume that the two populations are equally likely and the costs of misclassification are equal. In this situation, we consider the bootstrap confidence intervals of the error rate in the circular models when the covariance matrices are equal and not equal.

Keywords: Block of missing observations, bootstrap confidence interval, circular model, error rate, linear combination classification statistic, Monte Carlo study.

1. Introduction

In discriminant analysis the problem is to classify a px1 observation X of unknown origin to one of several distinct populations using an appropriate classification rule. In this paper it will be assumed that there are two distinct populations which are multivariate normal. We also assume that the two populations are equally likely and the costs of misclassification are equal. The classification rule depends on the situation when the training samples include missing values or not. Assuming that the covariance matrices are circular, we make an appropriate transformation which reduces the circular matrices to canonical forms. The discriminant function is given when the populations are multivariate normal with different circular matrices and the linear combination statistic is used when a block of observations is missing. We consider the bootstrap confidence intervals of the error rate in the circular models when the covariance matrices are equal and not equal.

[1] Corresponding author: Associate professor, Department of e-business, Gwangju University, Gwangju 503-703, Korea. E-mail: hcc@gwangju.ac.kr
[2] Professor, Department of Mathematics, University of Texas at Arlington, TX 76019-0408, U.S.A.

2. Discriminant analysis in circular models

Let X, a px1 vector, be an observation which is known to have come from one of two multivariate normal populations. Denote the ith population by π_i which is $N(\mu_i, \Sigma_i)$ for $i = 1, 2$. We assume that Σ_i is positive definite and circular, i.e. Σ_i is of the form

$$\sigma_i^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \cdot & \rho_2 & \rho_1 \\ \rho_1 & 1 & \rho_1 & \rho_2 & \cdot & \rho_3 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 & \cdot & \rho_4 & \rho_3 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_1 & \rho_2 & \rho_3 & \rho_4 & \cdot & \rho_1 & 1 \end{bmatrix}. \tag{2.1}$$

It is given in Wise (1955) that the matrix in (2.1) can be transformed into canonical form. Thus there exists an orthogonal matrix L with (m,n)th element

$$l_{mn} = p^{-1/2} \left\{ \cos \frac{2\pi}{p}(m-1)(n-1) + \sin \frac{2\pi}{p}(m-1)(n-1) \right\}$$

such that $L'\Sigma_i L = diag(\sigma_{i1}^2, \sigma_{i2}^2, \ldots, \sigma_{ip}^2)$. Since L is independent of the elements in Σ_1 and Σ_2, the discriminant function is equivalent to that when the covariance matrices are diagonal. This is true because the discriminant function derived by the likelihood ratio procedure is invariant under any linear transformation.

2.1. Discriminant function with complete data

Since the circular matrix can be transformed into canonical form, we may let $\Sigma_i = \text{diag}(\sigma_{i1}^2, \sigma_{i2}^2, \ldots, \sigma_{ip}^2)$, $i = 1, 2$. Han (1970) derived the discriminant function by using the likelihood ratio procedure. It is proportional to

$$(X - \mu_2)'\Sigma_2^{-1}(X - \mu_2) - (X - \mu_1)'\Sigma_1^{-1}(X - \mu_1).$$

Substituting μ_i and Σ_i we obtain, apart from a constant,

$$V = \sum_{j=1}^{p} \left\{ \left(\frac{1}{\sigma_{2j}^2} - \frac{1}{\sigma_{1j}^2} \right) \left(x_j - \frac{\mu_{2j}/\sigma_{2j}^2 - \mu_{1j}/\sigma_{1j}^2}{1/\sigma_{2j}^2 - 1/\sigma_{1j}^2} \right)^2 \right\}, \tag{2.2}$$

