INT. J. SYSTEMS SCI., 1982, VOL. 13, NO. 5, 559–570

# Bayes' error and its sensitivity in statistical pattern recognition in noisy environment

B. B. CHAUDHURI†

This work is concerned mainly with statistical pattern recognition in noisy environment. Analytical expressions have been found for misclassification probability under Bayes' rule for multivariate case under certain conditions of noise statistics. The case when noise density is normal has been considered in detail and the properties have been studied with numerical results. The sensitivity of misclassification probability has been formulated to study the effect of small inaccuracy in parameter estimation on misclassification probability both for ideal and noisy environment and numerical result presented for both cases. Also, the problems of unsupervised learning and recognition, e.g. clustering, has been discussed for noisy environment. The work is useful and important in practical pattern recognition problems.

## 1. Introduction

One of the purposes of statistical pattern classification is to find the class membership of a sample from a set of measurements made on the sample. Assuming that the feature selection is complete, the object may be fulfilled in two stages: (1) classifier modelling and training, whereby class feature statistics are estimated using a set of prototypes or training samples and decision criteria are optimally chosen; and, (2) classification with the trained classifier, which accepts the feature measurements of the test sample and classifies it to one of the possible classes, as shown in the block diagram of Fig. 1. If the statistics estimation is perfect and the measurements are error-free, Bayes' decision rule provides the minimum misclassification error probability.

Bayes' rule, however, has found wider applications in statistical pattern recognition problems. Its usefulness in parameter estimation, error analysis, feature selection and hypothesis testing is well documented (Andrews 1972, Fukunaga 1972) and is still a subject of research. Some of the recent work includes the application of Bayes' rule to stochastic automata (Tsai and Fu 1979), sensitivity of error rule in feature selection (Ben-Bessat 1980, Ben-Bessat

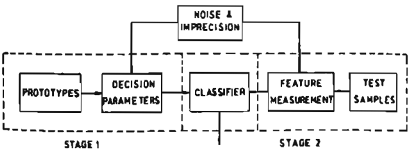



Figure 1. Stages of a pattern recognition system.

Received 5 February 1982.

† Currently on leave at Pure and Applied Physics Department, Queen's University of Belfast as Leverhulme Visiting Fellow. Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta-700 035, India.

0020-7721/82/1305 0559 $02·00 © 1982 Taylor & Francis Ltd

*et al.* 1980), classification in a time-varying environment (Swain 1978) and decision tree approaches including feature measurement cost (Dattatreye and Sarma 1981). Apart from its various direct applications, Bayes' rule also provides a standard scale to find the efficiency of other decision rules.

In practice, neither of the stages stated above can be performed ideally. In stage 1 the deviation from ideal conditions is mainly due to (*a*) finite number of prototypes and (*b*) noise or imprecision in feature measurement of the prototypes, while in stage 2 noise or imprecision may be the main problem. Another problem, i.e. the fuzziness in class description due to vagueness of features falls under the premises of fuzzy set theory and is not dealt with here.

In general, therefore, there are four cases ; where either stage 1 or 2 or both are perfect or imperfect. Of them the ideal case, when both 1 and 2 are perfect, and when 1 has deviation (*a*) stated above have been evaluated for error performance in standard literature. The present work, on the other hand, includes noise or imprecision of the measurement system using their statistical property and proceeds to evaluate the effect of the deviation (*a*) by sensitivity analysis.

Sensitivity considerations have long been of concern with dynamic systems, especially with modern control systems, where sensitivity analysis and synthesis has been used for optimization and adaptation (Horowitz 1963). The work has since been extended to diverse area, such as, microwave (Shamasundra and Gupta 1978) and fibre-optic (Chaudhuri 1979) designs—where the basic idea is to study the effectiveness of working parameters for better performance. Ben-Bessat (1980) used the sensitivity analysis to test the effectiveness of Bayes' rule to feature selection. In the present work the tolerance of misclassification probability due to fluctuation of the parameters has been found using the sensitivity analysis. The tolerance reflects the additional penalty to be paid for the deviation (*a*) in stage 1 when the deviations are small. The purpose of present analysis, therefore, is different from that presented by Ben-Bessat. However, the problem of feature selection may also be included in a more practical environment of noise and imprecision.

