Ann. Inst. Statist. Math.
38 (1986), Part A, 529-538

ON THE ERRORS OF MISCLASSIFICATION BASED ON DICHOTOMOUS AND NORMAL VARIABLES

N. BALAKRISHNAN, S. KOCHERLAKOTA AND K. KOCHERLAKOTA

(Received Aug. 2, 1985; revised Nov. 6, 1985)

Summary

The distribution of the errors of misclassification in procedures based on dichotomous and normal variables is derived. The expressions for $E(e_{12})$ and $E(e_{21})$ are also obtained. The results in the paper extend those of Chang and Afifi (1974, *J. Amer. Statist. Ass.*, **69**, 336–339), using the earlier papers due to John (1961, *Ann. Math. Statist.*, **32**, 1125–1144), Subrahmaniam and Chinganda (1978, *J. Statist. Plann. Inf.*, **2**, 79–91).

1. Preliminaries

Let us consider two random variables X and Y where X is dichotomous with the probability function

$$\mathrm{P}\{X=x\}=\theta^x(1-\theta)^{1-x}, \qquad x=0,1$$

and Y, conditional on $X=x$, has the normal distribution $N(\mu^{(x)}, \sigma^{(x)^2})$. We assume that

$$\mu^{(x)}=\mu+\delta x, \qquad \sigma^{(x)^2}=\sigma^2+\gamma^2 x .$$

Then the vector $\boldsymbol{W}'=(X, Y)$ has the probability density function

$$f(\boldsymbol{w})=\theta^x(1-\theta)^{1-x}\left[\frac{1}{\sqrt{2\pi}\,\sigma^{(x)}}\exp-\frac{1}{2\sigma^{(x)^2}}\{y-\mu^{(x)}\}^2\right], \tag{1.1}$$

$$x=0,1;\ -\infty<y<\infty .$$

It can be seen directly that

$$E(\boldsymbol{W})=\begin{pmatrix}\theta\\ \mu+\delta\theta\end{pmatrix} \tag{1.2}$$

Key words and phrases: Errors of misclassification, dichotomous variables, normal distribution, distribution of the errors of misclassification.

and

$$\operatorname{Var}(\boldsymbol{W})=\begin{pmatrix}\theta(1-\theta) & \theta(1-\theta)\delta \\ \cdot\ \cdot & \sigma^2+\gamma^2\theta+\theta(1-\theta)\delta^2\end{pmatrix} .$$

The marginal of Y can be determined from (1.1) to be

$$g(y)=\theta\left\{\frac{1}{\sqrt{2\pi(\sigma^2+\gamma^2)}}\exp-\frac{1}{2(\sigma^2+\gamma^2)}(y-\mu-\delta)^2\right\} \\ +(1-\theta)\left\{\frac{1}{\sqrt{2\pi}\,\sigma}\exp-\frac{1}{2\sigma^2}(y-\mu)^2\right\} , \tag{1.3}$$

a mixture of two normals.

The problem we examine here is concerned with classification of an observation $\boldsymbol{w}$ into one of the two populations Π_1, Π_2 where Π_i has the pdf of the form

$$f_i(\boldsymbol{w})=\theta_i^x(1-\theta_i)^{1-x}\phi(y\,;\,\mu_i^{(x)},\,\sigma^{(x)^2}) , \tag{1.4}$$

for $i=1, 2$. It should be noted that we are assuming here that the conditional variances of Y under Π_1 and Π_2 are equal. However, the $\operatorname{Var}(\boldsymbol{W}|\Pi_1)$ and $\operatorname{Var}(\boldsymbol{W}|\Pi_2)$ are not equal.

Such a model has been examined by Chang and Afifi [2]. However, they have not considered the errors of misclassification. The reader is referred for related work on the distribution of the misclassification errors to Chinganda and Subrahmaniam [3], Subrahmaniam and Chinganda [6], John [5], among others. For a discussion of the models involving discrete random variables, reference may be made to Goldstein and Dillon [4].

