The Annals of Statistics
1974, Vol. 2, No. 2, 347–352

LARGE SAMPLE DISCRIMINATION BETWEEN TWO GAUSSIAN PROCESSES WITH DIFFERENT SPECTRA[1]

BY ULF GRENANDER

Brown University

We study the probability of error asymptotically for testing one Gaussian stochastic process against another when the mean vectors are zero and we have the choice between two given covariance matrices. It is shown that under certain conditions the probabilities of error form asymptotically a geometric progression with a ratio that is derived. The approach employs Laplace's method of approximating integrals and does not appeal to Fourier analysis; in this sense it can be said to be elementary.

1. Asymptotic error probability. In connection with the problem of optimal feature selection in pattern recognition one encounters the difficulty that the resulting probability of error is difficult to evaluate. This is so particularly when the distributions are Gaussian and the possible patterns correspond to different covariance matrices. D. McClure (1972) has recently obtained bounds for the probability of error and these bounds were decreasing geometrically as the sample size n tends to infinity.

In this paper we shall approach the problem from another point of view and show that the error probabilities themselves tend to zero geometrically and obtain an expression for the ratio in this geometric progression.

Let y be an n-vector with normal distribution, mean zero, and covariance matrices R_1 and R_2 under hypotheses H_1 and H_2 respectively. We want to test H_1 against H_2 in a Bayesian manner assuming the a priori probabilities π_1 and π_2. Let us deal with the case $\pi_1 = \pi_2 = \frac{1}{2}$.

Then, in the usual manner, if w is the critical region for rejecting H_1,

$$2P_n(\text{error}) = \int_w p_1\,dy + \int_{w^C} p_2\,dy = 1 + \int_w (p_1 - p_2)\,dy \tag{1}$$

where p_1, p_2 are the joint normal densities. The optimal w is given by

$$w = \{y \mid p_1 - p_2 < 0\} \tag{2}$$

or

$$w = \left\{y \,\middle|\, 1 < \left(\frac{\det R_1}{\det R_2}\right)^{\frac{1}{2}} \exp -\tfrac{1}{2} y^T[R_2^{-1} - R_1^{-1}]y\right\}. \tag{3}$$

Without loss of generality we can let $R_1 = I$ and R_2 be the diagonal matrix

$$R_2 = \operatorname{diag}(d_1, d_2, \cdots, d_n) = D. \tag{4}$$

Received October 1972; revised April 1973.

[1] Work partially supported by N.S.F. Grant GJ-710.

AMS 1970 *subject classification.* 62M99.

Key words and phrases. Pattern discrimination, error probability, stationary processes.

Then w is given by the inequality

$$y^T Qy = y^T[I - D^{-1}]y > \log \det D . \tag{5}$$

Let us study the integral

$$I_n = \int_w p_1\, dy \tag{6}$$

and note that this is the probability that

$$y^T Qy > \log \det D . \tag{7}$$

To achieve asymptotic results we must assume something about the asymptotic distribution of the d's and we shall do this by assuming that their distribution converges weakly to some distribution on an interval (m, M) and with the distribution function F; $0 < m < M < \infty$.

To begin with, let us assume, instead, that the (finite) distribution of d has k spikes only, so that

$$\begin{aligned} d_1 = d_2 = \cdots = d_{\rho_1 n} &= \lambda_1 \\ d_{\rho_1 n+1} = \cdots = d_{(\rho_1+\rho_2)n} &= \lambda_2 \\ &\cdots \\ d_{(\rho_1+\cdots+\rho_{k-1})n+1} = \cdots = d_{(\rho_1+\cdots+\rho_k)n} &= \lambda_k \end{aligned} \tag{8}$$

with positive ρ's such that $\rho_1 + \rho_2 + \cdots + \rho_k = 1$ and positive λ's with at least one of them different from 1. Then the entries of Q (in the main diagonal) are

$$q_i = 1 - \frac{1}{\lambda_j}, \qquad (\rho_1 + \cdots + \rho_{j-1})n < i \leqq (\rho_1 + \cdots + \rho_j)n \tag{9}$$

and

$$\log \det D = n \textstyle\sum_{i=1}^k \rho_i \log \lambda_i = n \cdot c . \tag{10}$$

Therefore we can write

$$I_n = P\{\textstyle\sum_{i=1}^k q_i S_i \geqq cn\} , \tag{11}$$

where the q_i denote the k distinct values on the diagonal of Q and where S_i is χ^2-distributed with $\rho_i n$ degrees of freedom.

