COMPUTERS AND BIOMEDICAL RESEARCH 17, 352–361 (1984)

Multivariate Time Series Discrimination in the Spectral Domain

ROBERT R. RAWLINGS,* MICHAEL J. ECKARDT,† AND HENRI BEGLEITER‡

*Division of Biometry and Epidemiology and †Laboratory of Clinical Studies, Intramural Research Program, National Institute on Alcohol Abuse and Alcoholism, 5600 Fishers Lane, Rockville, Maryland 20857, and ‡Department of Psychiatry, Downstate Medical Center, State University of New York, Brooklyn, New York 11203

Received December 13, 1983

Multivariate time series discrimination for evoked potentials from two cognitively different populations is discussed. A method for constructing linear and quadratic discriminant functions in the spectral domain using a stepwise frequency selection procedure is described and applied. Further, the effects of rank transformations on the estimated nonerror rates are examined. Finally, the validity of the signal plus stationary noise model (SSN) is examined. The performance of the SSN model is compared with the performance of a model (SNN) which consists of a signal plus nonstationary noise and a model (Mix) which consists of mixtures of subpopulations.

In this paper we will discuss multivariate time series discrimination for evoked potentials (*1*) from two different populations. The approach that we will use is that of spectral analysis as described in Shumway (*2–4*) when transient signals are imposed upon stationary noise series. An important aspect of the time series discrimination problem is the selection of variables to be used for classification of future observations. We will present methods for variable selection for the linear and quadratic discriminant functions which are based upon certain likelihood ratio tests. Another aspect of the time series discrimination that must be considered is the dependence of results on the assumption of multivariate normality. We will examine the use of rank-transformed data as a method of obtaining time series that are more nearly Gaussian. The use of rank transformations in discriminant analysis has been discussed by Conover and Iman (*5*).

Finally, we will compare the transient signal imposed upon a stationary noise model (SSN) described above with two other models. Gersch and Brotherton (*6*) concluded that noise is not stationary or uniform over subjects and that the average evoked potential is not as important in discrimination as the noise. We will examine a signal embedded in a nonstationary noise model (SNN) by the classical method of discrimination on the vector of observations collected over time (*7*). The third model (Mix) is based on the assumption that each population is a mixture of an unknown number of subpopulations. The approach that we

0010-4809/84 $3.00

will use to evaluate this model is nonparametric discrimination on the vector of observations collected over time. The nonparametric discrimination method used will be the kernel density method as described by Habbema *et al.* (*8*).

Mathematical Background

Consider the p-dimensional multivariate time series model

$$\mathbf{x}_{jl}(t) = \mathbf{u}_j(t) + \mathbf{n}_{jl}(t)$$

where $t = 0, \ldots, T-1; j = 1, \ldots, q; l = 1, \ldots, N_j$ such that q is the number of groups and N_j denotes the number of independent time series in the jth group. $\mathbf{n}_{jl}(t)$ are zero-mean stationary Gaussian processes with the cross-correlation matrix $R_j(t-u) = [\gamma_j^{mn}(t-u); m, n = 1, \ldots, p]$ and $\mathbf{u}_j(t)$ denotes the evoked potential for group j. The spectral density matrix, $F_j(\lambda)$, for group j is def ~d by the representation $R_j(t-u) = (1/2\pi)\int_{-\pi}^{\pi} e^{i\lambda(t-u)}F_j(\lambda)d\lambda$. The multivariate finite Fourier transform (FFT) of $\mathbf{x}(t)$ is defined by $\hat{\mathbf{x}}(k) = (1/\sqrt{T})\sum_{t=0}^{T-1}\mathbf{x}(t)e^{-i\lambda_k t}$ where $\lambda_k = 2\pi k/T$ for $k = 0, \ldots, T-1$. When the FFT is applied to the above time series model, we obtain $\hat{\mathbf{x}}_{jl}(k) = \hat{\mathbf{u}}_j(k) + \hat{\mathbf{n}}_{jl}(k)$. From Hannan (*9*), it is known that $\hat{\mathbf{n}}_{jl}(k)$ has a multivariate complex normal distribution, $N(0, F_j(\lambda))$, and for large T the vectors $\hat{\mathbf{n}}_{jl}(k)$ and $\hat{\mathbf{n}}_{jl}(k')$ are independent for $k \neq k'$. Hence, the above model is the complex analog of the general linear model problem, and Goodman (*10*) and Khatri (*11*) derive the usual likelihood ratio tests. Shumway (*2*) discussed the use of the likelihood ratio tests. $L(\lambda_k)$ and $L'(\lambda_k)$, for testing the equality of group means at each frequency as well as testing the equality of group spectral matrices at each frequency, respectively. In Chang (*12*), the likelihood ratio test, $L''(\lambda_k)$, is derived for testing the hypothesis of homogeneity of complex multivariate normal populations.