In the following section Bayes' error probability has been formulated for perfect and noisy systems. At first it is assumed that stage 1 is perfect and a statistical density function exists to describe the feature measurement noise. The noise density is assumed to be symmetrical about its mean, where the mean is the expected correct feature measurement. Properties of the error probability under the special case of class conditional feature density as well as noise density being normal have been discussed. In § 3 the sensitivity of Bayes' error for both the ideal and noisy cases has been formulated for normal densities, thus including small deviation in stage 1 and noise in stage 2. The case when stage 1 is also noisy is discussed in § 5 while numerical results have been presented in § 4. Another problem of learning in unsupervised environment, e.g. clustering, has also been discussed in § 5 for noisy measurement.

## 2. Bayes' error for noisy and noise-free measurements

### 2.1. *Bayes' error for noise-free measurements*

Consider multiclass pattern recognition problem with class probability $P(S_i)$, $S_i$ being the $i$th among $I$ classes. Let $\mathbf{x}$ denote feature vector with components $x_k$, $k=1, 2, \ldots, n$ optimally chosen and measured with infinite precision and resolution in a noise-free environment. Let the class conditional feature probability density $p(\mathbf{x}|S_i)$ as well as $P(S_i)$ be known exactly over $D^n \subseteq R^n$. Let

$$A_i = \{\mathbf{x} : \; P(S_i)p(\mathbf{x}|S_i) > P(S_j)p(\mathbf{x}|S_j), \; \forall j \neq i\} \tag{1}$$

Bayes' decision rule with symmetric cost function assigns sample $s$ to $S_i$ if the corresponding $\mathbf{x}(s) \in A_i$. The probability of misclassification is

$$P = \sum_{i=1}^{I} \int_{A_i^c} P(S_i)p(\mathbf{x}|S_i)\, d\mathbf{x} \tag{2}$$

where $c$ denotes complement.

It can be shown that the decision rule is best under the symmetric cost function. It minimizes the average misclassification probability for any given $P(S_i)$ and $p(\mathbf{x}|S_i)$ over $D^n$. Equation (2) accounts for error when both stage 1 and 2 are perfect.

### 2.2. *Bayes' error for noisy measurements*

Consider, now, the case when $P(S_i)$ and $p(\mathbf{x}|S_i)$ are exactly known, i.e. stage 1 is perfect but feature measurement at the classification phase is noisy, i.e. stage 2 has deviation (*b*) stated above. Considering this measurement as $\mathbf{y}$ with noise $\mathbf{e} = \mathbf{y} - \mathbf{x}$ we assume that:

(i) the distribution of $\mathbf{e}$ is independent of position in $D^n$;
(ii) $p(\mathbf{y}|\mathbf{x})$ is independent of the class to which $\mathbf{x}$ belongs; and
(iii) $p(\mathbf{y}|\mathbf{x})$ is symmetric about the true measurement $\mathbf{x}$.

For zero-mean gaussian noise the assumptions are valid. If the decision rule is same on $\mathbf{y}$, i.e. for an observed $\mathbf{y}$ we assign it to class $S_i$ if

$$P(S_i)p(\mathbf{x}|S_i)|_{\mathbf{x}=\mathbf{y}} > P(S_j)p(\mathbf{x}|S_j)|_{\mathbf{x}=\mathbf{y}}, \quad \forall j \neq i \tag{3}$$

then the probability of misclassification is (Chaudhuri *et al.* 1982)

$$P_e = \sum_{i=1}^{I} \int_{D^n} P(S_i)p(\mathbf{x}|S_i) \int_{A_i^c} p(\mathbf{y}|\mathbf{x})\, d\mathbf{y}\, d\mathbf{x} \tag{4}$$

We have, after some algebraic manipulations

$$\begin{aligned} \Delta P &= P_e - P \\ &= \sum_{i=1}^{I} \int_{A_i} \sum_{j \neq i} [P(S_i)p(\mathbf{x}|S_i) - P(S_j)p(\mathbf{x}|S_j)] \int_{A_i^c} p(\mathbf{y}|\mathbf{x})\, d\mathbf{y}\, d\mathbf{x} \end{aligned} \tag{5}$$

