Moments of the Interclass Mahalanobis Distance

R. C. GONZALEZ, SENIOR MEMBER, IEEE, AND C. G. WAGNER

Abstract—**It is shown that the moments of the interclass Mahalanobis distance between elements of two *d*-variate Gaussian populations can be**

Manuscript received May 22, 1983. This work was supported by the U.S. Department of Energy under contract with the Union Carbide Corporation, and in part by the Office of Naval Research, Arlington, VA.

R. C. Gonzalez is with the Electrical Engineering Department, the University of Tennessee, Knoxville, TN 37996.

C. G. Wagner is with the Department of Mathematics, the University of Tennessee, Knoxville, TN 37996.

0018-9472/83/1100-1135$01.00 ©1983 IEEE

expressed in a simple polynomial form. The nth moment is expressible as a polynomial of order n whose variable depends on the mean vectors and eigenvalues of the covariance matrices. A closed-form solution is given for computing the coefficients of the polynomial expressions.

I. Introduction

Pattern recognition and image processing techniques based on the Mahalanobis distance have found wide applicability, ranging from nuclear reactor surveillance and automated analysis of image texture data to discrimination problems in biomedical observations [1], [2], [3].

The importance of the Mahalanobis distance classifier lies in the fact that, under a Gaussian assumption, it is an optimal discriminant in the Bayes sense [4]. The estimation of the probability density function (pdf) of the interclass Mahalanobis distance has been a topic of active interest for a number of years because of its direct relation to the probability of error of Bayes' classifier [5]. For Gaussian data with equal covariance matrices, the solution of this problem is straightforward [6]. When the covariance matrices are not equal, however, the problem becomes considerably more complicated, requiring the use of numerical integration techniques for computing the pdf [7].

In many applications (e.g., cluster seeking, texture analysis, and measuring spatial stationarity of multivariate data) it is often of interest to compute the moments of the interclass Mahalanobis distance without having to estimate its underlying pdf as an intermediate step. It is shown in this paper that these moments can be expressed directly in terms of a polynomial whose coefficients are given by a straightforward closed-form expression. The relative simplicity of these results has important implications in terms of implementation in a digital computer.

II. Background

Consider two d-dimensional Gaussian vector populations $\{x\}$ and $\{y\}$ with mean vectors and covariance matrices m_x, m_y, C_x, and C_y, respectively. The *intraclass Mahalanobis distance*[1] between any member of $\{x\}$ and m_x is given by the familiar equation [1]

$$R(x, m_x) = (x - m_x)^T C_x^{-1}(x - m_x), \tag{1}$$

and, similarly,

$$R(y, m_y) = (y - m_y)^T C_y^{-1}(y - m_y), \tag{2}$$

where "T" indicates the transpose.

As indicated in the previous section, (1) and (2) have been applied extensively in pattern recognition. In this paper, we are interested in characterizing the *interclass Mahalanobis distance* between members of x and the mean m_y, which is given by

$$R(x, m_y) = (x - m_y)^T C_y^{-1}(x - m_y) \tag{3}$$

and similarly,

$$R(y, m_x) = (y - m_x)^T C_x^{-1}(y - m_x). \tag{4}$$

For any nonsingular, real transformation matrix A it is easily shown that if

$$r = Ax \tag{5}$$

and

$$s = Ay, \tag{6}$$

then r and s are Gaussian random variables with mean vectors

$$m_r = Am_x \tag{7}$$

$$m_s = Am_y \tag{8}$$

and covariance matrices

$$C_r = AC_xA^T \tag{9}$$

$$C_s = AC_yA^T. \tag{10}$$

It is also easily shown that

$$R(r, m_s) = R(x, m_y) \tag{11}$$

and

$$R(s, m_r) = R(y, m_x). \tag{12}$$

Furthermore, as described in [6] and [16], the transformation matrix A can be chosen so that

$$C_r = AC_xA^T = I \tag{13}$$

and

$$C_s = AC_yA^T = D, \tag{14}$$

where I is the identity matrix and D is a diagonal matrix with elements $\gamma(i)$, $i = 1, 2, \cdots, d$, along the main diagonal. The elements $\gamma(i)$ are the eigenvalues of $C_x^{-1}C_y$. From (13), it is noted that the elements of r are uncorrelated which, in view of our Gaussian assumption, implies statistical independence. The same holds true for the elements of s.

