Some Results on Multicategory Pattern Recognition

W. W. Bledsoe

Panoramic Research, Inc., Palo Alto, California

Abstract. Some problems arising in multicategory (many pattern types) pattern recognition are treated mathematically, and formulas are derived which describe some inherent limitations associated therewith. The principal results concern the "dimensional" and "correlational" effects and their degradation of a multimeasurement recognition system.

1. *Introduction*

In simplified character recognition one is usually dealing with a limited number of pattern types or categories. Thus, we might have ten or eleven categories for numeral recognition, and 36–100 categories for alphanumeric recognition. However, there are a large number of applications, such as the recognition of fingerprints, military targets, chinese characters, facial photographs, paintings, etc., where the number of categories is enormous. In fact, most instances of human recognition fall in this case.

The purpose of this paper is to treat mathematically some problems that arise in many-category pattern classification, and to derive formulas describing the inherent limitations associated therewith. The principal results obtained on the "dimensional" and "correlational" effects in multimeasurement systems are presented in Section 3. In these results it is shown how the overall effectiveness of such a multimeasurement system can be predicted from elementary statistics.

2. *One-Dimensional Case*

Suppose we have a population of pattern types (categories):

$$P = \{p_1, p_2, \cdots, p_Q\}.$$

When a particular representation, $\hat{p}_i$, is given, it is to be identified or recognized as a member of the category p_i. Further assume that X is a real function of P.[1] The collection of all $X(p_i)$'s can be described by a probability density[2] function f where

$$\int_{-\infty}^{\infty} f(X)\, dX = 1.$$

Now suppose that we attempt to determine X by a measurement, x, and that the measurement error, $x - X$, is governed by the probability density function g_X for each X. (See Figure 1.)

Question: If X is unknown, how much uncertainty is eliminated by such a measurement x? The answer to this question can be expressed in terms of informa-

[1] For example, P might represent the collection of all people in the world, p could be a photograph of a person, and $x(p_i)$ could be the (absolute) width of the head of person p_i.

[2] X will be treated as a continuous variable in this discourse. This is a realistic assumption when the collection, P, of categories is large.

Journal of the Association for Computing Machinery, Vol. 13, No. 2 (April, 1966), pp. 304–316

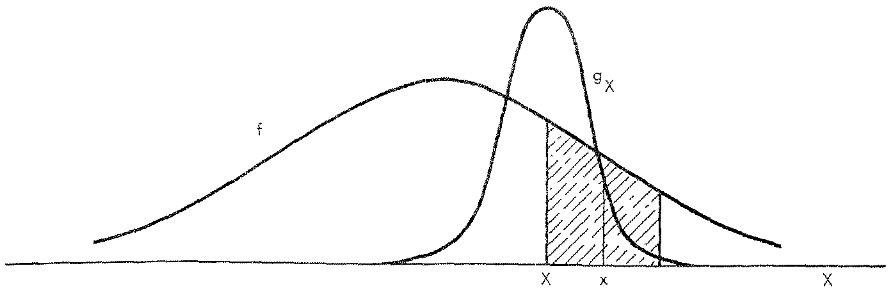

FIG. 1

tion theory [1, p. 54–64] as follows:

$$R(f, g_X) = H(f) - H(g_X)$$

where

$$H(f) = -\int_{-\infty}^{\infty} f(X) \log f(X)\, dX$$

is the entropy (or information) of the function f, and R is the transmission rate.[3]

If g_X is independent of X then $g_X = g$, and it follows immediately that [1, Th. 16, p. 66]

THEOREM 1. $R(f, g) = H(f) - H(g)$.

Also if f and g are normal with standard deviations σ and σ_E then it is easily shown [1, Th. 14, p. 66] that

THEOREM 2.

$$\begin{aligned} H(f) &= \log [(2\pi e)^{\frac{1}{2}}\sigma] \\ H(g) &= \log [(2\pi e)^{\frac{1}{2}}\sigma_E] \\ R(f, g) &= \log (\sigma/\sigma_E). \end{aligned}$$

Thus we could say that the reduction in uncertainty is $\log (\sigma/\sigma_E)$, or that $\log (\sigma/\sigma_E)$ bits of information have been transmitted by such a measurement. This result gives a general idea of how the uncertainty is reduced, but in a particular application the actual reduction in uncertainty depends on the method used for recognition.

What is the fractional reduction of uncertainty, F_{Xx}, associated with a given X and its approximation x? Let us define[4]

$$F_{Xx} = \int_{|z-x| \le |x-X|} f(z)\, dz.$$

Definition 1. Let

$$F(f, g) = \int_{-\infty}^{\infty} f(X) \int_{-\infty}^{\infty} g_X(x - X) \int_{|z-x| \le |x-X|} f(z)\, dz\, dx\, dX.$$

Thus $F(f, g)$ is the *average* fractional reduction in uncertainty.

[3] Base 2 logarithms will be assumed unless otherwise indicated.

[4] F_{Xx} is indicated by the hatched area in Figure 1, and represents the "fractional amount that has to be searched (starting from x) in order to find X."

Some of the properties of F are derived here in Section 2. The principal results of this paper are obtained in Section 3 where F is generalized to the multimeasurement case.

Definition 2. Let $U_\sigma(x) = \frac{1}{2}a$, $-a \leq x \leq a$, where $a = 3^{\frac{1}{2}}\sigma$, and note that U_σ is a uniform probability density function with mean O and variance σ^2. Let N_σ denote the normal probability density function

$$N_\sigma(x) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right), \qquad -\infty < x < \infty,$$

with mean O and variance σ^2.

THEOREM 3.

$$F(U_\sigma, U_{\sigma_E}) = \frac{1}{2}\frac{\sigma_E}{\sigma} - \frac{1}{6}\left(\frac{\sigma_E}{\sigma}\right)^2.$$

PROOF. Apply definitions 1 and 2 and integrate.