Using inequality (3) it can be easily proved that

$$\Delta P \geqslant 0 \tag{6}$$

It should be noted that eqn. (3) does not lead to minimum $P_e$ or $\Delta P$ if $p(x|S_i)$ for all $i$ do not have identical shape. An example of identical $p(x|S_i)$ is the normal densities having equal covariance matrix when $A_i$ is the same as in § 2.1. The integral in eqn. (4) is actually a convolution of $p(x|S_i)$ and $p(y|x)$ leading to $p(y|S_i)$. However, the present form of eqn. (4) is used because it allows a direct comparison with $P$ through eqn. (5) and the sensitivity of error with density parameters—as described in § 3—can be explicitly expressed in this form. Furthermore, some computational economy is possible in the general decision making for classification.

For $p(\mathbf{x}|S_i)$ and $p(\mathbf{y}|S_i)$ normal, the evaluation of eqns. (4) and (5) requires multiple infinite integration over quadratic hypersurfaces. For two-class univariate problem with equal covariance we have

$$P_e = \int_{-\infty}^{\infty} p(x|S_1) \int_{x_1}^{\infty} p(y|x)\, dy\, dx \tag{7}$$

where $P(S_1) = P(S_2) = 0{\cdot}5$ and

$$p(x|S_i) = \frac{1}{\sqrt{(2\pi)}\sigma} \exp -\left(\frac{x-\mu_i}{\sqrt{2}\sigma}\right)^2, \quad \forall i \tag{8}$$

$$p(y|x) = \frac{1}{\sqrt{(2\pi)}\sigma_e} \exp -\left(\frac{y-x}{\sqrt{2}\sigma_e}\right)^2 \tag{9}$$

$$x_1 = (\mu_1 + \mu_2)/2 \tag{10}$$

have been assumed. It can be shown that $\Delta P$ is monotonic increasing with $\sigma_e$. Furthermore, using the limiting conditions $\sigma_e \to 0$ and $\sigma_e \to \infty$ it can be shown that the upper and lower limit of $\Delta P$ is $0{\cdot}5 - P$ and 0, respectively. Also, the results are true for other densities $p(y|x)$ satisfying the conditions (i)–(iii) given above.

When $p(\mathbf{y}|S_i)$ rather than $p(\mathbf{x}|S_i)$ is given, a knowledge of $p(\mathbf{y}|\mathbf{x})$ can lead to $p(\mathbf{x}|S_i)$ from $p(\mathbf{y}|S_i)$ if the error $\mathbf{e}$ is assumed independent of $\mathbf{x}$ and all density functions normal. The above formulation and strategy can then be used to find the performance of a classifier both trained and working in a noisy environment. However, a complete relaxation of the requirement of large number of prototypes at the training phase may not be accommodated in the present framework and may be treated in terms of bounds of error. But if it is known with high confidence that the fluctuation in parameters defining $p(\mathbf{y}|S_i)$ or $p(\mathbf{x}|S_i)$ due to small number of prototypes is small, the following sensitivity analysis can be used to advantage for classifier performance.

## 3. Sensitivity analysis

### 3.1. *Parameter sensitivity*

The sensitivity of one parameter $x$ with that of $y$ may be defined as

$$S_{x,y} = \frac{\partial x}{\partial y}$$

It is to be noted that the definition given above is not, usually, dimensionless. Apart from this there are other problems regarding the definition of sensitivity. Excellent review regarding these problems in control theory is given in Kukotovic and Rutman (1965).

We do, however, accept the above definition modified so that it becomes dimensionless. Thus we accept (Peikari 1974)

$$S_{x,y} = \frac{y}{x}\frac{\partial x}{\partial y} \tag{11}$$

If a characteristic $C$ is a function of several independent parameters $y_i$, $i = 1, 2, \ldots, n$, then the change in the characteristics $\Delta C$ is related to the tolerance $\Delta y_i$ of $y_i$ by the relation

$$\Delta C = C \sum_{i=1}^{n} (\Delta y_i | y_i) S_{C,y_i} \tag{12}$$

### 3.2. *Sensitivity of Bayes' error for ideal case*

To find the sensitivity of Bayes' error probability in pattern classification, we shall consider the density $p(x|S_i)$ normal. Since mean and standard deviation are the parameters completely specifying the density, it is only necessary to find the sensitivity with these parameters.