where x_j and μ_{ij} are the jth component of X and μ_i, $i = 1, 2$, respectively. We classify X into π_1 if $V > k$ and into π_2 if $V \le k$ for some suitable choice of the constant k. To find the distribution of V, we shall assume that $\sigma_{1j}^2 > \sigma_{2j}^2$ for all j, or equivalently $\Sigma_1 - \Sigma_2$ is positive definite. Hence $1/\sigma_{2j}^2 - 1/\sigma_{1j}^2 > 0$. Let

$$Z_j = \sqrt{\frac{1}{\sigma_{2j}^2} - \frac{1}{\sigma_{1j}^2}} \left(x_j - \frac{\mu_{2j}/\sigma_{2j}^2 - \mu_{1j}/\sigma_{1j}^2}{1/\sigma_{2j}^2 - 1/\sigma_{1j}^2} \right).$$

Then $V = \sum_{j=1}^{p} Z_j^2$. When X comes from π_i, i=1 or 2, Z_j are independently distributed as $N(\zeta_{ij}, \tau_{ij}^2)$ where

$$\zeta_{ij} = \sqrt{\frac{1}{\sigma_{2j}^2} - \frac{1}{\sigma_{1j}^2}} \left(\mu_{ij} - \frac{\mu_{2j}/\sigma_{2j}^2 - \mu_{1j}/\sigma_{1j}^2}{1/\sigma_{2j}^2 - 1/\sigma_{1j}^2} \right), \qquad \tau_{ij}^2 = \sigma_{ij}^2 \left(\frac{1}{\sigma_{2j}^2} - \frac{1}{\sigma_{1j}^2} \right).$$

Therefore V is distributed as the sum of $\tau_{ij}^2 \cdot \chi_1'^2(\delta_{ij}^2)$, where $\chi_1'^2(\delta_{ij}^2)$ is a non-central χ^2 distribution with 1 degree of freedom and non-centrality parameters $\delta_{ij}^2 = \zeta_{ij}^2/\tau_{ij}$. It is not easy to obtain the distribution in a closed form. Patnaik (1949) has considered a χ^2 approximation to the distribution of the sum by fitting the first two moments. Thus the distribution may be approximated by $c_\alpha \chi_{\nu_\alpha}^2$ where

$$c_\alpha = \frac{\sum_j \tau_{ij}^4 + 2\sum_j \tau_{ij}^2 \zeta_{ij}^2}{\sum_j \tau_{ij}^2 + \sum_j \zeta_{ij}^2}, \quad \nu_\alpha = \frac{1}{c_\alpha}(\tau_{ij}^2 + \sum_j \zeta_{ij}^2).$$

2.2. Discriminant function with incomplete data

In this paper we consider a special pattern which contains a block of missing observations in circular models. Instead of estimating the parameters, we construct two different discriminant functions from the complete data and incomplete data, respectively, and then a linear combination of these two linear discriminant functions is used to obtain the classification rule.

When the populations are multivariate normal with equal covariance matrix, that is, $\pi_i : N(\mu_i, \Sigma)$, Chung and Han (2000) derived the linear combination statistic when a block of observations is missing.

Let us partition the p x 1 observation X as follows.

$$X = \begin{bmatrix} Y \\ Z \end{bmatrix},$$

where Y is a $k \times 1$ vector and Z is a $(p-k) \times 1$ vector $(1 \le k < p)$. Suppose random samples of sizes m_i, containing no missing values,

$$X_{ij} = \begin{bmatrix} Y_{ij} \\ Z_{ij} \end{bmatrix},\ i = 1, 2, ;\ j = 1, 2, \ldots, m_i, \text{are available from}$$

$$N_p(\mu_i, \Sigma) = N_p\left(\begin{bmatrix} \mu_{yi} \\ \mu_{zi} \end{bmatrix}, \begin{bmatrix} \sum_{yy} & \sum_{zy} \\ \sum_{yz} & \sum_{zz} \end{bmatrix} \right),$$

and random samples of sizes $n_i - m_i$, which contain only the first k-components $Y_{ij(kx1)}, i = 1, 2; j = m_i + 1, \ldots, n_i$, are available from $N_k(\mu_{yi}, \sum_{yy})$. We denote by $X_{ij}, i = 1, 2; j = 1, \ldots, m_i$, the complete observations, and by $Y_{ij}, i = 1, 2; j = 1, \ldots, n_i$, the incomplete observations. Hence the data have the special pattern of missing values where a block of variables is missing on $n_i - m_i$ observations, and the remaining observations are all complete.