LEMMA 1.

$$(1) \quad N_\sigma(x) \cdot N_\sigma(x + s) = N_{2^{\frac{1}{2}}\sigma}(s) N_{\sigma/2^{\frac{1}{2}}}(x + s/2)$$

$$(2) \quad N_\sigma(x) \cdot N_{\sigma'}(x) = \frac{1}{(2\pi c)^{\frac{1}{2}}} N_{\sigma''}(x), \quad \text{where} \quad c = (\sigma^2 + \sigma'^2)^{\frac{1}{2}}, \quad \sigma'' = \frac{\sigma\sigma'}{c}.$$

$$(3) \quad N_\sigma(sx) = \frac{1}{s} N_{\sigma/s}(x)$$

$$(4) \quad \int_0^\infty x N_\sigma(x) N_{\sigma'}(x)\, dx = \sigma\sigma'/2\pi(\sigma^2 + \sigma'^2).$$

THEOREM 4.[5]

$$F(N_\sigma, N_{\sigma_E}) = \frac{1}{\pi} \operatorname{Arctan}\left(2^{\frac{1}{2}} \cdot \frac{\sigma_E}{\sigma}\right).$$

PROOF. Let $f = N_\sigma$, $g = N_{\sigma_E}$. Using definitions 1 and 2, we have

$$\begin{aligned}
F(N_\sigma, N_{\sigma_E}) &= \int_{-\infty}^\infty f(X) \int_{-\infty}^\infty g(x - X) \int_{|z-x| \leq |X-x|} f(z)\, dz\, dx\, dX \\
&= \int_{-\infty}^\infty g(u) \int_{-\infty}^\infty f(X) \int_{|z-(X+u)| \leq |u|} f(z)\, dz\, dX\, du \\
&= 2\int_0^\infty g(u) \int_{-\infty}^\infty f(X) \int_X^{X+2u} f(z)\, dz\, dX\, du \\
&= 2\int_0^\infty g(u) \int_{-\infty}^\infty f(X) \int_0^1 f(X + 2ut) \cdot 2u\, dt\, dX\, du \\
&= 4\int_0^1 \int_0^\infty u g(u) \int_{-\infty}^\infty N_{2^{\frac{1}{2}}\sigma}(2ut) N_{\sigma/2^{\frac{1}{2}}}(X + ut)\, dX\, du\, dt \quad \text{by Lemma 1.1} \\
&= 4\int_0^1 \int_0^\infty u g(u) N_{2^{\frac{1}{2}}\sigma}(2ut)\, du\, dt \\
&= 4\int_0^1 \int_0^\infty u N_{\sigma_E}(u) \frac{1}{2t} N_{(2^{\frac{1}{2}}\sigma/2t)}(u)\, du\, dt \qquad \text{by Lemma 1.3}
\end{aligned}$$

[5] D. B. Owen, Graduate Center of the Southwest, first brought this result to my attention.

$$= 4 \int_0^1 \frac{1}{2t} \frac{\sigma_E \dfrac{\sigma}{2^{\frac{1}{2}} t}}{2\pi \left(\sigma_E^{\,2} + \dfrac{\sigma^2}{2t^2} \right)} \, dt \qquad \text{by Lemma 1.4}$$

$$= \frac{\sigma}{2^{\frac{1}{2}} \pi \sigma_E} \int_0^1 \frac{dt}{t^2 + \dfrac{\sigma^2}{2\sigma_E^{\,2}}} = \frac{\sigma}{2^{\frac{1}{2}} \pi \sigma_E} \cdot \frac{2^{\frac{1}{2}} \sigma_E}{\sigma} \operatorname{Arctan} \left(2^{\frac{1}{2}} \cdot \frac{\sigma_E}{\sigma} \right)$$

$$= \frac{1}{\pi} \operatorname{Arctan} \left(2^{\frac{1}{2}} \cdot \frac{\sigma_E}{\sigma} \right). \qquad \text{Q.E.D.}$$

THEOREM 5.
(1) $H(N_\sigma) = \log\, ((2\pi e)^{\frac{1}{2}} \sigma)$,
(2) $H(U_\sigma) = \log(2 \cdot 3^{\frac{1}{2}} \sigma)$,
(3) $R(N_\sigma, N_{\sigma_E}) = R(U_\sigma, U_{\sigma_E}) = -\log\, (\sigma_E/\sigma)$.

PROOF. Let $a = 3^{\frac{1}{2}} \sigma$. $H(U_\sigma) = -\int_{-a}^{a} (1/2a) \log\, (1/2a)\, dx = \log\, (2 \cdot 3^{\frac{1}{2}} \sigma)$. (1) is given in Theorem 2; (3) is a consequence of (1), (2) and Theorem 1.

THEOREM 6. *If* $\sigma_E \ll \sigma$, *and if* $F = F(N_\sigma, N_{\sigma_E})$ *or* $F = F(U_\sigma, U_{\sigma_E})$, *and* $R = F(N_\sigma, N_{\sigma_E})$ *or* $R = R(U_\sigma, U_{\sigma_E})$, *then* $F \simeq \frac{1}{2} \sigma_E/\sigma$, $R = -\log\, (\sigma_E/\sigma)$, *and* $F \simeq \frac{1}{2} \cdot 2^{-R}$.