Recalling the functional form of the χ^2-distribution, we have

$$I_n = C_n \int_{E_k} \textstyle\prod_{i=1}^k e^{-y_i/2} y_i^{(n_i/2)-1}\, dy \tag{12}$$

with $n_i = \rho_i n$, and where the region of integration is

$$E_k = \{y \mid y_1, y_2, \cdots, y_k \geqq 0, \textstyle\sum q_i y_i \geqq cn\} . \tag{13}$$

Also,

$$C_n^{-1} = \textstyle\prod_{i=1}^k \Gamma\left(\frac{n_i}{2}\right) 2^{n_i/2} . \tag{14}$$

Changing variables $y_i/n = u_i$, we get

$$I_n = C_n n^k n^{\Sigma(n_i/2-1)} \int_{F_k} \textstyle\prod_{i=1}^k \exp(-nu_i/2) u_i^{(n_i/2)-1}\, du \tag{15}$$

with

$$F_k = \{u \mid u_1, u_2, \cdots, u_k \geqq 0, \textstyle\sum q_i u_i \geqq c\} . \tag{16}$$

Putting

$$\phi(u) = \textstyle\prod_{i=1}^k e^{-u_i/2} u_i^{\rho_i/2} , \tag{17}$$

we can write

$$I_n = C_n n^{n/2} \textstyle\int_{F_k} \phi^{n-\alpha}(u) q(u)\, du \tag{18}$$

where

$$q(u) = \prod e^{-\alpha u_i/2} u_i^{-1+\alpha\rho_i/2}$$

is integrable $\alpha > 0$. Then

$$I_n^{1/n} = \frac{n^{\frac{1}{2}}}{2^{\frac{1}{2}}[\prod_1^k \Gamma(n_i/2)]^{1/n}} \{\textstyle\int_{F_k} \phi^{n-\alpha} q\, du\}^{1/n} . \tag{19}$$

Using Stirling's asymptotic expression for the Γ-function

$$\Gamma(\alpha + 1) \sim \left(\frac{\alpha}{e}\right)^\alpha (2\pi\alpha)^{\frac{1}{2}} , \tag{20}$$

we get

$$\left\{\prod_1^k \Gamma\left(\frac{n_i}{2}\right)\right\}^{1/n} \sim \left(n \frac{\prod_1^k \rho_i^{\rho_i}}{2e}\right)^{\frac{1}{2}} , \tag{21}$$

which takes care of the coefficient on the right-hand side of (19).

Let us consider the behavior of ϕ in the region F_k. The unconstrained maximum of ϕ occurs for $u_i = \rho_i$, $i = 1, 2, \cdots, k$, but this point does not belong to F_k since

$$\sum q_i \rho_i = \sum \rho_i \left(1 - \frac{1}{\lambda_i}\right) < \sum \rho_i \log \lambda_i = c . \tag{22}$$

The constrained maximum cannot be attained when one or more of the coordinates vanish, since then $\phi = 0$, and hence we only have the possibility left that it is attained on the simplex

$$G_k = \{u \mid u_1, u_2, \cdots, u_k > 0, \textstyle\sum q_i u_i = c\} . \tag{23}$$

Maximizing $\sum [-u_i + \rho \log u_i] + \mu[C - \sum q_i \mu_i]$ by the method of Lagrangian multipliers, we then get

$$\alpha_i = \frac{\rho_i}{1 + \mu q_i} \tag{24}$$

where μ should be chosen to satisfy the equality in (23) when α's are substituted for u's, so that if we introduce the function

$$\alpha(t) = \sum_i \frac{\rho_i q_i}{1 + t q_i} , \tag{25}$$

we should have

$$\alpha(\mu) = c . \tag{26}$$

The denominators in (24) must be positive so that t is contained in an interval (a, A) with

$$a = -\frac{1}{\max_i q_i^+} \tag{27}$$
$$A = -\frac{1}{\min_i q_i^-},$$

where the notation indicates the set of positive and negative q's respectively, if one of these sets is empty a or A should be replaced by $-\infty$ or $+\infty$ respectively. In the interval (a, A) the function $\alpha(t)$ defined in (25) is continuous and decreasing and an examination of the range of α shows that (26) has just one root which we denote by μ. Summing up we get

LEMMA 1. *The function ϕ attains its maximum over F_k in the point P with coordinates given by* (24), (25) *and* (26).

Denote the maximum by M,

$$M = \exp \tfrac{1}{2} \textstyle\sum_i (\rho_i \log \alpha_i - \alpha_i)\,. \tag{28}$$

We also note that the gradient of ϕ at P points outwards from F_k.

The reason we make our integral in the form (19) is that we can now apply the classical method of Laplace (see, e.g., Hardy, Littlewood and Pólya (1964), pages 134, 143, where p corresponds to our integrable function q). Combining (19), (21) and (22), we get

LEMMA 2. *For eigenvalue distributions of the type* (8) *we have*

$$\lim_{n\to\infty} I_n^{1/n} = \exp \tfrac{1}{2} \textstyle\sum_i \rho_i(\log a_i - a_i + 1) \tag{29}$$

with

$$a_i = \frac{1}{1 + \mu q_i} = \alpha_i/\rho_i\,. \tag{30}$$

To treat the integral

$$I_n' = \textstyle\int_{w^C} p_2\, dy \tag{31}$$

we transform the y-variables by $y_i/d_i^{1/2} = z_i$ so that in the z-variables

$$w^C = \{z \mid z^T[D - I]z < \log \det D\} \tag{32}$$

and z have the identity matrix as covariance. We can now proceed in the same way as before and get a similar expression

$$\lim_{n\to\infty} (I_n')^{1/n} = \exp \tfrac{1}{2} \textstyle\sum_i \rho_i(\log a_i' - a_i' + 1) \tag{33}$$

but with

$$a_i' = \frac{1 - q_i}{1 + \mu q_i}, \tag{34}$$

where μ is given by (26).