We can construct two linear discriminant functions. The first linear discriminant function is based on the observations, $X_{ij}, i = 1, 2; j = 1, \ldots, m_i$. We have

$$W_x = (\bar{X}_1 - \bar{X}_2)' S_{xx}^{-1} \left[X - \frac{1}{2}(\bar{X}_1 - \bar{X}_2) \right],$$

where

$$\bar{X}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} X_{ij} = \begin{bmatrix} \bar{Y}_{i1} \\ \bar{Z}_i \end{bmatrix}, \quad \bar{Y}_{i1} = \frac{1}{m_i} \sum_{j=1}^{m_i} Y_{ij}, \quad \bar{Z}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} Z_{ij}, \ i = 1, 2,$$
$$S_{xx} = \sum_{i=1}^{2} \sum_{j=1}^{m_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'/\nu_x, \quad \nu_x = m_1 + m_2 - 2.$$

The second linear discriminant function is based on the incomplete observations, $\bar{Y}_{ij(kx1)}, i = 1, 2; j = 1, 2, \ldots, n_i$. We have

$$W_y = (\bar{Y}_1 - \bar{Y}_2)' S_{yy}^{-1} \left[Y - \frac{1}{2}(\bar{Y}_1 - \bar{Y}_2) \right], \text{ where } \bar{Y}_i = \frac{1}{n_i} \left[m_i \bar{Y}_{i1} + (n_i - m_i) \bar{Y}_{i2} \right],$$
$$\bar{Y}_{i2} = \frac{1}{n_i - m_i} \sum_{j=m_i+1}^{n_i} Y_{ij}, \quad S_{yy} = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)(Y_{ij} - \bar{Y}_i)'/\nu_y, \quad \nu_y = n_1 + n_2 - 1.$$

Now we combine the two linear discriminant functions and construct the classification rule which is a linear combination of W_x and W_y , namely

$$W_c = cW_x + (1 - c)W_y, \quad 0 \le c \le 1. \tag{2.3}$$

Now we consider the linear combination statistic when the populations are multivariate normal with unequal covariance matrix. When a block of observations is missing in circular models, we have the discriminant function.

$$W_c = cW_x + (1 - c)W_y, \quad 0 \le c \le 1,$$
$$W_x = (X - \bar{X}_2)' S_2^{-1} (X - \bar{X}_2) - (X - \bar{X}_1)' S_1^{-1} (X - \bar{X}_1),$$
$$W_y = (Y - \bar{Y}_2)' S_{y2} (Y - \bar{Y}_2) - (Y - \bar{Y}_1)' S_{y1}^{-1} (Y - \bar{Y}_1),$$

where

$$c = \frac{\left(\frac{1}{m_1} + \frac{1}{m_2} \right)^{-1} D^2}{\left(\frac{1}{m_1} + \frac{1}{m_2} \right)^{-1} D^2 + \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1} D_y^2},$$

$$D^2 = (\bar{X}_1 - \bar{X}_2)' S^{-1} (\bar{X}_1 - \bar{X}_2), \quad D_y^2 = (\bar{Y}_1 - \bar{Y}_2)' S_y^{-1} (\bar{Y}_1 - \bar{Y}_2),$$
$$S = \frac{S_1}{m_1} + \frac{S_2}{m_2}, \quad S_i = \frac{1}{m_i - 1} \sum_{j=1}^{m_i} (X_{ij} - \bar{X}_i),$$
$$S_y = \frac{S_{y1}}{n_1} + \frac{S_{y2}}{n_2}, \quad S_{yi} = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i).$$