However, even for the simplest two class univariate problem, there are two means and two standard deviations to be accounted. It is rather useful to combine the two mean $\mu_1$ and $\mu_2$ as a single parameter $\mu_2-\mu_1=\Delta\mu$ because it is the difference rather than individual values that are sensitive to misclassification probability. For the simplest case $P(S_1)=P(S_2)=0{\cdot}5$ we have, using eqns. (2), (8) and (11)

$$S_{P,\Delta\mu}=-\frac{\Delta\mu}{2P}p(x_1|S_1) \tag{13}$$

which means that the error sensitivity is proportional to the density at decision boundary $x_1$ and decreases as $\Delta\mu$, the separation between the two means increases.

Under the same conditions as above, we have

$$S_{P,\sigma}=\frac{\Delta\mu}{2P}p(x_1|S_1) \tag{14}$$

which means that the error sensitivity increases at the same rate as $p(x_1|S_i)$ increases with $\sigma$. It is interesting to note that eqns. (13) and (14) are the same except for sign. However the variable and constants are different in the two cases.

We can, however, combine the two parameters together in the form $\Delta\mu/\sigma=z$ and see that

$$S_{P,z}=-\frac{\Delta\mu}{2P}p(x_1|S_1) \tag{15}$$

which again has the same form as eqn. (13) but tolerance to both $\Delta\mu$ and $\sigma$ fluctuation has been accounted here.

Let, now, $p(\mathbf{x}|S_1)$ be a multivariate normal density with mean $\boldsymbol{\mu}_1$ and covariance matrix $[\sigma_1^2]$. Let $[\sigma_1^2]=[\sigma_2^2]=[\sigma^2]$. Again, a straightforward evaluation of error sensitivity is difficult because of multiple integration involved over decision boundary. The misclassification probability derived from the distribution of likelihood ratio (Fukunaga 1972) may, however, be used to advantage.

The minus-log likelihood ratio $h_{12}(\mathbf{x})$ for two classes $S_1$ and $S_2$ are defined as

$$h_{12}(\mathbf{x})=(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^{\mathrm{T}}[\sigma^2]^{-1}\mathbf{x}+\tfrac{1}{2}\{\boldsymbol{\mu}_1^{\mathrm{T}}[\sigma^2]^{-1}\boldsymbol{\mu}_1-\boldsymbol{\mu}_2^{\mathrm{T}}[\sigma^2]^{-1}\boldsymbol{\mu}_2\} \tag{16}$$

and the decision rule is

$$h(\mathbf{x})\lessgtr \ln\frac{P(S_1)}{P(S_2)}\to\mathbf{x}\in\begin{cases}S_1\\S_2\end{cases}$$

The distribution of $h_{12}(\mathbf{x}|S_i)$ is also normal and its expected value and covariance are given, respectively, by

$$\begin{aligned}\eta_1&=\mathscr{E}\{h_{12}(\mathbf{x}|S_1)\}\\&=-\tfrac{1}{2}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^{\mathrm{T}}[\sigma^2]^{-1}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)\triangleq-\eta\end{aligned} \tag{17}$$

with

$$\eta_2 = -\eta_1$$

and

$$\sigma_{h1}^2 = \mathscr{E}[\{h_{12}(\mathbf{x}|S_1) - \eta_1\}^2] = 2\eta \tag{18}$$

with

$$\sigma_{h1}^2 = \sigma_{h2}^2 = \sigma_h^2$$

Note that $2\eta = z^2$ for univariate case. To be consistent with this we define $2\eta = z^2$ and find the error sensitivity with respect to this combined parameter $z$. Now, the error probability is

$$P = P(S_1) \int_t^{\infty} p_h(h_{12}|S_1)\, dh_{12} + P(S_2) \int_{-\infty}^{t} p_h(h_{12}|S_2)\, dh_{12}$$

$$= \frac{1}{\sqrt{(2\pi)}} \left[ P(S_1) \int_{(\eta + t/\sigma_h)}^{\infty} \exp -\frac{\xi^2}{2}\, d\xi + P(S_2) \int_{(\eta - t/\sigma_h)}^{\infty} \exp -\frac{\xi^2}{2}\, d\xi \right] \tag{19}$$

where

$$t = \ln \{P(S_1)/P(S_2)\}$$

from (12) and (19) we have,

$$S_{P,z} = -\frac{z}{\sqrt{(2\pi)}P} \left[ P(S_1) \left( \tfrac{1}{2} - \frac{t}{z^2} \right) \exp -\tfrac{1}{2} \left( \frac{(z^2/2) + t}{z} \right)^2 + P(S_2) \left( \tfrac{1}{2} + \frac{t}{z^2} \right) \exp -\tfrac{1}{2} \left( \frac{(z^2/2) - t}{z} \right)^2 \right] \tag{20}$$