PROOF. Since $\sigma_E \ll \sigma$, it follows from Theorem 4 that

$$F(N_\sigma, N_{\sigma_E}) = \frac{1}{\pi} \operatorname{Arctan} \left(2^{\frac{1}{2}} \frac{\sigma_E}{\sigma} \right)$$
$$= \frac{1}{\pi} \left[\frac{2^{\frac{1}{2}} \sigma_E}{\sigma} - \frac{1}{3} \left(\frac{2^{\frac{1}{2}} \sigma_E}{\sigma} \right)^3 + \cdots \right]$$
$$\simeq \frac{2^{\frac{1}{2}}}{\pi} \frac{\sigma_E}{\sigma} \simeq 0.451 \frac{\sigma_E}{\sigma} \simeq \frac{1}{2} \frac{\sigma_E}{\sigma},$$

and from Theorem 3 that

$$F(U_\sigma, U_{\sigma_E}) = \frac{1}{2} \frac{\sigma_E}{\sigma} - \frac{1}{6} \left(\frac{\sigma_E}{\sigma} \right)^2 \simeq \frac{1}{2} \frac{\sigma_E}{\sigma}.$$

Apply Theorem 5 and the definition of log.

By Theorem 6 it can be seen that the quantities F and R hold relatively constant for two widely differing types of density functions (uniform and normal). Also, $F \simeq \frac{1}{2} \cdot 2^{-R}$ shows an intuitively satisfying connection between R, the information transferred, and F the fractional reduction in uncertainty.

3. *Multidimensional Case*

Suppose instead of one function X, we have several functions[6] $X_1, X_2, \cdots, X_K$ on P. These are approximated by measurements $x_1, x_2, \cdots, x_K$ respectively, and X and x are governed by the K-dimensional probability density functions f and g. Definition 1, for the average functional reduction of uncertainty, is generalized in a natural way as follows.

[6] If we again refer to the photograph recognition example mentioned earlier, the numbers $X_1(p_i)$, $X_2(p_i)$, $\cdots$ might represent the width of the head, the length of the left ear, etc., for the person p_i.

Definition 3.

$$F_K(f, g) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(X_1, X_2, \cdots X_K) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}$$

$$\cdot g(x_1 - X_1, \cdots, x_K - X_K) \int \cdots \int_{|z-x| \leq |x-X|} f(z_1, \cdots z_K)\, dz_K \cdots dX_1.$$

Question: If we obtain an average fractional reduction in uncertainty of F_K for *each* of the K measurements, do we get a reduction proportional to F for the combined set of measurements? Or more precisely, if f_i and g_i are the marginal densities of f and g on dimension i, and if

$$F(f_i, g_i) = F \simeq \frac{1}{2}\frac{\sigma_E}{\sigma} \qquad (i = 1, \cdots, K),$$

do we get $F_K(f, g) \simeq \frac{1}{2}(\sigma_E/\sigma)^K$?

This simple relation is obtained only in the extreme case where f and g are uniformly distributed. In the "natural" case, when f and g are normal, we shall present results which show that $F_K = D \cdot C \cdot S$ where $S = (\sigma_{E1}/\sigma_1)(\sigma_{E2}/\sigma_2) \cdots (\sigma_{EK}/\sigma_K)$ is a term which tends to reduce F_K, and D and C are "dimensional" and "correlational" effects which tend to increase F_K.

Let us denote by $N_{K\sigma}$ the K-dimensional normal probability density function.

Definition 4.

$$N_{K_\sigma}(X) = \frac{(|\sigma^{ij}|)^{\frac{1}{2}}}{(2\pi)^{K/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{K}\sum_{j=1}^{K} \sigma^{ij} X_i X_j\right).$$

We will use the standard notation, σ_{ij}, to represent the element of the symmetric positive definite covariance matrix,[7] and σ^{ij} for elements of its inverse. $|A|$ represents the determinant of the matrix A. Also we use the notation $\sigma_i = (\sigma_{ii})^{\frac{1}{2}}$ and $\rho_{ij} = \sigma_{ij}/\sigma_i\sigma_j$, the correlation coefficients.

Definition 5.

$$V_K(r) = \int \cdots \int_{x_1^2 + \cdots + x_K^2 \leq r^2} dx_1, \cdots dx_K$$

Thus $V_K(r)$ is the volume of the K-dimensional sphere of radius r, and we have [2, p. 305],

LEMMA 2.

$$V_K(r) = \frac{\pi^{K/2}}{\Gamma[(K+2)/2]} \cdot r^K.$$

Definition 6.

$$U_{K\sigma}(X_1, \cdots X_K) = \begin{cases} 1/V_K(a), & x_1^2 + \cdots + x_K^2 \leq a^2 \\ 0, & \text{otherwise} \end{cases}$$

where $a = (K + 2)^{\frac{1}{2}}\sigma$.

Thus $U_{K\sigma}$ is the K-dimensional uniform probability density function with variance K_σ^2.

[7] See standard texts on probability and statistics for definition of terms.

Assumption. Throughout the remainder of this paper it will be assumed that:

1. The measurement errors, $x - X$, which are governed by g, are independent of X.
2. f is either uniform, or normal with covariance matrix (σ_{ij}).
3. g is either uniform, or normal with covariance matrix $(\sigma_{Eij}) = \sigma_E^2 I$. Thus we assume that the measurement errors are uncorrelated and have equal variance.

Definition 7. Let

$$C_K = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_1^2 + \cdots x_K^2)^{k/2} N_{K1}(x_1, \cdots, x_K)\, dx_1 \cdots dx_K .$$

The following lemma is well known.[8]

LEMMA 3. $C_K = K!2^{(K/2)-1}/\Gamma[(K+2)/2]$.

3.1 *The Dimensional Effect.* We shall use the notation

$$x = (x_1, \cdots x_K), \qquad |x|^2 = x_1^2 + \cdots + x_K^2, \qquad dx = dx_1 \cdots dx_K, \quad \text{etc.}$$

We will first treat the case in which the X_i are uncorrelated and identically distributed.