Substituting $\bar{X}_i$, S_i, $\bar{Y}_i$ and S_{yi} we obtain apart from a constant,

$$V_x = \sum_{j=1}^{p} \left\{ \left(\frac{1}{s_{2j}^2} - \frac{1}{s_{1j}^2} \right) \left(x_j - \frac{\bar{x}_{2j}/s_{2j}^2 - \bar{x}_{1j}/s_{1j}^2}{1/s_{2j}^2 - 1/s_{1j}^2} \right)^2 \right\} \text{ for } W_x,$$
$$V_y = \sum_{j=1}^{k} \left\{ \left(\frac{1}{s_{y2j}^2} - \frac{1}{s_{y1j}^2} \right) \left(x_j - \frac{\bar{y}_{2j}/s_{y2j}^2 - \bar{y}_{1j}/s_{y1j}^2}{1/s_{y2j}^2 - 1/s_{y1j}^2} \right)^2 \right\} \text{ for } W_y.$$

Then $W_c = cV_x + (1-c)V_y, \ 0 \le c \le 1$.

The unconditional cdf of V_x and V_y are given (Han, 1970). The distribution of W_c is very complicated.

Now we consider the conditional distribution given the sample statistics. Given sample statistics, V_x is distributed as the sum of $\tau_{ij}^{*2} \cdot \chi_1'^2(\delta_{ij}^{*2})$, where $\chi_1'^2(\delta_{ij}^{*2})$ is a non-central distribution χ^2 with 1 degree of freedom and non-centrality parameters

$$\delta_{ij}^{*2} = \frac{\zeta_{ij}^{*2}}{\tau_{ij}^{*2}},$$

where $\zeta_{ij}^* = \sqrt{\frac{1}{s_{2j}^2} - \frac{1}{s_{1j}^2}} \left(\mu_{ij} - \frac{\bar{x}_{2j}/s_{2j}^2 - \bar{x}_{1j}/s_{1j}^2}{1/s_{2j}^2 - 1/s_{1j}^2} \right)$, $\tau_{ij}^{*2} = \sigma_{ij}^2 \left(\frac{1}{s_{2j}^2} - \frac{1}{s_{1j}^2} \right)$.

Also, given the statistics, V_y is distributed as the sum of $\tau_{yij}^{*2} \cdot \chi_1'^2(\delta_{yij}^{*2})$, where $\chi_1'^2(\delta_{yij}^2)$ has non-centrality parameters $\delta_{yij}^{*2} = \zeta_{yij}^{*2}/\tau_{yij}^{*2}$, where

$$\zeta_{yij}^* = \sqrt{\frac{1}{s_{y2j}^2} - \frac{1}{s_{y1j}^2}} \left(\mu_{yij} - \frac{\bar{y}_{2j}/s_{y2j}^2 - \bar{y}_{1j}/s_{y1j}^2}{1/s_{y2j}^2 - 1/s_{y1j}^2} \right), \quad \tau_{yij}^{*2} = \sigma_{yij}^2 \left(\frac{1}{s_{y2j}^2} - \frac{1}{s_{y1j}^2} \right).$$

The conditional probability of misclassifying an observation X from π_1 and π_2 by W_c is given by

$$\beta_{1n}^* = Pr(W_c < k | \bar{x}_{ij}, \bar{y}_{ij}, s_{ij}^2, s_{yij}^2, x, y \in \pi_1)$$
$$= Pr\left\{ \left(c \sum_{j=1}^{p} \tau_{ij}^{*2} \chi_1'^2(\delta_{ij}^{*2}) + (1-c) \sum_{j=1}^{k} \tau_{yij}^{*2} \chi_1'^2(\delta_{yij}^{*2}) \right) < k \right\}.$$