$$L''(\lambda_k) = \frac{n^{pn}\prod_{j=1}^{q}|C_j|^{n_j}}{\prod_{j=1}^{q} n_j^{pn_j}\left|C + \sum_{j=1}^{q} N_j(\hat{\mathbf{x}}_{j\cdot}(k) - \hat{\mathbf{x}}_{\cdot\cdot}(k))(\hat{\mathbf{x}}_{j\cdot}(k) - \hat{\mathbf{x}}_{\cdot\cdot}(k))^*\right|}$$

where

$$C_j = \sum_{l=1}^{N_j}(\hat{\mathbf{x}}_{jl}(k) - \hat{\mathbf{x}}_{j\cdot}(k))(\hat{\mathbf{x}}_{jl}(k) - \hat{\mathbf{x}}_{j\cdot}(k))^* \quad \text{and} \quad C = \sum_{j=1}^{q} C_j.$$

By using the expression for the hth moment of $L''(\lambda_k)$ given in (*12*) and by using Box's asymptotic expansion, as in Anderson (*7*), we obtained the distribution for the test statistic. The distribution function is given by

$$P(-2\rho \ln L'' \leq z) = G_f(z) + w_2(G_{f+4}(z) - G_f(z)) + O(n^{-3})$$

where

$$\rho = 1 - \left(\sum_{j=1}^{q} \frac{1}{n_j} - \frac{1}{n}\right) \frac{2p^2 + 3p - 1}{6(q-1)(p+3)} + \frac{1}{n}\left(\frac{p-q+2}{p+3}\right)$$

$$w_2 = \frac{p}{1152\rho^2}\left[6\left(\sum_{j=1}^{q} \frac{1}{n_j^2} - \frac{1}{n^2}\right)(p+1)(p-1)(p+2) - \left(\sum_{j=1}^{q} \frac{1}{n_j} - \frac{1}{n}\right)^2 \frac{(2p^2+3p-1)^2}{(q-1)(p+3)} - 12\left(\sum_{j=1}^{q} \frac{1}{n_j} - \frac{1}{n}\right) \frac{(2p^2+3p-1)(p-q+2)}{n(p+3)} - \frac{36(q-1)(p-q+2)^2}{n^2(p+3)} - \frac{12(q-1)}{n^2}(-2q^2 + 7q + 3pq - 2p^2 - 6p - 4)\right]$$

$$f = p^2(q^2 - 1) + p(2q - 2)$$

and $G_f(z)$ is the cumulative distribution function for a chi-square distribution with f degrees of freedom.

Now define

$$F_T(\lambda_k) = \frac{1}{n}\sum_{j=1}^{q} n_j F_{jT}(\lambda_k)$$

as the pooled estimator of the spectral density matrix where

$$F_{jT}(\lambda_k) = \frac{1}{n_j}\sum_{l=1}^{N_j} (\hat{\mathbf{x}}_{jl}(k) - \hat{\mathbf{x}}_{j\cdot}(k))(\hat{\mathbf{x}}_{jl}(k) - \hat{\mathbf{x}}_{j\cdot}(k))^*$$

is the estimate of the spectral density matrix for the jth population. When we construct the quadratic discriminant function, the estimate of the density function for the lth time series in group j at frequency k is given by

$$f_j(\hat{\mathbf{x}}_{jl}(k)) = \pi^{-p}|F_{jT}(\lambda_k)|^{-1} e^{-(\hat{\mathbf{x}}_{jl}(k) - \hat{\mathbf{x}}_{j\cdot}(k))^* F_{jT}^{-1}(\lambda_k)(\hat{\mathbf{x}}_{jl}(k) - \hat{\mathbf{x}}_{j\cdot}(k))}.$$