THEOREM 7. *If* $(\sigma_{ij}) = \sigma^2 I$, $(\sigma_{Eij}) = \sigma_E^2 I$, $\sigma_E \ll \sigma$, and $f = N_{K\sigma}$ or $f = U_{K\sigma}$ and $g = N_{K\sigma_E}$ or $g = U_{K\sigma_E}$ then

$$F_K(f, g) \simeq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f^2(X)\, dX \cdot \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} V(|u|) g(u)\, du.$$

PROOF. Since $\sigma_E \ll \sigma$, we can make the approximation[9] $f(X) \simeq f(z)$ in step 3 below because the function f tends to remain relatively constant in the range: $|z - X| \leqq 3K\sigma_E$.

$$\begin{aligned}
F_K(f, g) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(X) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x - X) \underset{|z-x| \le |x-X|}{\int \cdots \int} f(z)\, dz\, dx\, dX \\
&\simeq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(X) \underset{|x-X| \le 3(K)^{\frac{1}{2}}\sigma_E}{\int \cdots \int} g(x - X) \underset{|z-x| \le |x-X|}{\int \cdots \int} f(z)\, dz\, dx\, dX \\
&\simeq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(X) \underset{|x-X| \le 3(K)^{\frac{1}{2}}\sigma_E}{\int \cdots \int} g(x - X) \underset{|z-x| \le |x-X|}{\int \cdots \int} f(X)\, dz\, dx\, dX \\
&\simeq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(X) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x - X) \underset{|z-x| \le |x-X|}{\int \cdots \int} f(X)\, dz\, dx\, dX \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(X) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x - X) V_K(|x - X|) f(X)\, dx\, dX \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f^2(X)\, dX \cdot \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} V_K(|x - X|) g(x - X)\, dx \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f^2(X)\, dX \cdot \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} V_K|\mu| g(u)\, du \qquad \text{Q.E.D.}
\end{aligned}$$

[8] The proof is given in the Appendix of [4]. See [3, p. 234].

[9] This particular approximation is used here for ease in presentation. A better approximation is used in [4, App., Th. A11].

THEOREM 8. *If* $(\sigma_{ij}) = \sigma^2 I$, $(\sigma_{Eij}) = \sigma_E^2 I$, *then*

$$(1)\quad \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} N_{K\sigma}^2(X)\,dX = 1/\pi^{K/2}2^K\sigma^K,$$

$$(2)\quad \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} U_{K\sigma}^2(X)\,dX = 1/V_K(a), \quad \text{where} \quad a = (K+2)^{\frac{1}{2}}\sigma,$$

$$(3)\quad \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} V_K(|u|)N_{K\sigma_E}(u)\,du = \frac{K!\,\pi^{K/2}}{\Gamma^2[(K+2)/2]}\,2^{(K/2)-1}\sigma_E^{\,K},$$

$$(4)\quad \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} V_K(|u|)U_{K\sigma_E}(u)\,du = \frac{1}{2}\,V_K(a_E),$$

$$\text{where} \quad a_E = (K+2)^{\frac{1}{2}}\sigma_E.$$

The proofs, which are straightforward, are given in [4, Appendix].

Using Theorems 7 and 8, one can approximate $F_K(f, g)$ for various combinations of normal and uniform distributions, under the hypotheses of Theorem 7. These are shown in Table I.

Table II gives approximate values for the ratio $F_K(f, g)/F_{K-1}(f, g)$ for the same combinations when K is large.

From Tables I and II it can be seen that, under the hypothesis of Theorem 7 and in the normal case,

$$\begin{aligned} F_K &\simeq \frac{K!}{\Gamma^2[(K+2)/2]2^{(K/2)+1}}\left(\frac{\sigma_E}{\sigma}\right)^K \\ &\simeq F_{K-1}\cdot 2^{\frac{1}{2}}\,\frac{\sigma_E}{\sigma} \\ &\simeq F_{K-1}\cdot\pi F_1, \end{aligned}$$

where F_1 is the one-dimensional result, $F_1 = F(N_\sigma, N_{\sigma_E})$.

The results of Tables I and II can be easily extended to cover the case in which f and g are normal, the matrix (σ_{ij}) is diagonal (no correlation), but where the σ_i's are nonequal. The following result is then obtained, which is proved in [4, App., Th. A12]:

TABLE I. APPROXIMATE VALUES OF $F_K(f, g)$ FOR DIFFERENT FUNCTIONS f AND g, WHERE $\sigma_E \ll \sigma$

	$f = N_{K\sigma}$	$f = U_{K\sigma}$
$g = N_{K\sigma_E}$	$\dfrac{K!}{\Gamma^2[(K+2)/2]2^{(K/2)+1}}\left(\dfrac{\sigma_E}{\sigma}\right)^K$	$\dfrac{K!\cdot 2^{(K/2)-1}}{\Gamma[(K+2)/2](K+2)^{K/2}}\left(\dfrac{\sigma_E}{\sigma}\right)^K$
$g = U_{K\sigma_E}$	$\dfrac{(K+2)^{K/2}}{\Gamma[(K+2)/2]2^{K+1}}\left(\dfrac{\sigma_E}{\sigma}\right)^K$	$\dfrac{1}{2}\left(\dfrac{\sigma_E}{\sigma}\right)^K$

TABLE II. THE RATIO $F_K(f, g)/F_{K-1}(f, g)$ FOR DIFFERENT FUNCTIONS f AND g, WHERE $\sigma_E \ll \sigma$, AND K IS LARGE