Similarly,

$$\beta_{2n}^* = Pr(W_c \ge k | \bar{x}_{ij}, \bar{y}_{ij}, s_{ij}^2, s_{yij}^2, x, y \in \pi_2)$$

$$= Pr\left\{\left(c\sum_{j=1}^{p}\tau_{ij}^{*2}\chi_1'^2(\delta_{ij}^{*2}) + (1-c)\sum_{j=1}^{k}\tau_{yij}^{*2}\chi_1'^2(\delta_{yij}^{*2})\right) \geq k\right\}.$$

Hence the conditional error rate, with equal prior probability, is defined as

$$\beta_n^* = \frac{1}{2}(\beta_{1n}^*\beta_{2n}^*). \tag{2.4}$$

We use Patnaik's method to approximate it by a constant multiple of a central chi-square distribution. The distribution of V_x may be approximated by $c_x\chi^2_{\nu_x}$, where

$$c_x = \frac{\sum_j \tau_{ij}^{*4} + 2\sum_j \tau_{ij}^{*2}\zeta_{ij}^{*2}}{\sum_j \tau_{ij}^{*2} + \sum_j \zeta_{ij}^{*2}}, \quad \nu_x = \frac{1}{c_x}(\sum_j \tau_{ij}^{*2} + \sum_j \zeta_{ij}^{*2}).$$

Also, the distribution of V_y may be approximated by $c_y\chi^2_{\nu_y}$, where

$$c_y = \frac{\sum_j \tau_{yij}^{*4} + 2\sum_j \tau_{yij}^{*2}\zeta_{yij}^{*2}}{\sum_j \tau_{yij}^{*2} + \sum_j \zeta_{yij}^{*2}}, \quad \nu_y = \frac{1}{c_y}(\sum_j \tau_{yij}^{*2} + \sum_j \zeta_{yij}^{*2}).$$

The conditional probability of misclassifying an observation X from $\pi_1$and π_2 by W_c is given by

$$\beta_{1c}^* = Pr(W_c < k|\bar{x}_{ij}, \bar{y}_{ij}, s_{ij}^2, s_{yij}^2, x, y \in \pi_1) = Pr\left\{\left(c\cdot c_x\chi^2_{\nu_x} + (1-c)c_y\chi^2_{\nu_y}\right) < k\right\},$$
$$\beta_{2c}^* = Pr(W_c < k|\bar{x}_{ij}, \bar{y}_{ij}, s_{ij}^2, s_{yij}^2, x, y \in \pi_2) = Pr\left\{\left(c\cdot c_x\chi^2_{\nu_x} + (1-c)c_y\chi^2_{\nu_y}\right) \geq k\right\}.$$

Hence the conditional error rate, with equal prior probability, is defined as

$$\beta_c^* = \frac{1}{2}(\beta_{1c}^* + \beta_{2c}^*). \tag{2.5}$$

3. Bootstrap confidence interval when training samples do not contain missing values

Let the populations be multivariate normal with equal covariance matrix, that is, $\pi_i : N(\mu_i, \Sigma), i = 1, 2$.

We now consider the bootstrap confidence interval for the conditional error rate, which is defined as $\alpha = \Phi(-\Delta/2)$, where $\Delta^2 = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$, when the training samples contain no missing values. The bootstrap method is a resampling technique using Monte Carlo simulation (Efron, 1982). In our situation, independent random samples of sizes n_1 and n_2 with replacement are taken from the two training samples respectively. An estimator $\hat{\alpha^*}$ of α based on the bootstrap sample is obtained by using $\hat{\alpha} = \Phi(-D/2)$, where $D^2 = (\bar{X}_1 - \bar{X}_2)'S^{-1}(\bar{X}_1 - \bar{X}_2)$ which is the Mahalanobis squared distance Δ^2. This process is repeated independently a large number B of times. Then bootstrap confidence interval for α can be obtained from the B values of $\hat{\alpha}^*$. Let $\hat{\alpha}^*_{(i)}$ denote the i-th ordered value, so that $\hat{\alpha}^*_{(1)} \leq \hat{\alpha}^*_{(2)} \leq \cdots \leq \hat{\alpha}^*_{(B)}$.