When we construct the linear discriminant function, we obtain the same expression with $F_{jT}(\lambda_k)$ replaced by $F_T(\lambda_k)$. The density estimate for frequencies $k_1, \ldots, k_m$ is given by $f_j(\hat{\mathbf{x}}_{jl}(k_1)) \ldots f_j(\hat{\mathbf{x}}_{jl}(k_m))$ since different k's have independent statistics. The posterior probability for assigning time series x to group i is given by

$$P(i|x) = \frac{\pi_i f_i(x)}{\sum_{j=1}^{q} \pi_j f_j(x)}$$

where π_i, denotes the a priori probability for membership in group i, and $f_i(x)$ denotes the estimate of the density function in group i for a specified set of frequencies. We shall obtain resubstitution and leaving-one-out estimates of the nonerror rates. In order to select frequencies for the linear discriminant function, compute $L(\lambda_k)$ for each λ_k. Since the statistics are uncorrelated at the different frequencies, we need only select the m frequencies corresponding to the m smallest values of $L(\lambda_k)$. Likewise, in order to select frequencies for the quadratic discriminant function, compute $L''(\lambda_k)$ for each λ_k and select the m frequencies corresponding to the m smallest values of L''. Only a small value of m should be chosen; otherwise, the frequency selection process can lead to biased estimates of the nonerror rates. The selection of an appropriate value for N_j is dependent upon the value of p. From empirical runs in the present study, it appears that we can obtain reliable estimates of the nonerror rates for $N_j \geq 7p$ for the quadratic discriminant function. The linear discriminant function would be expected to require fewer observations.

Consider the T-dimensional multivariate normal vector of random variable, $\mathbf{Y}_{jl}$, where $j = 1, \ldots, q$ and $l = 1, \ldots, N_j$ are defined as above. The SNN model does not assume that the covariance matrices, Σ_j, possess the special form required for a stationary process. The linear form of the SNN model assumes that $\Sigma_1 = \cdots = \Sigma_q$ while the quadratic form of the SNN model does not make this assumption. The Mix model assumes that an observation, $\mathbf{Y}_{jl}$, from group, j, has arisen from a population consisting of a mixture of distributions. The BMDP7M linear discriminant analysis program (*13*) is used for variable selection for the linear SNN model. For the quadratic SNN model, the variable selection method, as described in Rawlings *et al.* (*14*), is used. The variable selection procedure used for the Mix model is the procedure described in (*8*) for the kernel discriminant method.

Experimental Design

Investigators studying human brain potentials called event-related potentials (ERPs) have reported that the amplitude and/or latency of the P300 component is related to the "significance" or "utility" a stimulus has for the subject (*15*, *16*).

We designed a P300 study in the visual modality, with task difficulty determined by the complexity of processing equally physically deviant stimuli. The subject was seated in a sound-attenuated chamber facing a computer-controlled display (CRT), with his head resting on an adjustable chin rest. He was told to look at a fixation point displayed in the center of the screen. The experimental design consisted of a visual head orientation task. The nontarget stimulus was a frequently occurring circle presented in the center of the CRT, to which the subject did not respond. The target stimulus was an aerial view of the head with the nose and only one ear drawn in on either the right or left side; the subject pressed the corresponding button indicating whether a right or left ear was present as quickly as possible (reaction time). In the "easy" condition, the head was facing forward (nose up on screen), and left or right ear appeared directly on the side corresponding to the appropriate button. In the "hard" condition, the head was facing back (nose down on screen), and either the left or right ear appeared on the opposite side of the screen to the corresponding