2. Classification: All parameters known

The log-likelihood ratio gives the following procedure for the observation $\boldsymbol{w}$. Classify it as belonging to Π_1 or Π_2 according as

$$\left\{y-\frac{1}{2}(\mu_1^{(x)}+\mu_2^{(x)})\right\}\frac{\mu_1^{(x)}-\mu_2^{(x)}}{\sigma^{(x)^2}}+x\ln\left(\frac{\theta_1}{\theta_2}\right)+(1-x)\ln\left(\frac{1-\theta_1}{1-\theta_2}\right)\gtrless 0 . \tag{2.1}$$

2.1. *Errors of misclassification*

There are two types of errors of misclassification with any classification procedure. Their probabilities are

$$e_{12}=\Pr[\text{misclassifying } \boldsymbol{w} \text{ into } \Pi_2 \text{ when in fact } \boldsymbol{w} \text{ is from } \Pi_1]$$

and

$$e_{21}=\Pr[\text{misclassifying } \boldsymbol{w} \text{ into } \Pi_1 \text{ when in fact it is from } \Pi_2] .$$

Since there is no difference in the techniques involved in their determination, we shall present details for e_{12}. Thus, writing $Z(x)$ for the quantity on the left-hand side of the inequality in (2.1),

$$\begin{aligned} e_{12} &= \Pr\left[Z(x)<0 \mid \boldsymbol{w}\in\Pi_1\right] \\ &= (1-\theta_1)\Pr\left[\left\{y-\frac{\mu_1^{(0)}+\mu_2^{(0)}}{2}\right\}\frac{(\mu_1^{(0)}-\mu_2^{(0)})}{\sigma^{(0)^2}}+\ln\left(\frac{1-\theta_1}{1-\theta_2}\right)<0 \,\middle|\, y\sim N(\mu_1^{(0)},\sigma^{(0)^2})\right] \\ &\quad +\theta_1\Pr\left[\left\{y-\frac{\mu_1^{(1)}+\mu_2^{(1)}}{2}\right\}\frac{(\mu_1^{(1)}-\mu_2^{(1)})}{\sigma^{(1)^2}}+\ln\left(\frac{\theta_1}{\theta_2}\right)<0 \,\middle|\, y\sim N(\mu_1^{(1)},\sigma^{(1)^2})\right]. \end{aligned}$$

Algebraic manipulation yields

$$\text{(2.2)}\qquad \begin{aligned} e_{12} &= (1-\theta_1)\Phi\left\{-\frac{1}{2}\sqrt{\alpha^{(0)}}-\frac{1}{\sqrt{\alpha^{(0)}}}\ln\frac{(1-\theta_1)}{(1-\theta_2)}\right\} \\ &\quad +\theta_1\Phi\left\{-\frac{1}{2}\sqrt{\alpha^{(1)}}-\frac{1}{\sqrt{\alpha^{(1)}}}\ln\left(\frac{\theta_1}{\theta_2}\right)\right\} \end{aligned}$$

where $\Phi(\cdot)$ stands for the distribution function of the standard normal distribution and, for $i=0, 1$,

$$\alpha^{(i)}=\frac{(\mu_1^{(i)}-\mu_2^{(i)})^2}{\sigma^{(i)^2}} .$$

Similarly, the expression for e_{21} may be derived.

3. Classification when the parameters are unknown

Consider independent random samples of sizes n_1, n_2 from each of the populations. Denote by $y_{ij}^{(x)}$ the j-th observation from the i-th population, of which the value of X is x ($x=0, 1$). Let $n_i=n_i^{(0)}+n_i^{(1)}$ and

$$\bar{y}_i^{(x)}=\frac{1}{n_i^{(x)}}\sum_{j=1}^{n_i^{(x)}} y_{ij}^{(x)} ,$$

$$s^{(x)^2}=\frac{1}{n_1^{(x)}+n_2^{(x)}}\left\{\sum_{j=1}^{n_1^{(x)}}(y_{1j}^{(x)}-\bar{y}_1^{(x)})^2+\sum_{j=1}^{n_2^{(x)}}(y_{2j}^{(x)}-\bar{y}_2^{(x)})^2\right\} .$$