There are several methods to construct the bootstrap confidence interval. We will consider the percentile method, bias-corrected percentile method, accelerated bias-corrected percentile method to construct the confidence interval (Efron, 1982, 1987; Buckland, 1983, 1984, 1985; Hall, 1986a, 1986b; Hinkley, 1988; DiCiccio and Romano, 1988; among others). These three types of $100(1-2\eta)\%$ confidence interval are presented as follows:

Percentile method. The confidence interval is given by $(\hat{\alpha}^*_{(r)}, \hat{\alpha}^*_{(s)})$, where $r = (B+1)\eta$, and $s = (B+1)(1-\eta)$, both rounded to nearest integer, subject to $r + s = B + 1$.

Bias-corrected percentile method. Suppose $\hat{\alpha}^*_{(q)} < \hat{\alpha} < \hat{\alpha}^*_{(q+1)}$, where $\hat{\alpha}$ is calculated from the original samples. That is, q of the B bootstrap estimates for α are smaller than $\hat{\alpha}$. Define

$$z_o = \Phi^{-1}(q/B),\ \eta_{BL} = \Phi(2z_o - z_\eta) \text{ and } \eta_{BR} = \Phi(2z_o + z_\eta),$$

where $\Phi(z_\eta) = 1 - \eta$ and Φ denotes the cumulative standard normal distribution. Then the confidence interval is given by $(\hat{\alpha}^*_{(j)}, \hat{\alpha}^*_{(k)})$, where $j = (B+1)\eta_{BL}$ and $k = (B+1)\eta_{BR}$.

Accelerated bias-corrected percentile method. Define

$$\eta_{AL} = \Phi(z_o + \frac{z_o - z_\eta}{1 - a(z_o - z_\eta)}),\ \text{and } \eta_{AR} = \Phi(z_o + \frac{z_o + z_\eta}{1 - a(z_o + z_\eta)}),$$

where

$$a = \frac{1}{6}\left[\sum_{i=1}^{B}(\hat{\alpha}^*_i - \bar{\hat{\alpha}}^*)^3 \Big/ \left[\sum_{i=1}^{B}(\hat{\alpha}^*_i - \bar{\hat{\alpha}}^*)^2\right]^{\frac{3}{2}}\right],$$

which is called the acceleration constant, and $\bar{\hat{\alpha}}^*$ is the mean of the B bootstrap estimates for $\hat{\alpha}^*_i, i = 1, \ldots, B$.

Then the confidence interval is given by $(\hat{\alpha}^*_{(u)}, \hat{\alpha}^*_{(v)})$, where $u = (B+1)\eta_{AL}$ and $v = (B+1)\eta_{AR}$.

Note that η_{AR} and η_{AL} become η_{BR} and η_{BL} if a equals 0. If z_o is zero, then η_{BR} and η_{BL} become η.

We evaluate the bootstrap confidence intervals for the conditional error rate (Chung and Han, 2000) when the training samples come from the circular models in which the covariance matrix has the equal covariance matrix.

4. Bootstrap confidence interval when training samples contain missing values

First, when the covariance matrices are equal, we consider the bootstrap confidence intervals for the conditional error rates using W_c (Chung and Han, 2000) in circular models.