Note that eqn. (20) is the same as (15) if $P(S_1) = P(S_2) = 0{\cdot}5$ and $\mathbf{x}$ becomes univariate.

### 3.3. *Sensitivity of Bayes' error for noisy measurement*

For the practical case of noisy measurement there are three parameters to be considered for sensitivity analysis. If however, the noise statistics is assumed accurately known we are left with $\Delta\mu$ and $\sigma$. Error sensitivity with respect to $\Delta\mu$ and $\sigma$ can be derived individually. But a close look into eqn. (7)–(9) clarifies that no closed form expression exists for the sensitivity with respect to $z = \Delta\mu/\sigma$ defined earlier. This is because $\Delta\mu$ is involved through $x_1$ in the limit of second integration while $\sigma$ is a parameter of the first integral. However the tolerance of error probability when both $\Delta\mu$ and $\sigma$ are fluctuating is found from individual sensitivities using eqn. (12) above.

It can be shown by straightforward calculation that for two-class univariate case with equal variance

$$S_{P_n,\Delta\mu} = -\frac{\Delta\mu}{2P} \int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)}\sigma} \exp -\frac{(x-\mu)^2}{2\sigma^2}\, p(x_1|x)\, dx \tag{21}$$

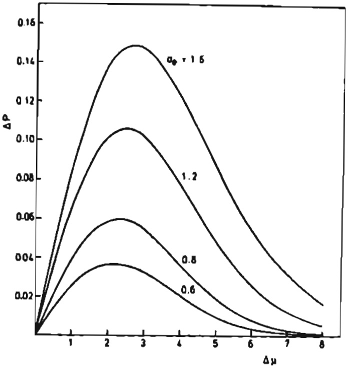
and

$$S_{P_n,\sigma} = -\frac{1}{P} \int_{-\infty}^{\infty} \left[ 1 - \frac{x^2}{\sigma^2} \right] p(x|S_1) \int_{x_1}^{\infty} p(y|x)\, dy\, dx \tag{22}$$

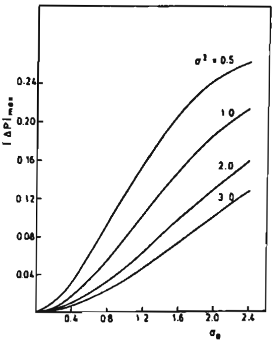
where $p(x_1|x)$ is the value of $p(y|x)$ at $y = x_1$ with mean at $x$.

It is seen that the eqns. (21) and (22) require integration over infinite range and hence a simple extension of the equation to multivariate case will involve multiple such integrations. A straightforward simplification of the problem even for $[\sigma_1^2]=[\sigma_2^2]$ as in eqn. (19) for the ideal case is not possible because of double integration involved in $P_e$.

But if $\sigma^2$ and $\sigma_e^2$ are added so that $\sigma_{yi}^2=\sigma_i^2+\sigma_e^2$ and $\mu_{yi}=\mu_i+\mu_e$ the sensitivity with respect to these parameters can be found using (13)–(15) and (20).




(*a*)




(*b*)

Figure 2. (*a*) Excess Bayes' error for noisy measurement ($\sigma=1{\cdot}0$). (*b*) Maxima of excess Bayes' error for noisy measurement.