It can be shown that the maximum likelihood estimates of the parameters are, for $i=1, 2$,

$$\hat{\theta}_i=n_i^{(1)}/n_i , \qquad \hat{\mu}_i^{(x)}=\bar{y}_i^{(x)} , \qquad \hat{\sigma}^{(x)^2}=s^{(x)^2} .$$

The likelihood classification rule is in this case to classify an observation w as belonging to Π_1 or Π_2 according as

$$Z(x)=\left\{y-\frac{1}{2}(\bar{y}_1^{(x)}+\bar{y}_2^{(x)})\right\}\frac{(\bar{y}_1^{(x)}-\bar{y}_2^{(x)})}{s^{(x)^2}}+\ln\left\{\frac{n_1^{(x)}n_2}{n_2^{(x)}n_1}\right\}\gtrless 0 .$$

For simplicity, in order to discuss the errors of misclassification, we shall assume that $\theta_1=\theta_2$. In this case, the last term in $Z(x)$ is zero. The classification statistic can be reduced to

$$Z(x)=\left\{y-\frac{1}{2}(\bar{y}_1^{(x)}+\bar{y}_2^{(x)})\right\}(\bar{y}_1^{(x)}-\bar{y}_2^{(x)}) .$$

3.1. *Errors of misclassification*

Defining the errors of misclassification as usual, it can be seen that

$$e_{12}=\begin{cases}\Phi\left\{\left[\dfrac{\bar{y}_1^{(x)}+\bar{y}_2^{(x)}}{2}-\mu_1^{(x)}\right]\Big/\sigma^{(x)}\right\} & \text{if } \bar{y}_1^{(x)}>\bar{y}_2^{(x)} \\ \Phi\left\{\left[\mu_1^{(x)}-\dfrac{\bar{y}_1^{(x)}+\bar{y}_2^{(x)}}{2}\right]\Big/\sigma^{(x)}\right\} & \text{if } \bar{y}_1^{(x)}<\bar{y}_2^{(x)} .\end{cases} \tag{3.1}$$

and a similar expression is obtained for e_{21}.

3.2. *The distribution function of* e_{12}

Let

$$G(z)=\Pr\{e_{12}\leqq z\} .$$

Then from (3.1), we have

$$\begin{aligned}G(z)=&(1-\theta)\left\{\Pr\left[\Phi\left\{\left[\frac{1}{2}(\bar{y}_1^{(0)}+\bar{y}_2^{(0)})-\mu_1^{(0)}\right]\Big/\sigma^{(0)}\right\}\leqq z ,\right.\right.\\ &\left.\left.\bar{y}_1^{(0)}-\bar{y}_2^{(0)}>0\right]\right\}\\ &+\theta\left\{\Pr\left[\Phi\left\{\left[\frac{1}{2}(\bar{y}_1^{(1)}+\bar{y}_2^{(1)})-\mu_1^{(1)}\right]\Big/\sigma^{(1)}\right\}\leqq z ,\right.\right.\\ &\left.\left.\bar{y}_1^{(1)}-\bar{y}_2^{(1)}>0\right]\right\} .\end{aligned} \tag{3.2}$$