	$f = N_{K\sigma}$	$f = U_{K\sigma}$
$g = N_{K\sigma_E}$	$2^{\frac{1}{2}} \frac{\sigma_E}{\sigma} \simeq \pi F_1$	$\frac{2}{e^{\frac{1}{2}}} \frac{\sigma_E}{\sigma} \simeq 2\left(\frac{2\pi}{e}\right)^{\frac{1}{2}} F_1$
$g = U_{K\sigma_E}$	$\left(\frac{e}{2}\right)^{\frac{1}{2}} \frac{\sigma_E}{\sigma} \simeq 2\left(\frac{\pi e}{6}\right)^{\frac{1}{2}} F_1$	$\frac{\sigma_E}{\sigma} \simeq 2F_1$

$$
\begin{aligned}
F_K &\simeq \frac{K!}{\Gamma^2[(K+2)/2]2^{(K/2)+1}} \left(\frac{\sigma_E}{\sigma_1} \cdot \frac{\sigma_{E2}}{\sigma_2} \cdot \cdots \cdot \frac{\sigma_{EK}}{\sigma_K}\right) \\
&\simeq F_{K-1} 2^{\frac{1}{2}} \frac{\sigma_{EK}}{\sigma} \qquad (1) \\
&\simeq F_{K-1} \cdot \pi F^{(K)},
\end{aligned}
$$

where $F^{(K)}$ is the one-dimensional result

$$F^{(K)} = F(N_{\sigma_K}, N_{\sigma_{EK}}).$$

We draw the conclusion that if we have a $(K-1)$-dimensional measurement system, $(X_1, X_2, \cdots X_{K-1})$, and we wish to add an additional noncorrelated measurement, X_K, then this new measurement will contribute only about $2^{\frac{1}{2}}\sigma_{EK}/\sigma_K$ to the reducing power of F_K. Or, put another way, there is not much point in adding measurements, X, whose sigma-ratio σ/σ_E is considerably less than $2^{\frac{1}{2}}$. This conclusion holds for the important and "natural" case of normal distributions.

The factor $2^{\frac{1}{2}}$ which creeps in with the factors σ_i/σ_{Ei} is called the "dimensional-effect" because it is the penalty we pay for trying to combine the measurements to get a K-dimensional reduction.

In case of uniform distributions this dimensional effect is reduced from $2^{\frac{1}{2}}$ to 1 (see Table II). However, it should be pointed out that a uniform distribution is quite unlikely in a real pattern recognition system, either for f or g. It is true that the measurements X_i can be transformed to $Y_i = T_i(X_i)$ to obtain new distributions. In this manner, for example, the function f might be converted from normal to uniform, but this transformation would have a marked influence on the measurement errors $x_i - X_i$ (i.e., would change the function g), and the resulting effect on $F(f, g)$ would probably be deleterious.

3.2 *Correlational Effect.*

The following theorem expresses the intuitive satisfying result that F_K remains essentially unchanged when we add additional measurements x's for which σ/σ_E is very small.

THEOREM 9. *If $(\sigma'_{ij}) = \sigma'^2 I$ is L-dimensional, $(\sigma''_{ij}) = \sigma''^2 I$ is $(K - L)$-dimensional,*

$$(\sigma_{ij}) = \begin{pmatrix} (\sigma'_{ij}) & 0 \\ 0 & (\sigma''_{ij}) \end{pmatrix} \textit{ is K-dimensional, } \sigma'_{Eij} = \sigma_E^2 I$$

is L-dimensional, $\sigma_{Eij} = \sigma_E^2 I$ is K-dimensional, $\sigma'' \ll \sigma_E \ll \sigma'$, $f = N_{K\sigma}$, $g = N_{K\sigma_E}$, $f' = N_{L\sigma'}$, $g' = N_{L\sigma'_E}$, then $F_K(f, g) \simeq F_K(f', g')$.

The proofs of Theorems 9 and 10 are given in [4, App.].

THEOREM 10. *If* $F = N_{K\sigma}$, $g = N_{K\sigma_E}$, $(\sigma_{Eij}) = \sigma_e^2 I$, $\sigma_{Eii} \ll \sigma_{ii}$, $-1 \ll \rho_{ij} \ll 1$ *where* $\rho_{ij} = \sigma'_{ij}/(\sigma_{ii}\sigma_{jj})^{\frac{1}{2}}$, *for each* $i \neq j$, *and if* $f' = N_{K\sigma'}$, *where*

$$\sigma'_{ij} = \begin{cases} \sigma_{ij} & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$

then

$$F_K(f, g) \simeq F_K(f', g)/(|\ \rho_{ij}\ |)^{\frac{1}{2}}.$$

This theorem states, in effect, that if measurements X_i are correlated, with correlation matrix (ρ_{ij}), we obtain an *increase* in F_K by the factor $1/(|\ \rho_{ij}\ |)^{\frac{1}{2}}$. Thus

$$F_K \simeq D \cdot C \cdot S$$

where D and S are given in Section 3.1, and $C = 1/(|\ \rho_{ij}\ |)^{\frac{1}{2}}$.

This expression for C is not very accurate when the correlation ρ_{ij} between *any* two values of X is large, because in this case the determinant $|\ \rho_{ij}\ |$ is very small.

If the correlation is high it is better to use the conventional factor-analysis technique to determine new coordinates in which there is no correlation, and then proceed by the methods of Section 3.1, to estimate the dimensional effect in the new coordinate system.

Suppose $f = N_{K\sigma}$, $g = N_{K\sigma_E}$, (σ_{ij}) is symmetric positive definite, and (σ_{Eij}) is diagonal positive definite,

$$(\sigma_{Eij}) = \begin{pmatrix} \sigma_{E1}^2 & & & & \\ & \sigma_{E2}^2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot & \\ 0 & & & & & \sigma_{EK}^2 \end{pmatrix}.$$

Let $\sigma'_{ij} = \sigma_{ij}/\sigma_{Ei}\sigma_{Ej}$, for $i, j = 1, K$ and let

$$\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_L^2 \geq 1 > \lambda_{L+1}^2 \geq \cdots \geq \lambda_K^2$$

be the eigenvalues (characteristic roots) of the matrix (σ'_{ij}).