## 4. Numerical results

In Fig. 2 (*a*) the excess probability of misclassification $\Delta P$ has been plotted against $\Delta\mu$ for different $\sigma_\varepsilon$ using eqns. (7)–(9). The plots show that $\Delta P$ increases with $\sigma_\varepsilon$ for a fixed $\Delta\mu$. For $\Delta\mu = 0$, $\Delta P = 0$ because the two classes are statistically indistinguishable and $P_\varepsilon = P = 0{\cdot}5 = P(S_i)$ for any value of $\sigma_\varepsilon$, the worst possible case of classification. The decision strategy is to assign $x$ or $y$ to $S_1$ or $S_2$ according as $x > \mu_1 = \mu_2$ or not and hence $\sigma_\varepsilon$ is immaterial. The other limiting case $\Delta P \to 0$ as $\Delta\mu \to \infty$ occurs because $P_\varepsilon \to 0$, $P \to 0$, the order of convergence of the former depending on $\sigma_\varepsilon$. The function $\Delta P$ shows a maximum which can be found by differentiating it with respect to $\Delta\mu$. The maximum occurs at lower $\Delta\mu$ with $\sigma_\varepsilon$ decreasing and vanishes with $\sigma_\varepsilon$ vanishing with $\Delta P \to 0$ also.

In Fig. 2 (*b*) the maximum of $\Delta P$ is plotted against $\sigma_\varepsilon$ for different values of $\sigma$. The maximum is quite significant in pattern recognition problems. It describes the optimum sensitivity of excess misclassification probability at the particular situation and indicates the noise immunity necessary for the system to overcome it.

The magnitude of sensitivity of misclassification probability $S_{P,\Delta\mu}$ has been plotted against $\Delta\mu$ for different values of $\sigma$ in Fig. 3. It is shown that the magnitude increases with $\mu$ for a fixed $\sigma$, i.e. $S_{P,\Delta\mu}$ decreases with $\Delta\mu$. For a fixed $\Delta\mu$, the magnitude increases with $\sigma$ decreasing, the error being less sensitive. This is because the densities $p(x|S_i)$ becomes more flat and absolute value of its slope at any point decreases with $\sigma$ decreasing.

In Fig. 4 the sensitivity of misclassification probability with $\sigma$, i.e. $S_{P,\sigma}$ has been plotted against $\sigma$ for different values of $\Delta\mu$. The sensitivity sharply decreases around $\sigma = 0{\cdot}2$ and falls linearly thereafter. It decreases also with $\Delta\mu$ for a fixed $\sigma$, the reason again being flattening of slope for large $\Delta\mu$.

The magnitude of sensitivity of misclassification probability with respect to $z$ has been plotted in Fig. 5. The nature of the curve is similar to those in Fig. 3 and can be used for multivariate case also. However, for multivariate

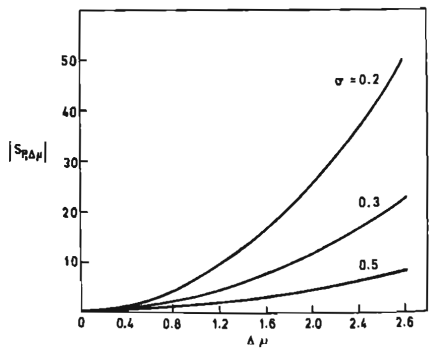



Figure 3. Sensitivity of Bayes' error with mean separation (ideal case).

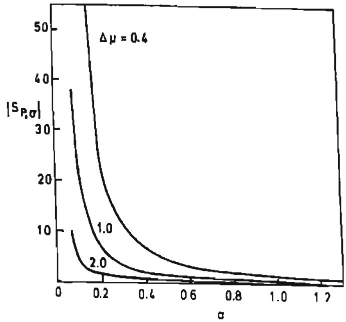



Figure 4. Sensitivity of Bayes' error with standard deviation (ideal case).

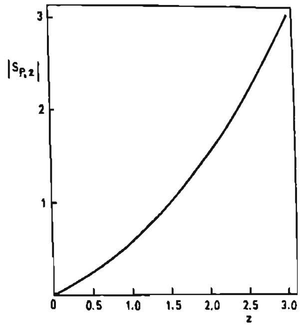



Figure 5. Sensitivity of Bayes' error with variable $z$ (ideal case).

case the relationship with the direct parameters $\mu_i$ and $[\sigma_i^2]$ is implicit and it is necessary to use eqns. (17)–(18) to find the exact tolerance of misclassification probability to these parameters.