Also we will consider the bootstrap confidence interval for the conditional error rate β^*_n and β^*_c in (2.4) and (2.5) using W_c. The conditional error rate can be estimated by substituting the estimate $\hat{\sigma}^2_{ij}$, $\hat{\mu}_{ij}$, $\hat{\sigma}^2_{yij}$ and $\hat{\mu}_{yij}$for σ^2_{ij}, μ_{ij}, σ^2_{yij} and μ_{yij} in (2.4) and (2.5) respectively. Let $\hat{\mu}_i = [\bar{Y}^{(i)}, \bar{Z}^{(i)}]'$ and $\hat{\mu}_{yi} = \bar{Y}^{(i)}$ be the estimate of μ_i and μ_{iy} from (2.4) and (2.5). For the variances, let $\hat{\sigma}^2_{cij} = \sum_{l=1}^{m_i}(x_{ijl} - \bar{x}_{ij})^2/(m_i - 1), i = 1, 2, j =$

$1, 2, \cdots, p$, be the estimates from the complete observations of sizes m_i. Also let $\hat{\sigma}^2_{Iij} = \sum_{l=m_i+1}^{n_i} (y_{ijl} - \bar{y}_{ij})^2/(n_i - m_i)$ be the estimates from the incomplete observations of sizes $n_i - m_i$ using only Y observations, $i = 1, 2, j = 1, 2, \ldots, k$. Then for σ^2_{ij}, we suggest the combined estimates $\hat{\sigma}^2_{ij} = m_i \hat{\sigma}^2_{cij}/n_i + (n_i - m_i)\hat{\sigma}^2_{Iij}/n_i$, $i = 1, 2, j = 1, 2, \ldots, k$, and $\hat{\sigma}^2_{ij} = \hat{\sigma}^2_{cij}, i = 1, 2, j = k + 1, \ldots, p$. For σ^2_{yij}, we use $\hat{\sigma}^2_{ij}, i = 1, 2, j = 1, 2, \ldots, k$.

We will use these estimates in the construction of the bootstrap confidence intervals for the conditional error rate when the training samples contain missing observations. Basically the same procedure described for α is applied in this situation for getting the three types of the bootstrap confidence intervals, i.e., the percentile method, the bias-corrected percentile method, and the accelerated bias-corrected method. In order to evaluate the properties of the confidence intervals, we conduct a similar Monte Carlo study described for the optimal error rate.

References

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*, John Wiley and Sons, New York.

Buckland, S. T. (1983). Monte Carlo methods for confidence interval estimation using the bootstrap technique. *Bias,* **10**, 194-212.

Buckland, S. T. (1984). Monte Carlo confidence intervals. *Biometrics,* **40**, 811-817.

Buckland, S. T. (1985). Calculation of Monte Carlo confidence intervals. *Royal Statistical Society*, **Algorithm AS214**, 297-301.

Chung, H. C. and Han, C. P. (2000). Discriminant analysis when a block of observations is missing. *Annals of the Institute of Statistical Mathematics*, **52**, 544-556.

Cho, J. J., Han, J. H. and Lee, I. P. (1999). Better bootstrap confidence intervals for process incapability index. *Journal of the Korean Data & Information Science Society*, **10**, 341-357.

Dempster, A. P., Laird, N. M. and Rubin, R. J. A. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 302-306.

DiCiccio, T. J. and Romano, J. P. (1988). A review of bootstrap confidence intervals. *Journal of the Royal Statistical Society, Series B,* **50**, 338-354.

Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. *CBMS-NSF Regional Conference Series in Applied Mathematics,* **38**, *Society for Industrial and Applied Mathematics(SIAM),* Philadelphia.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association,* **82**, 171-200.

Hall, P. (1986a). On the bootstrap and confidence intervals. *Annals of Statistics,* **14**, 1431-1452.

Hall, P. (1986b). On the number of bootstrap simulations required to construct a confidence interval. *Annals of Statistics*, **14**, 1453-1462.

Han, C. P. (1970). Distribution of discriminant function in circular models. *Annals of the Institute of statistical Mathematics*, **22,** 117-125.

Hinkley, D. V. (1988). Bootstrap methods. *Journal of the Royal Statistical Society, Ser. B*, **50**, 321-337.

Patnaik, D. B. (1949). The non-central and F distributions and their applications. *Biometrika*, **36**, 202-232.

Wise, J. (1955). The autocorrelation function and the spectral density function. *Biometrika*, **42**, 151-159.