Thus the normalized numbers $X_i' = X_i/\sigma_{Ei}$ and $x_i' = x_i/\sigma_{Ei}$ can be linearly transformed to a new space $Y_i = AX_i'$, $y_i = Ax_i'$ in which the Y_i and y_i are governed by the probability density functions $f' = N_{K\Lambda}$ and $g' = N_{KI}$, where

$$\Lambda = \begin{pmatrix} \lambda_1^2 & & & & & 0 \\ & \lambda_2^2 & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ 0 & & & & & \lambda_K^2 \end{pmatrix}.$$

We now consider the λ's in two groups:

$$\lambda_1^2 \geq \cdots \geq \lambda_L^2 \geq 1 \quad \text{and} \quad 1 > \lambda_{L+1}^2 \geq \cdots \geq \lambda_K^2.$$

To the first group we apply the results of Section 3.1 (formula (1)) and to the second group we apply Theorem 9 to obtain

$$F_K \simeq \frac{1}{2} \frac{L!}{\Gamma^2[(L+2)/2]2^{(L/2)+1}} \frac{1}{\lambda_1} \cdot \frac{1}{\lambda_2} \cdot \cdots \cdot \frac{1}{\lambda_L}. \tag{2}$$

Of course, the formula (2) will be more valid when the λ_i are bounded away from 1, but even without this restriction formula (2) should give a good estimate for F_K in many practical applications.

4. *Summary and Comments*

In Sections 3.1 and 3.2 formulas were derived for the effectiveness of a pattern recognition system in which K measurements $(x_1, x_2, \cdots, x_K)$ are made. The expression F_K gives the average fractional reduction in uncertainty that results from these measurements. We have endeavored to express F_K in terms of elementary statistics, so that the behavior of a large, multicategory system can be predicted from experiments on small subsystems, where the variances σ_i^2 and $\sigma_{E\,i}^2$ and the correlation matrix (ρ_{ij}) are determined.

These results were derived for recognition systems in which there are a large number of categories, but it is believed that something like the "dimensional-effect" is experienced in *any* recognition system where a series of measurements are used.

The present recognition scheme uses a K-dimensional *distance*, but other schemes could be used. If the measurements $(x_1, x_2, \cdots, x_K)$ are used in a *decision-tree* recognition scheme, then one is led naturally to integration over K-dimensional cubes instead of the K-dimensional spheres that are encountered in Section 3. In another paper [6] it is shown that for Symmetric normal densities f and g, the spherical schemes of the type used here, are more efficient than a whole class of other schemes, including the cubical ones. Thus the dimensional effect will be at least as large for a decision-tree scheme.

5. *Exact Formulas and Monte Carlo Evaluations*

In Section 3 we derived the approximation

$$F_K' = \frac{K!}{\Gamma^2[(K+2)/2]2^{K/2+1}} \left(\frac{\sigma_E}{\sigma}\right)^K \tag{3}$$

for the case when f and g are normal. A somewhat better approximation

$$F_K'' = \frac{K!\, 2^{K/2-1}}{\Gamma^2[(K+2)/2]} \left(\frac{\sigma\sigma_E}{2\sigma^2+\sigma_E^2}\right)^K = \left(\frac{1}{1+\frac{1}{2}\left(\frac{\sigma_E}{\sigma}\right)^2}\right)^K F'_K \tag{4}$$

is given in [4, App., Th. A11].

In Section 2 we derived the exact expression for F_K,

$$F_1 = \frac{1}{\pi} \operatorname{Arctan} (\sqrt{2}\cdot\sigma_E/\sigma) \tag{5}$$

in the case when f and g are normal and $K = 1$. We know of no such expression for $K > 1$. However, we have evaluated F_K by Monte Carlo simulations for certain values of K and several σ-ratios.

Figure 2a gives a comparison between the exact formula, (5), and the approximations (3) and (4), for $K = 1$. Figures 2b–d compares formulas (3) and (4) with

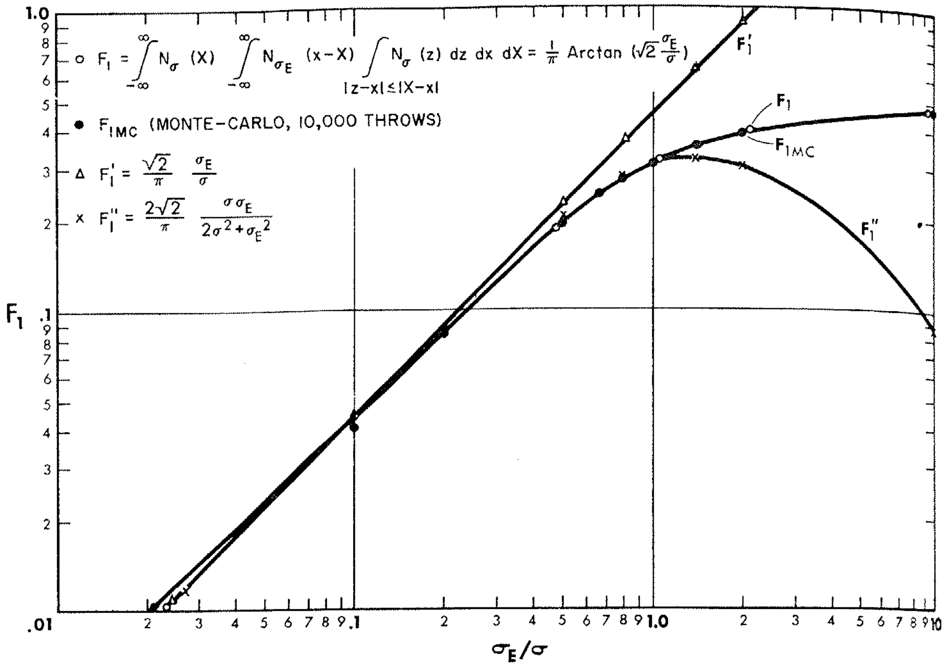

FIG. 2a. Comparison of the function F_1 with its approximations F_1', F_1'', and the Monte Carlo evaluation F_{1MC}