Using (3), (11), and (14), it follows that

$$\begin{aligned} R(x, m_y) &= R(r, m_s) \\ &= (r - m_s)^T D^{-1}(r - m_s) \\ &= \sum_{i=1}^{d} (r_i - m_{si})^2 \gamma^{-1}(i), \end{aligned} \tag{15}$$

where r_i and m_{si}, $i = 1, 2, \cdots, d$, are the components of vectors r and m_s, respectively. Since r is a Gaussian random vector and $C_r = I$, we have that the variable $z_i = (r_i - m_{si})$ is Gaussian with mean $(m_{ri} - m_{si})$ and unit variance. It then follows [9] that

$$w_i = z_i^2 = (r_i - m_{si})^2 \tag{16}$$

is a noncentral chi-square variable with density

$$p(w_i) = e^{-\lambda_i} \sum_{k=0}^{\infty} \frac{\lambda_i^k w_i^{(1+2k)/2} e^{-w_i/2}}{k! 2^{(1+2k)/2} \Gamma\left(\frac{1+2k}{2}\right)} \tag{17}$$

and moment generating function

$$\phi_{w_i}(t) = e^{-\lambda_i} \sum_{k=0}^{\infty} \frac{\lambda_i^k}{k!}(1 - 2t)^{-(1+2k)/2}, \tag{18}$$

where

$$\lambda_i = \frac{1}{2}(m_{ri} - m_{si})^2. \tag{19}$$

Since r_i, $i = 1, 2, \cdots, d$, are statistically independent, it follows that the w_i defined in (16) are also statistically independent.

A similar development can be carried out for $R(y, m_x)$:

$$\begin{aligned} R(y, m_x) &= R(s, m_r) \\ &= (s - m_r)^T I^{-1}(s - m_r) \\ &= \sum_{i=1}^{d} (s_i - m_{ri})^2. \end{aligned} \tag{20}$$

The variable $z_i = (s_i - m_{ri})/\sqrt{\gamma(i)}$ is Gaussian with mean $(m_{si} - m_{ri})/\sqrt{\gamma(i)}$ and unit variance. As above, the variable

$$w_i = z_i^2 = \frac{1}{\gamma(i)}(s_i - m_{ri})^2 \tag{21}$$

[1]This is in reality a *squared* distance. However, it has become customary to refer to this measure simply as the *Mahalanobis distance*.

has the density and moment generating function given in (17) and (18), but λ_i is now given by

$$\lambda_i = \frac{1}{2\gamma(i)}(m_{si} - m_{ri})^2. \tag{22}$$

III. Moments of the Interclass Mahalanobis Distance.

It is shown in this section that the moments about zero of w_i can be expressed in terms of λ_i. Once these moments have been obtained, they will be used to obtain the moments of the interclass Mahalanobis distance.

A. Moments about Zero of w_i

We begin the development of noting that the nth moment about zero of w_i is given by

$$\alpha_n(w_i) = E\{w_i^n\} = \phi_{w_i}^{[n]}(0) \tag{23}$$

where $\phi_{w_i}^{[n]}(0)$ is the nth derivative of (18) with respect to t, evaluated at $t = 0$[10]. Evaluating (23) with the moment generating function given in (18) leads to the following theorem.

Theorem 1: Let $\alpha_n(w_i)$ denote the nth moment about zero of w_i. Then

$$\alpha_n(w_i) = \sum_{r=0}^{n} c(n,r)\lambda_i^r \tag{24}$$

where

$$c(n,r) = 2^r\binom{n}{r}\prod_{j=r}^{n-1}(2j+1) \tag{25}$$

for all $n \geqslant 1$ and $0 \leqslant r \leqslant n-1$, and

$$c(n,n) = 2^n \tag{26}$$

for all $n \geqslant 0$.

Proof: From (18) and (23),

$$\alpha_n(w_i) = \phi_{w_i}^{[n]}(0)$$

$$= e^{-\lambda_i}\sum_{k=0}^{\infty}\frac{\lambda_i^k}{k!}\frac{\partial^n}{\partial t^n}(1-2t)^{-(1+2k)/2}\Big|_{t=0} \tag{27}$$

$$= e^{-\lambda_i}\sum_{k=0}^{\infty}\frac{\lambda_i^k}{k!}\prod_{s=1}^{n}(2k+2s-1) \tag{28}$$

$$= \sum_{j=0}^{\infty}\frac{(-1)^j\lambda_i^j}{j!}\sum_{k=0}^{\infty}\frac{\lambda_i^k}{k!}\prod_{s=1}^{n}(2k+2s-1) \tag{29}$$

$$= \sum_{r=0}^{\infty}\frac{\lambda_i^r}{r!}\sum_{m=0}^{r}(-1)^{r-m}\binom{r}{m}\prod_{s=1}^{n}(2m+2s-1), \tag{30}$$

where (30) follows from (29) by the standard rule for multiplying Taylor series.