Writing

$$u^{(i)}=\bar{y}_1^{(i)}+\bar{y}_2^{(i)} , \qquad v^{(i)}=\bar{y}_2^{(i)}-\bar{y}_1^{(i)} ,$$

the expressions on the right-hand side of (3.2) can be written in terms of these random variables. Hence

$$\begin{aligned}G(z)=&(1-\theta)[\Pr\{u^{(0)}\leqq k_1^{(0)}, v^{(0)}<0\}+\Pr\{u^{(0)}\geqq k_2^{(0)}, v^{(0)}>0\}]\\ &+\theta[\Pr\{u^{(1)}\leqq k_1^{(1)}, v^{(1)}<0\}+\Pr\{u^{(1)}\geqq k_2^{(1)}, v^{(1)}>0\}] ,\end{aligned} \tag{3.3}$$

where

$$k_1^{(i)}=2\{\mu_1^{(i)}+\sigma^{(i)}\Phi^{-1}(z)\}\ , \qquad k_2^{(i)}=2\{\mu_1^{(i)}-\sigma^{(i)}\Phi^{-1}(z)\}$$

for $i=0, 1$. Since $(u^{(i)}, v^{(i)})$ are jointly BVN $[\mu_u^{(i)}, \mu_v^{(i)}; \sigma_u^{(i)^2}, \sigma_v^{(i)^2}, \rho_{uv}^{(i)}]$ where, for $i=1, 2$,

$$\mu_u^{(i)}=\mu_1^{(i)}+\mu_2^{(i)}\ , \qquad \mu_v^{(i)}=\mu_2^{(i)}-\mu_1^{(i)}\ ,$$

$$\sigma_u^{(i)^2}=\sigma_v^{(i)^2}=\sigma^{(i)^2}\left(\frac{1}{n_1^{(i)}}+\frac{1}{n_2^{(i)}}\right)$$

and

$$\rho_{uv}^{(i)}=(n_1^{(i)}-n_2^{(i)})/(n_1^{(i)}+n_2^{(i)})\ .$$

If $H(z_1, z_2; \rho)$ denotes the distribution function in the standard BVN distribution with the coefficient of correlation ρ, then (3.3) is

$$\text{(3.4)}\quad G(z)=(1-\theta)\left[H\left(\frac{k_1^{(0)}-\mu_u^{(0)}}{\sigma_u^{(0)}}, \frac{-\mu_v^{(0)}}{\sigma_v^{(0)}}; \rho_{uv}^{(0)}\right)+H\left(\frac{\mu_u^{(0)}-k_2^{(0)}}{\sigma_u^{(0)}}, \frac{\mu_v^{(0)}}{\sigma_v^{(0)}}; \rho_{uv}^{(0)}\right)\right]$$
$$+\theta\left[H\left(\frac{k_1^{(1)}-\mu_u^{(1)}}{\sigma_u^{(1)}}, \frac{-\mu_v^{(1)}}{\sigma_v^{(1)}}; \rho_{uv}^{(1)}\right)+H\left(\frac{\mu_u^{(1)}-k_2^{(1)}}{\sigma_u^{(1)}}, \frac{\mu_v^{(1)}}{\sigma_v^{(1)}}; \rho_{uv}^{(1)}\right)\right]\ .$$

From (3.4), the density function of e_{12} is obtained by directly differentiating the distribution function. Thus

$$\text{(3.5)}\quad g(z)=2\sum_{i=0}^{1}\theta^i(1-\theta)^{1-i}\frac{\sigma^{(i)}}{\sigma_u^{(i)}}\left[\Phi((-1)C_1^{(i)})\left(\phi\left(\frac{k_1^{(i)}-\mu_u^{(i)}}{\sigma_u^{(i)}}\right)\Big/\phi\left(\frac{k_1^{(i)}-2\mu_1^{(i)}}{2\sigma^{(i)}}\right)\right)\right.$$
$$\left.+\Phi(C_2^{(i)})\left(\phi\left(\frac{k_2^{(i)}-\mu_u^{(i)}}{\sigma_u^{(i)}}\right)\Big/\phi\left(\frac{k_2^{(i)}-2\mu_1^{(i)}}{2\sigma^{(i)}}\right)\right)\right]\ ,$$

where

$$C_j^{(i)}=\left\{\frac{\mu_v^{(i)}}{\sigma_v^{(i)}}+\rho_{uv}^{(i)}\left(\frac{k_j^{(i)}-\mu_u^{(i)}}{\sigma_u^{(i)}}\right)\right\}\Big/(1-\rho_{uv}^{(i)^2})^{1/2}\ .$$

A similar expression can be derived for e_{21} in each of the cases of the distribution function and the density function.