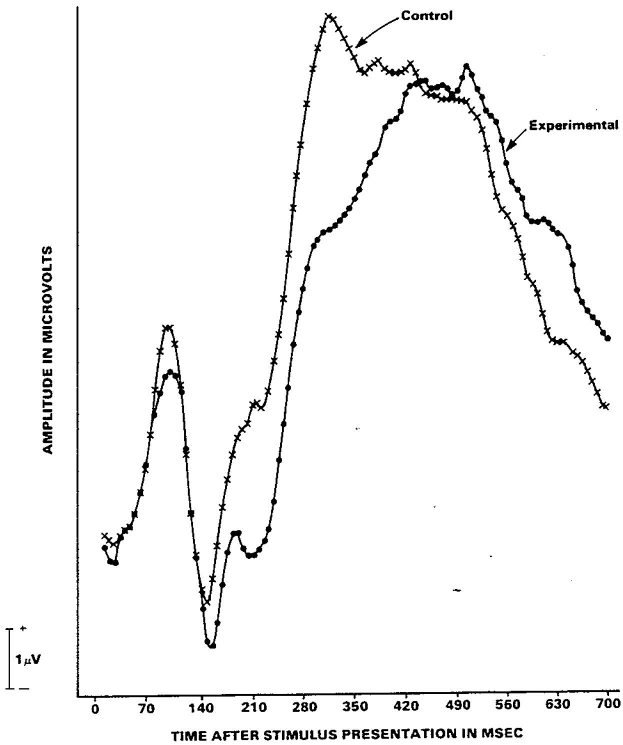

FIG. 1. Averaged event-related potentials recorded from a midline parietal location for control ($N = 24$) and experimental ($N = 23$) populations. CHAN = Pz.

button. A total of 240 stimuli were randomly presented—160 nontargets and 80 targets (20/target condition). The stimuli were 25 msec in duration and subtended 2.9° of arc; interstimulus intervals varied randomly between 2 to 4 sec.

Monopolar ERPs were recorded from midline frontal (Fz), central (Cz), parietal (Pz), and occipital (Oz) scalp leads. The linked ears served as reference and the nasion served as ground. ERPs were sampled by a PDP 11/40 computer for 49 msec preceding the stimulus (baseline) and for 700 msec (142 points/sec; bandwidth 0.01–100 Hz) following the stimulus. The prestimulus baseline voltage level was subtracted from each ERP recording at each electrode.

APPLICATION

We shall consider two groups of subjects ($N_1 = 24$, $N_2 = 23$) who were suspected of being cognitively different and for whom we have obtained average evoked potentials (over 20 repetitions). For brevity of presentation, we shall restrict our discussion to the results obtained for the four-lead and single-lead analyses for the "hard left ear" stimulus condition. An example of the population-averaged potentials for Pz is shown in Fig. 1.

TABLE I

SPECTRAL DOMAIN RESULTS

Location	Quadratic discriminant function		Linear discriminant function	
	Resubstitution nonerror rate	Leaving-one-out nonerror rate	Resubstitution nonerror rate	Leaving-one-out nonerror rate
Fz, Cz, Pz, Oz	100	81.0	79.0	70.0
	k = 18, 44, 27, 10		k = 3, 19, 48	
Fz	69.9	67.8	74.4	72.2
	k = 13, 15, 26		k = 3, 17, 18	
Cz	61.1	61.1	70.1	63.7
	k = 44, 23, 40		k = 3, 18, 1	
Pz	65.6	63.4	68.1	66.0
	k = 32		k = 10	
Oz	74.5	72.5	66.0	64.0
	k = 13, 1		k = 1, 10	

Note. The k values indicate the frequencies used in constructing the discriminant functions in units of 1.42 Hz.

In Table I, we present the results for the linear and quadratic discriminant analyses. In order to arrive at each entry in Table I, we obtained a test of equal means, a test for equal spectral matrices, and a test for equal populations for each frequency. Using the selected frequencies, the resubstitution and leaving-one-out nonerror rates were determined at each step of a forward-stepping procedure. By observing the number of variables at which the leaving-one-out estimate began to deteriorate, or diverge from the resubstitution estimate, a decision was made in each case as to the number of frequencies to include.

It can be seen that the four-lead analysis gave the best nonerror rate with the quadratic discriminant function. This result was anticipated since, when we performed the tests for equal spectral matrices, we found significant group differences at many frequencies. In the single-lead analysis, the linear discriminant function was superior in three out of four cases.

In order to examine the dependence of the results on distributional deviations from normality, we next present in Table II the same analyses as in Table I, except that we first have performed a rank transformation on the observed time series.