The magnitude of sensitivity of misclassification probability with noisy measurement has been plotted against $\Delta\mu$ in Fig. 6. It is seen that the magnitude decreases with increasing $\sigma_e$. The value of sensitivity for noiseless measurement has been shown for comparison of relative magnitudes. Although the change in sensitivity is low, it should be mentioned that $P_e$ is higher than $P$ for all values of $\Delta\mu$.

Figure 7 shows the sensitivity of misclassification probability with respect to $\sigma$ under noisy environment conditions. The sensitivity shows a peak that shifts towards larger $\sigma$ for higher $\sigma_e$. It increases with $\Delta\mu$ decreasing and vanishes as $\Delta\mu \to \infty$. The peaks denote the positions where the error is most sensitive.

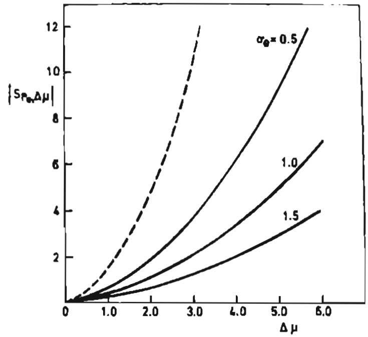



**Figure 6.** Sensitivity of Bayes' error with mean separation (for noisy measurement ..., denotes $S_{p,\mu}$ for $\sigma=0{\cdot}5$).

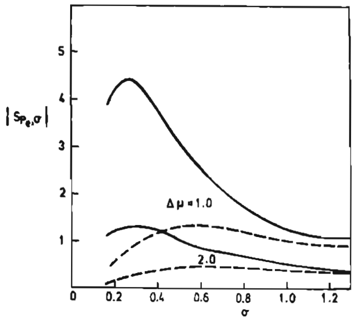



**Figure 7.** Sensitivity of Bayes' error with standard deviation (for noisy measurement, —for $\sigma_\varepsilon=0{\cdot}25$, ..., for $\sigma_\varepsilon=0{\cdot}5$).

| $\sigma_\varepsilon$ | $\Delta\mu$ | $P$ or $P_\varepsilon$ | $\|\Delta P\|$ or $\|\Delta P_\varepsilon\|$ |
|---|---|---|---|
| — | 0·6 | 0·38208 | $11{\cdot}44\times10^{-4}$ |
| — | 1·0 | 0·30854 | $10{\cdot}04\times10^{-4}$ |
| 0·5 | 0·6 | 0·39427 | $13{\cdot}58\times10^{-4}$ |
| | 1·0 | 0·32727 | $10{\cdot}59\times10^{-4}$ |
| 1·0 | 0·6 | 0·41592 | $11{\cdot}47\times10^{-4}$ |
| | 1·0 | 0·36176 | $9{\cdot}58\times10^{-4}$ |

The Table gives an idea about the tolerance of $P$ or $P_e$ to 1% fluctuation in $\Delta\mu$ from their true values. Thus, for $\sigma_e = 0{\cdot}5$, $\Delta\mu = 1{\cdot}0$, the system will suffer additional error $\pm 0{\cdot}059\%$ if $\Delta\mu$ deviates $\pm 1\%$ from its true value due to estimation from a finite number of prototypes.

## 5. Discussions

The formulation and underlying assumptions in § 2 is useful and important in pattern recognition problems. Situations like this occurs when the recognizer is trained in good laboratory conditions with large numbers of prototypes and utmost precaution can be taken to make the feature measurements noise-free while the sample data to be recognized may be telemetric and hence susceptible to noise and imprecision. An inverse situation is less likely to occur. However, when this occurs, the assumptions about $p(\mathbf{y}|\mathbf{x})$ allow an estimation of $p(\mathbf{x}|S_i)$ from $p(\mathbf{y}|S_i)$ and hence the formulation of § 2 can be used thereafter. Nevertheless, when the assumptions are invalid should be treated carefully.

The problem of recognition error due to finiteness of prototypes in estimating the parameters of $p(\mathbf{x}|S_i)$ can be treated using the sensitivity analysis. If it is known with a given risk that the deviation of the parameter $p$ from the true value $p_i$ has a small range $\Delta p_i$ about $p_i$, the excess misclassification probability can be found as discussed above.