4. Expected values of the error rates

The probabilities of misclassification have been found earlier. Thus

$$\text{(4.1)}\quad \begin{aligned} e_{12}&=\Pr\{2Y>\bar{y}_1^{(x)}+\bar{y}_2^{(x)}\,|\,\bar{y}_1^{(x)}, \bar{y}_2^{(x)}\} && \text{if } \bar{y}_1^{(x)}<\bar{y}_2^{(x)} \\ &=\Pr\{2Y<\bar{y}_1^{(x)}+\bar{y}_2^{(x)}\,|\,\bar{y}_1^{(x)}, \bar{y}_2^{(x)}\} && \text{if } \bar{y}_1^{(x)}>\bar{y}_2^{(x)}\ . \end{aligned}$$

Hence the expected value of e_{12} (also called the unconditional probability of misclassification) may be computed as

$$E[e_{12}]=\sum_{i=0}^{1}\theta^i(1-\theta)^{1-i}\Pr[\bar{y}_1^{(i)}<\bar{y}_2^{(i)},\ 2Y>\bar{y}_1^{(i)}+\bar{y}_2^{(i)}]$$
$$+\sum_{i=0}^{1}\theta^i(1-\theta)^{1-i}\Pr[\bar{y}_1^{(i)}>\bar{y}_2^{(i)},\ 2Y<\bar{y}_1^{(i)}+\bar{y}_2^{(i)}]\ . \tag{4.2}$$

We note that in (4.2) (X, Y) is assumed to be from Π_1. Also, the pdf of Y depends upon the value of i, the conditioning random variable. Thus, for $i=0, 1$,

$$f(y\,|\,i)=\frac{1}{\sqrt{2\pi}\,\sigma^{(i)}}\exp\left\{-\frac{(y-\mu_1^{(i)})^2}{2\sigma^{(i)^2}}\right\}\ .$$

In evaluating the four terms of (4.2) we have to use the appropriate density for Y. Thus, setting $i=0$ in the first term on the right-hand side, it can be seen that

$$\Pr\{\bar{y}_1^{(0)}<\bar{y}_2^{(0)},\ \bar{y}_1^{(0)}+\bar{y}_2^{(0)}<2Y\}$$
$$=\Phi\left[\frac{-\mu_v^{(0)}}{(4\sigma^{(0)^2}+\sigma_u^{(0)^2})^{1/2}}\right]-H\left(\frac{-\mu_v^{(0)}}{(4\sigma^{(0)^2}+\sigma_u^{(0)^2})^{1/2}},\ -\frac{\mu_v^{(0)}}{\sigma_v^{(0)}};\ \frac{\sigma_u^{(0)}\rho_{uv}^{(0)}}{(4\sigma^{(0)^2}+\sigma_u^{(0)^2})^{1/2}}\right)\ .$$

Using this and related results the expected value of e_{12} can be seen to be

$$E[e_{12}]=\sum_{i=0}^{1}\theta^i(1-\theta)^{1-i}\left[\Phi\left\{-\frac{\mu_v^{(i)}}{\sigma_v^{(i)}}\right\}+\Phi\left\{\frac{-\mu_v^{(i)}}{(4\sigma^{(i)^2}+\sigma_u^{(i)^2})^{1/2}}\right\}\right.$$
$$\left.-2H\left(\frac{-\mu_v^{(i)}}{(4\sigma^{(i)^2}+\sigma_u^{(i)^2})^{1/2}},\ -\frac{\mu_v^{(i)}}{\sigma_v^{(i)}};\ \frac{\sigma_u^{(i)}\rho_{uv}^{(i)}}{(4\sigma^{(i)^2}+\sigma_u^{(i)^2})^{1/2}}\right)\right]. \tag{4.3}$$