One simplification is possible observing that

$$\sum_i \rho_i(\log a_i - a_i + 1) - \sum \rho_i(\log a_i' - a_i' + 1)$$

$$(35) \qquad = \sum_i \rho_i\left[\log\left(\frac{a_i}{a_i'}\right) - (a_i - a_i')\right]$$

$$= \sum_i \rho_i\left[-\log(1 - q_i) - \frac{q_i}{1 + \mu q_i}\right].$$

Using the definition of μ in (25), (26) the sum reduces to

$$(36) \qquad -\sum_i \rho_i \log(1 - q_i) - c = \sum_i \rho_i \log \lambda_i - c = 0 .$$

Hence the two limits in (29) and (33) are identical, and we have proven

THEOREM 1. *Under the given assumption the probability of error decreases geometrically as the sample size tends to infinity and with the ratio*

$$(37) \qquad \lim_{n\to\infty} (P_n(\text{error}))^{1/n} = \exp \tfrac{1}{2} \sum_i \rho_i(\log a_i - a_i + 1) .$$

We can use this result to find the asymptotic decrease of the error probability when the distribution of

$$(38) \qquad 0 < m \leqq d_1^{(n)} \leqq d_2^{(n)} \leqq \cdots \leqq d_n^{(n)} \leqq M < \infty$$

converges weakly to a distribution over (m, M) with df F. We shall approximate the probability of the event in (7) from below and from above. We do this by shifting the $d_i^{(n)}$ a little to the left such that they form k spikes inside the interval (m, M). Let this affect only the left-hand side of (7); the right-hand side we leave as it is. In a similar way we shift the values to the right. The probabilities associated with the new quadratic forms replacing Q in (7) obviously form lower and upper bounds for the probability we study. Both of the bounds behave asymptotically in a way that is described by (29). For fixed, but arbitrarily large value of k, this gives us asymptotic bounds expressed through the ratios of geometric progressions. Since the a's depend in a continuous manner upon the q's we can make the difference between the two ratios arbitrarily small. This gives us the result we sought.

THEOREM 2. *If the eigenvalues $d_i^{(n)}$ are contained in the interval (m, M), $0 < m < M < \infty$, and their distribution converges weakly to the* df F *the optimal error probability* P_n(error) *satisfies*

$$(39) \qquad \lim_{n\to\infty} [P_n(\text{error})]^{1/n} = \exp \tfrac{1}{2} \int_m^M [\log a(x) - a(x) + 1]F(dx)$$

with

$$(40) \qquad a(x) = \frac{x}{x + \mu(x - 1)} ,$$

where μ should be determined from the equation

$$(41) \qquad \int_m^M \frac{x - 1}{x + \mu(x - 1)} F(dx) = \int_m^M \log x \cdot F(dx) .$$

It can be noted that the geometric ratio in (39) is strictly less than one unless $a(x) \equiv 1$ a.c. which is only possible if $x \equiv 1$ a.c. so that F has all its variation in $x = 1$.

2. Remarks. It is intuitively plausible that the role of the a priori probabilities π_1, π_2 is asymptotically negligible. This can be proven by noticing that if $\pi_1 \neq \pi_2$ and $\pi_1 \neq 0, 1$, the only change in (7) is the occurrence of an additional term $\log \pi_1 - \log \pi_2$ on the right-hand side. This term will not influence the asymptotic result, so that the theorems in the last section are valid in this case too.

We have scarcely exploited Laplace's method to its full strength. The problem was very well suited to it and we could apply it directly. It seems likely, however, that using a more refined analysis (see, e.g., Pólya and Szegö (1964), Volume 1, Chapter 5), we could get asymptotic expressions of the form

$$P_n(\text{error}) \sim B \cdot A^n . \tag{42}$$

The author had originally intended to use Laplace's method in the Fourier-domain before he observed that the frequency functions (and not only their Fourier transformed) could be written as high powers of a continuous function. This led to the above variation of an old theme in the theory of limiting distributions.

3. Acknowledgment. The author is indebted to D. McClure for his remarks and for reading the manuscript.

REFERENCES

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1964). *Inequalities.* Cambridge Univ. Press.

MCCLURE, D. (1972). Optimal and efficient finite linear representations of stationary time series. To appear.

PÓLYA, G. and SZEGÖ, G. (1964). *Aufgaben arnd Lehrsätze aus der Analysis.* Springer-Verlag, Berlin.

DEPARTMENT OF APPLIED MATHEMATICS
BROWN UNIVERSITY
PROVIDENCE, RHODE ISLAND 02912