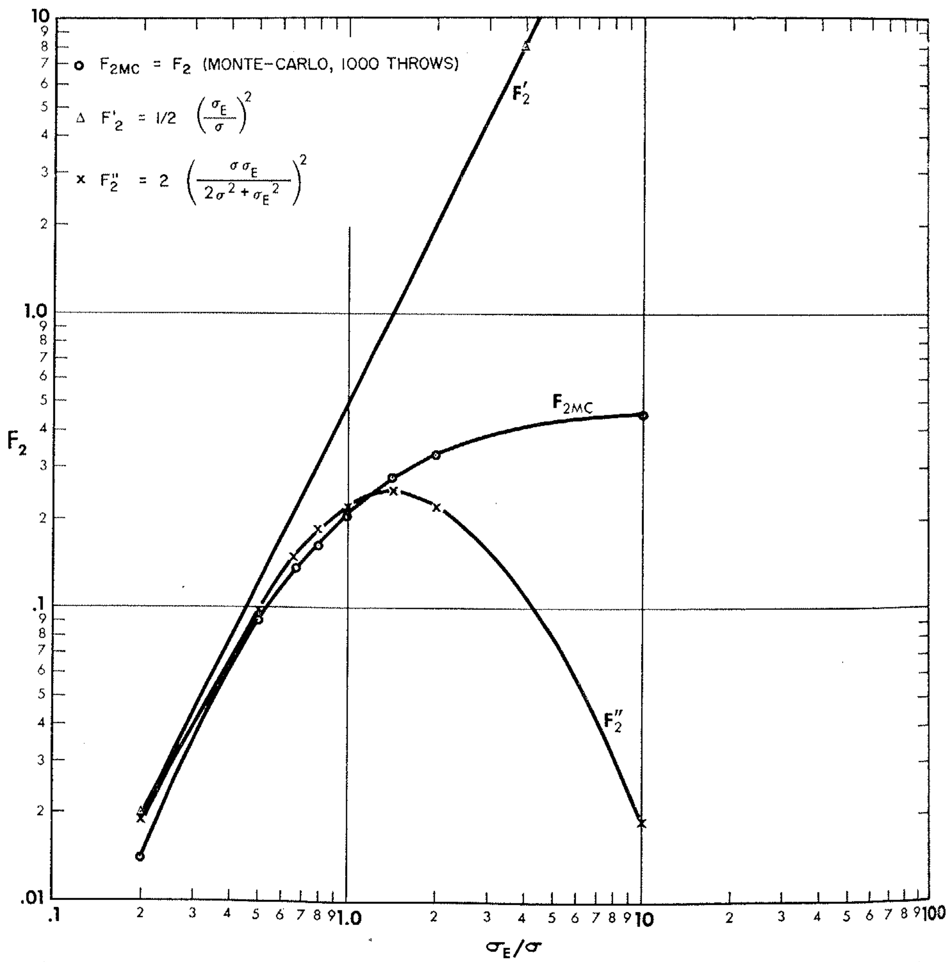

FIG. 2b. Comparison of the Monte Carlo evaluation F_{2MC} of F_2 with its approximations F_2' and F_2''