In (4.3), we note that, for $i=0, 1$,

$$\mu_u^{(i)}=\mu_1^{(i)}+\mu_2^{(i)}\ ,\qquad \mu_v^{(i)}=\mu_2^{(i)}-\mu_1^{(i)}\ ,$$
$$\sigma_u^{(i)^2}=\sigma_v^{(i)^2}=\sigma^{(i)^2}\left(\frac{1}{n_1^{(i)}}+\frac{1}{n_2^{(i)}}\right)\ ,$$
$$\rho_{uv}^{(i)}=\frac{n_1^{(i)}-n_2^{(i)}}{n_1^{(i)}+n_2^{(i)}}\ ,$$

as derived in an earlier section.

A similar expression may be derived for the expectation of e_{21}.

5. Numerical results

It would be appropriate to recall the parametric structure of the population under consideration. Under the model introduced, the parameters of the conditional distribution of Y given $X=x$ are:

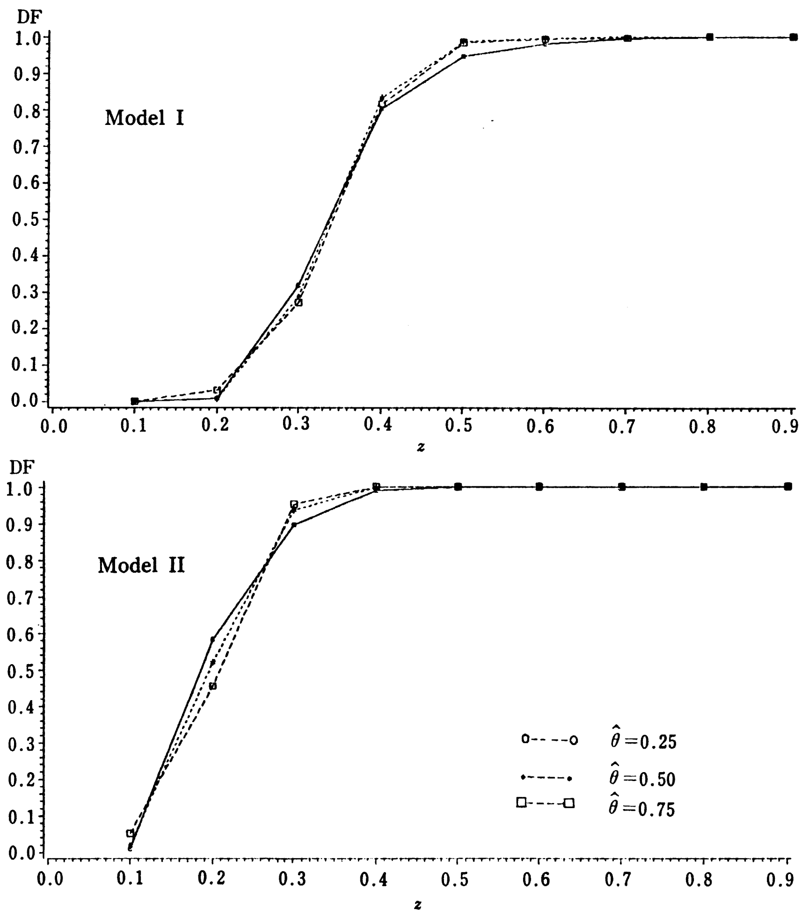
	Population I	Population II
$x=0$	μ_1 ; σ^2	μ_2 ; σ^2
$x=1$	$\mu_1+\delta$; $\sigma^2+\gamma^2$	$\mu_2+\delta$; $\sigma^2+\gamma^2$

This ensures that for either value of the conditioning variable X, the disturbances in the populations are through the means.