Sometimes Bayes' error probability is used for feature selection problems. The analysis presented in § 2.2 may be used for the more general case of noisy feature. However, as Ben-Bassat pointed out for the noise-free case, the sensitivity of Bayes' error is sometimes quite low and should be treated carefully. It should be mentioned that the sensitivity is lower for noisy than for noise-free case.

The numerical results presented with the help of a VAX 11/780 computer in double precision covers the workable range of parameters and can be used directly, if necessary. The formulation and results have been presented for normal $p(\mathbf{y}|\mathbf{x})$ so as to include the most common electrical white gaussian noise. Other forms of density functions may also be used in a similar manner. There is a problem of computing the misclassification probability for noisy multivariate case which involves several infinite integrals. The problem may be attempted either using the approach of Fukunaga (1972) or through the approaches of finding error bounds.

Another problem that needs attention is unsupervised learning and recognition under noisy environment. Of the unsupervised techniques clustering is one of the most important for its applications beyond pattern recognition problems. Both hierarchal and non-hierarchical agglomerative and devisive methods have found wide use and a new approach that has drawn considerable attention recently is the fuzzy set theoretic approach. While the former techniques define hard sets, the later describes each cluster as a fuzzy set with a grade of membership for each element over the demain. But none of the approaches considers noise, imprecision or ambiguity. Hence any such algorithm will produce inconsistent clustering of the data produced by a noisy generating system at different instances. Although there is no general solution to the problem, some properties may be studied if the noise statistics is known.

Let the noise density be zero-mean normal with a standard deviation $\sigma$ so that any noise-free datum remaining within $\pm 3\sigma$ of the observed value has a probability 0·99. Now if a clustering algorithm is used on two positions of the data $3\sigma$ away from the observed value, we get two sets of clusters. The intersection of the elements of the corresponding clusters in two cases may be called a consistent cluster with high confidence. A degree of inconsistency may be defined on the number of data in consistent cluster in relation to the total number in the original cluster. The region of consistent cluster is useful as a region more immune to the noise and the data falling in that region may be used as prototype in modelling a pattern recognizer. The problem of clustering for practical noisy data will be discussed elsewhere in details.

## Acknowledgments

The author wishes to thank Professor D. Dutta Majumder, Indian Statistical Institute and Dr. B. Bates, Queen's University of Belfast for their interest and the Leverhulme Trust for offering him a fellowship to work at Queen's University, Belfast. Help rendered by Mrs. M. Simpson and Mrs. N. McAllister in preparing the manuscript is gratefully acknowledged.

## References

Andrews, H. A., 1972, *Mathematical Techniques in Pattern Recognition* (New York: Wiley–Interscience).

Ben-Bessat, M., 1980, *I.E.E.E. Trans. Pattern Anal. Mach. Intell.*, **2**, 57.

Ben-Bessat, M., Klove, K. L., and Weil, M. H., 1980, *I.E.E.E. Trans. Pattern Anal. Mach. intell.*, **2**, 261.

Chaudhuri, B. B., 1979, *I.E.E.E. Jl quant. Electron.*, **15**, 545.

Chaudhuri, B. B., Murthy, C. A., and Dutta-Majumder, D., 1982, *I.E.E.E. Trans. Syst. Man Cyber.* (submitted).

Datatreya, G. R., and Sarma, V. V. S., 1980, *I.E.E.E. Trans. Pattern Anal. Mach. Intell.*, **3**, 293.

Fukunaga, K., 1972, *Introduction to Statistical Pattern Recognition* (New York: Academic Press).

Horowitz, I. M., 1963, *Synthesis of Feedback System* (New York: Academic Press).

Kukotovic, P. V., and Rutman, R. S., 1965, *Automatika i telemekhanika* (U.S.S.R.), **26**, 727.

Peikari, B., 1974, *Fundamentals of Network Analysis and Synthesis* (New Jersey: Prentice–Hall).

Shamusundra, S. D., and Gupta, K. C., 1978, *I.E.E.E. Trans. Microw. Theory Tech.*, **26**, 788.

Swain, P. H., 1978, *I.E.E.E. Trans. Syst. Man Cyber.*, **8**, 879.

Tsai, W. H., and Fu, K. S., 1979, *I.E.E.E. Trans. Syst. Man Cyber.*, **9**, 745.