It can be seen that the four-lead analysis again gave the best nonerror rate with the quadratic discriminant function. However, the nonerror rate was smaller than in the analysis with the untransformed data. In the single-lead analysis, the linear discriminant function was again superior in three out of four cases, and for the same leads. When comparing the leaving-one-out analyses using untransformed data with the analyses using rank-transformed data, it was found that rank transformation decreased the nonerror rate in five cases, increased the nonerror rate in four cases, and maintained the same results in one

TABLE II

SPECTRAL DOMAIN RESULTS WITH RANK DATA

	Quadratic discriminant function		Linear discriminant function	
Location	Resubstitution nonerror rate	Leaving-one-out nonerror rate	Resubstitution nonerror rate	Leaving-one-out nonerror rate
Fz, Cz, Pz, Oz	95.7	76.5	83.0	76.0
	$k = 0, 3, 10$		$k = 3, 15, 18$	
Fz	76.4	70.0	72.2	72.2
	$k = 15, 3$		$k = 3, 15$	
Cz	61.3	59.2	76.5	68.1
	$k = 26$		$k = 15, 25, 3$	
Pz	69.8	63.6	68.3	63.9
	$k = 3$		$k = 0, 3$	
Oz	72.5	70.0	66.0	63.9
	$k = 13, 1$		$k = 1, 10$	

case. It also should be noted that the rank transformation often resulted in different frequencies being selected. In particular, higher frequencies tended not to be selected.

We now consider the problem of model validation. We will first analyze the results from the SNN and Mix models in the time domain, after which we shall compare these results with the results of the SSN model in the spectral domain. Because of the large number of time points involved, we will only analyze the four single-lead problems. Presented in Table III are results of the linear and quadratic discriminant analyses for both the untransformed and rank-transformed data, as well as the results of the kernel discriminant function analyses. In all cases, we have selected only the first two time points, which were obtained by the appropriate variable selection process, because of the correlations among the variables at different time points. The results of the analyses with the untransformed data show that the quadratic discriminant function had the larger nonerror rate in three out of four cases. The results of the analyses with the rank-transformed data show, however, that the linear discriminant function had the larger nonerror rate in three out of four cases. It is also seen that the kernel discriminant function was not superior for any of the leads. When we compare the spectral domain (Table I) and time domain (Table III) results for the untransformed data, we find that for the linear discriminant functions the SNN model was superior to the SSN model in three out of four cases. The results for the quadratic discriminant functions show that the SSN model was superior in two out of four cases, inferior in one case, and had the same nonerror rate in one case. When we make similar comparisons for the rank-transformed data, we find that for the linear discriminant functions the SNN model was again superior to the SSN model in three out of four cases. The results for the quadratic discriminant functions show that the SNN model

TABLE III

TIME DOMAIN RESULTS

Location	Linear	Quadratic	Rank linear	Rank quadratic	Kernel
Fz	63.8	72.2	70.2	74.8	—
	63.8	67.8	66.0	68.9	63.8
	T = 105, 210	T = 462, 343	T = 350, 448	T = 441, 483	T = 175, 294
Cz	68.1	59.2	72.3	66.2	—
	66.0	50.7	72.3	64.1	66.0
	T = 567, 693	T = 336, 329	T = 588, 686	T = 567, 686	T = 336, 266
Pz	70.2	76.6	70.2	70.1	—
	70.2	76.6	70.2	65.9	72.3
	T = 189, 322	T = 315, 392	T = 322, 455	T = 315, 273	T = 301, 343
Oz	74.5	80.8	76.6	78.7	—
	68.1	72.3	74.5	72.3	70.2
	T = 315, 546	T = 315, 441	T = 322, 434	T = 322, 364	T = 329, 63

Note. The upper number is the resubstitution nonerror rate, and the lower number is the leaving-one-out nonerror rate. The values of T given are the time points, in milliseconds, of the processes which were selected.

was also superior to the SSN model in three out of four cases. When we compare the best nonerror rates for each lead, we find that the SNN model was superior in three out of four cases, the SSN model was superior in one case, and the Mix model was not superior in any case. From Table III, it can be seen that the linear discriminant model was superior for leads Cz and Oz, while the quadratic discriminant model was superior for leads Fz and Pz. These results imply that both the average evoked potentials and the noise can contribute to the discrimination for the SNN model. The Mix model performed reasonably well for all of the leads, although it did not result in superior nonerror rates. In Rawlings *et al.* (*17*), the kernel method was shown to perform very well in some mixture problems, but larger numbers of subjects were utilized in that investigation. In the present study, there were substantial between-subject variations, and there were statistically significant differences for many between-subject comparisons. This would suggest that a mixture model would be a more appropriate model, although more complicated. However, the SSN and SNN models appear to be adequate for the present problem.