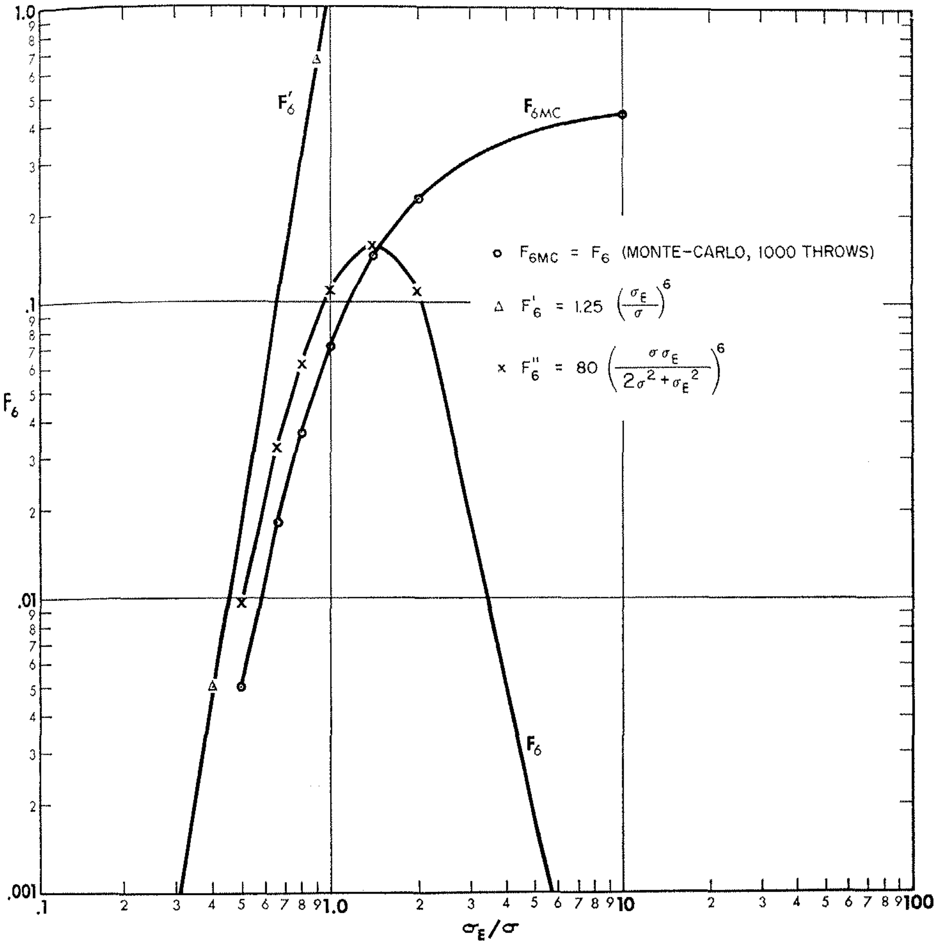

FIG. 2c. Comparison of the Monte Carlo evaluation F_{6MC} of F_6 with its approximations F_6' and F_6''

the Monte Carlo evaluations of F_K for $K = 2, 6, 10$. In these calculations f and g are normal with no correlation. At least 1000 throws were used in every Monte Carlo calculation, and 13,000 were used for certain crucial points, as indicated on the figures.

It appears, from a cursory look, that the "dimensional-effect" is a phenomenon peculiar to recognition systems utilizing a K-dimensional *distance*. However, it appears that a similar effect is also present in systems which utilize the measurements $(x_1, \cdots, x_K)$ in a *decision tree*.

It would be desirable, in the case where f and g are normal, to obtain an exact formula for $F_K(f, g)$ analogous to

$$F_1 = \frac{1}{\pi} \text{Arctan } (2^{\frac{1}{2}}\sigma_E/\sigma)$$

given when $K = 1$, but to date no such expression is known for F_K.

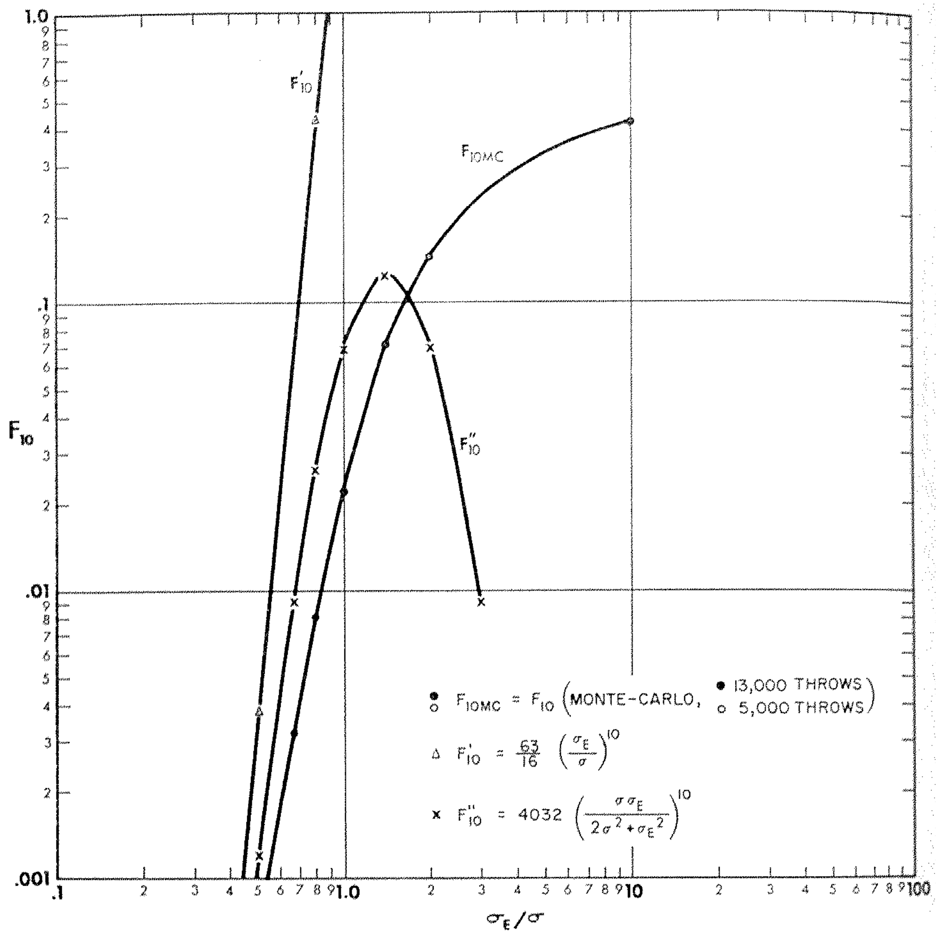

FIG. 2d. Comparison of the Monte Carlo evaluation F_{10MC} of F_{10} with its approximations F_1' and F_{10}''

Figure 2 gives a comparison of this exact formula and its approximations for $K = 1$.

RECEIVED MARCH, 1965; REVISED DECEMBER 1965

REFERENCES

1. SHANNON, CLAUDE E., AND WEAVER, WARREN. *The Mathematical Theory on Communication*. U. of Illinois Press, 1963.
2. COURANT, RICHARD. *Differential and Integral Calculus*, Vol. II. Interscience Publishers, Inc., New York.
3. CRAMÉR, HAROLD. *Mathematical Methods of Statistics*. Princeton U. Press, 1946.
4. BLEDSOE, W. W. Some results on multi-category pattern classification. Panoramic Res. Publ. PRI:16, May 1965.
5. GUPTA, S. S. Probability integrals of multivariate normal and multivariate *t*. *Ann. Math. Stat.* (1963), 792–838.
6. BLEDSOW, W. W. Optimum searching schemes. (Unpublished manuscript)