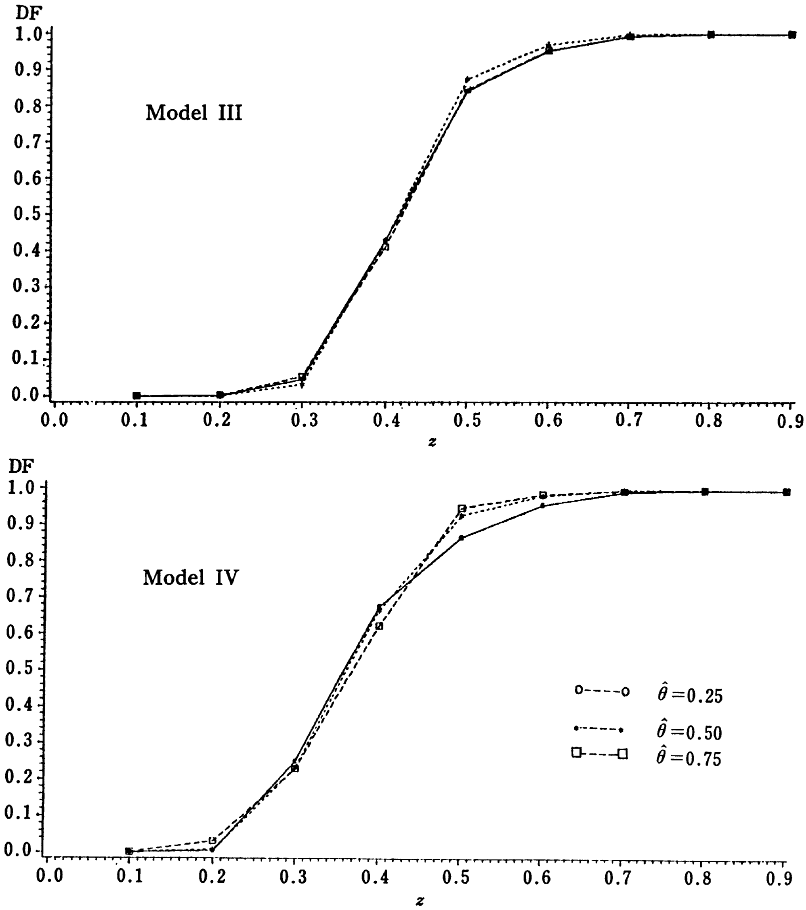
Four parameter combinations are considered in the tables:

Model I $\mu_1=0,\ \mu_2=1,\ \delta=0.5;\ \sigma^2=1,\ \gamma^2=1$
Model II $\mu_1=0,\ \mu_2=2,\ \delta=0.5;\ \sigma^2=1,\ \gamma^2=1$
Model III $\mu_1=0,\ \mu_2=1,\ \delta=0.5;\ \sigma^2=4,\ \gamma^2=1$
Model IV $\mu_1=0,\ \mu_2=1,\ \delta=0.5;\ \sigma^2=1,\ \gamma^2=4.$

Model I is essentially the standard model with mean of Π_2 shifted to the right. Model II represents a larger shift in the mean of Π_2. On the other hand, Model III introduces a larger variance (both when $x=0$ and $x=1$). Finally, Model IV is similar to Model I with the ex-



Figures I and II. Distribution Function of the Error Rate e_{12} for $n_1=n_2=40$.



Figures III and IV. Distribution Function of the Error Rate e_{12} for $n_1=n_2=40$.

ception that the conditional variance of Y is larger under $x=1$ than that under $x=0$. It should be noted that the results obtained are conditional on $n_i^{(x)}$.

Figures I–IV give the distribution function $G(z)$ of e_{12} under the four models for $n_1=n_2=40$ and $\theta=0.5$. Tables of $G(z)$ for these and a wider selection of θ and n are given in Balakrishnan et al. [1]. In Table I the expected value of e_{12} under the four models is given. In order to examine the effects of estimation of θ, three estimated values of θ are used for different sample sizes.

The results of the numerical studies can be discussed in terms of behaviour of $G(z)$.