CONCLUSIONS

We have demonstrated in this paper easily implemented methods for constructing linear and quadratic discriminant functions in the spectral domain which provide useful nonerror rates for multiple-lead problems. It has also been shown that the nonerror rates obtained are reasonably close to the nonerror rates obtained by more complex time domain models. The analyses of the SSN and SNN models indicate that both the average evoked potentials and the noise can provide discrimination. The results of the analyses of the SSN and SNN models with rank-transformed data indicate that these data transformations can enhance discrimination in some problems. Close examination of the observations from individual subjects strongly suggests the presence of a number of subpopulations making the Mix model most appropriate. However, application of the Mix model did not result in superior nonerror rates in any case. The success of the simpler SSN and SNN models may be a result of the small numbers of subjects that we utilized. With additional subjects, the Mix model might prove to give better classification accuracy. However, in studies of this nature, large numbers of subjects are generally not available. The extension of the SNN and Mix models to multiple leads is difficult at present because of the large numbers of variables involved, while the SSN model is easily applied to multiple leads.

REFERENCES

1. GEVINS, A. S. Pattern recognition of human brain electrical potentials. *IEEE Trans. Pattern Anal. machine intelligence* **PAMI-2**(5), 383 (1980).
2. SHUMWAY, R. H. Discriminant analysis for time series. *In* "Handbook of Statistics," Vol. 2, "Classification, Pattern Recognition and Reduction of Dimensionality." North-Holland, Amsterdam, 1982.
3. SHUMWAY, R. H. Applied regression and analysis of variance for stationary time series. *J. Amer. Stat. Assoc.* **65,** 1527 (1970).
4. SHUMWAY, R. H. On detecting a signal in N stationarily correlated noise series. *Technometrics* **13,** 499 (1971).
5. CONOVER, W., AND IMAN, R. The rank transformation as a method of discrimination with some examples. *Commun. Stat.-Theor. Methods,* **A9,** 465 (1980).
6. GERSCH, W., AND BROTHERTON, T. "Where Is the Information for Discrimination in Evoked Potentials?: Some Preliminary Observations," Proceedings, IEEE 1981 Frontiers of Engineering in Health Care, Engineering in Medicine and Biology Conference, Houston, Tex., (1981).
7. ANDERSON, T. W. "An Introduction to Multivariate Statistical Analysis." Wiley, New York, (1958).
8. HABBEMA, J., HERMANS, J., AND VAN DEN BROEK, K. A stepwise discriminant analysis program, using density estimation. *In* "Compstat 1974" (G. Grukman, Ed.), Proceedings in Computational Statistics, pp. 101–110. Physica-Verlag, Vienna, (1974).
9. HANNAN, E. J. "Multiple Time Series." Wiley, New York, (1970).
10. GOODMAN, N. R. Statistical analysis based on a certain multivariate complex distribution (an introduction). *Ann. Math. Stat.* **34,** 152 (1963).
11. KHATRI, C. G. Classical statistical analysis based on a certain multivariate complex Gaussian distribution. *Ann. Math. Stat.* **36,** 98 (1965).
12. CHANG, T. C., KRISHNAIAH, P. R., AND LEE, J. C. Approximations to the distribution of the likelihood ratio statistics for testing the hypotheses on covariance matrices and mean vectors simultaneously. *In* "Application of Statistics" (P. R. Krishnaiah, Ed.), pp. 97–108. North-Holland, Amsterdam, (1977).
13. DIXON, W. J. "BMDP Statistical Software." Univ. of California Press, Berkeley, (1981).
14. RAWLINGS, R., RAE, D., GRAUBARD, B., ECKARDT, M. J., AND RYBACK, R. S. A methodology for construction of a multivariate diagnostic instrument: An application to alcohol abuse screening. *Comput. Biomed. Res.* **15,** 228 (1982).

15. SQUIRES, N. K., DONCHIN, E., SQUIRES, K. C., AND GROSSBERG, S. Bisensory stimulation: Inferring decision-related processes from the P300 component. *J. Exp. Psychol.: Human Percept. Perform.* **3**, 299 (1977).
16. KUTAS, M., MCCARTHY, G., AND DONCHIN, E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science* **197**, 792 (1977).
17. RAWLINGS, R., GRAUBARD, B., RAE, D., ECKARDT, M., AND RYBACK, R. Two-group classification when both groups are mixtures of normals. *Biom. J.*, in press.