(i) A shift in the mean makes the distribution function $G(z)$ rise up more steeply. This implies that e_{12} remains small with a higher probability.

(ii) As the sample size increases, the distribution function ascends

Table I. $E(e_{12})=E(e_{21})$ under Models I-IV ($n_1=n_2=n$)

n	θ	$\hat{\theta}$	Model I	Model II	Model III	Model IV
20	0.25	0.15	0.3391	0.1908	0.4291	0.3503
		0.25	0.3351	0.1861	0.4286	0.3484
		0.35	0.3332	0.1847	0.4293	0.3476
	0.50	0.25	0.3577	0.2095	0.4378	0.3843
		0.50	0.3490	0.2050	0.4350	0.3793
		0.75	0.3544	0.2070	0.4392	0.3849
	0.75	0.65	0.3620	0.2247	0.4359	0.4079
		0.75	0.3628	0.2252	0.4357	0.4085
		0.85	0.3671	0.2277	0.4366	0.4124
40	0.25	0.15	0.3315	0.1833	0.4174	0.3454
		0.25	0.3276	0.1819	0.4157	0.3428
		0.35	0.3258	0.1816	0.4153	0.3411
	0.50	0.25	0.3452	0.2031	0.4241	0.3755
		0.50	0.3390	0.2020	0.4200	0.3687
		0.75	0.3407	0.2031	0.4243	0.3685
	0.75	0.65	0.3517	0.2221	0.4215	0.3946
		0.75	0.3522	0.2223	0.4219	0.3940
		0.85	0.3552	0.2231	0.4239	0.3961
100	0.25	0.15	0.3246	0.1804	0.4093	0.3398
		0.25	0.3231	0.1801	0.4070	0.3374
		0.35	0.3227	0.1800	0.4060	0.3363
	0.50	0.25	0.3370	0.2007	0.4121	0.3657
		0.50	0.3360	0.2004	0.4086	0.3618
		0.75	0.3363	0.2008	0.4111	0.3614
	0.75	0.65	0.3492	0.2205	0.4110	0.3871
		0.75	0.3493	0.2206	0.4117	0.3869
		0.85	0.3498	0.2210	0.4139	0.3872

rather steeply. Such a behaviour is to be expected. This indicates that e_{12} is large (with a high probability) when n is small while it is small (with high probability) when n is large.

(iii) There doesn't seem to be a marked effect of estimation of θ. This is seen by the distribution function remaining relatively stable over the different estimates of θ. The table of expected values of e_{12} (Table I) further reflects this pattern.

(iv) An examination of Table I shows that $E(e_{12})$ is smaller when the two populations are far apart. Other factors such as sample size, variances do not seem to play any role in this regard. Nor is $E(e_{12})$ affected by the value of the estimate of θ.

Acknowledgements

NSERC is to be thanked for financial support of this research.

UNIVERSITY OF MANITOBA

REFERENCES

[1] Balakrishnan, N., Kocherlakota, S. and Kocherlakota, K. (1985). On the errors of misclassification based on dichotomous and normal variables, *Technical Report*, 163, Department of Statistics, University of Manitoba.

[2] Chang, P. C. and Afifi, A. A. (1974). Classification based on dichotomous and continuous variables, *J. Amer. Statist. Ass.*, **69**, 336-339.

[3] Chinganda, E. F. and Subrahmaniam, K. (1979). Robustness of the linear discriminant function to nonnormality: Johnson's system, *J. Statist. Plann. Inf.*, **3**, 69-77.

[4] Goldstein, M. and Dillon, W. R. (1978). *Discrete Discriminant Analysis*, John Wiley, New York.

[5] John, S. (1961). Errors in discrimination, *Ann. Math. Statist.*, **32**, 1125-1144.

[6] Subrahmaniam, K. and Chinganda, E. F. (1978). Robustness of the linear discriminant function to nonnormality: Edgeworth series distribution, *J. Statist. Plann. Inf.*, **2**, 79-91.