By a well-known formula from the calculus of finite differences [15]

$$b(n,r) \equiv \sum_{m=0}^{r}(-1)^{r-m}\binom{r}{m}\prod_{s=1}^{n}(2m+2s-1)$$

$$= \Delta^r\prod_{s=1}^{n}(2x+2s-1)\Big|_{x=0} \tag{31}$$

where $\Delta^0 f(x) = f(x)$, $\Delta^1 f(x) = \Delta f(x) = f(x+1) - f(x)$, and

$$\Delta^r f(x) = \Delta(\Delta^{r-1}f(x)) \tag{32}$$

$$= \sum_{k=0}^{r}(-1)^{r-k}\binom{r}{k}f(x+k), \tag{33}$$

for $r \geqslant 2$.

Since $\prod_{s=1}^{n}(2x+2s-1)$ is a polynomial in x of degree n, it follows that $b(n,r) = 0$ if $r > n$, and hence from (30) that

$$\alpha_n(w_i) = \sum_{r=0}^{n} c(n,r)\lambda_i^r, \tag{34}$$

where

$$c(n,r) \equiv b(n,r)/r!. \tag{35}$$

Now

$$\prod_{s=1}^{n}(2x+2s-1) = 2^n\left(\left(x+\frac{1}{2}\right)\left(x+\frac{3}{2}\right)\cdots\left(x+\frac{2n-1}{2}\right)\right) \tag{36}$$

$$= 2^n u(u-1)\cdots(u-n+1) \tag{37}$$

where

$$u = x + (2n-1)/2. \tag{38}$$

It follows easily by induction from (33) that

$$\Delta^r 2^n u(u-1)\cdots(u-n+1) = 2^n n(n-1)\cdots \cdot(n-r+1)u(u-1)\cdots(u-n+r+1) \tag{39}$$

when $r < n$, and that

$$\Delta^n 2^n u(u-1)\cdots(u-n+1) = 2^n n! \tag{40}$$

Hence, (31), (37), (38), and (39) yield

$$b(n,r) = 2^r n(n-1)\cdots(n-r+1)\prod_{j=r}^{n-1}(2j+1) \tag{41}$$

if $n \geqslant 1$ and $0 \leqslant r < n$, and (31), (37), (38), and (40) yield

$$b(n,n) = 2^n n! \tag{42}$$

Dividing (41) by $r!$ and (42) by $n!$ yields (25) and (26), as desired.

In particular, with $c(0,0) = 1$,

$$c(n,0) = \prod_{j=0}^{n-1}(2j+1) = (2n-1)c(n-1,0) \tag{43}$$

for $n \geqslant 1$, and

$$c(n,r) = \frac{2(n-r+1)}{r(2r-1)}c(n,r-1) \tag{44}$$

for $n \geqslant 1$ and $1 \leqslant r \leqslant n$. The recurrence relations (43) and (44) enable one to generate the triangular array of numbers $c(n,r)$ with ease. Hence one may easily calculate the polynomials $\alpha_n(w_i)$, which are listed below for $0 \leqslant n \leqslant 5$:

$$\begin{aligned}
\alpha_0(w_i) &= 1\\
\alpha_1(w_i) &= 1 + 2\lambda_i\\
\alpha_2(w_i) &= 3 + 12\lambda_i + 4\lambda_i^2\\
\alpha_3(w_i) &= 15 + 90\lambda_i + 60\lambda_i^2 + 8\lambda_i^3\\
\alpha_4(w_i) &= 105 + 840\lambda_i + 840\lambda_i^2 + 224\lambda_i^3 + 16\lambda_i^4\\
\alpha_5(w_i) &= 945 + 9450\lambda_i + 12600\lambda_i^2 + 5040\lambda_i^3 + 720\lambda_i^4 + 32\lambda_i^5.
\end{aligned} \tag{45}$$

B. Moments about Zero of R

From (15) and (16),

$$R(\boldsymbol{x}, \boldsymbol{m}_y) = \sum_{i=1}^{d} w_i \gamma^{-1}(i). \tag{46}$$

The nth moment about zero of R is then

$$\alpha_n[R(\boldsymbol{x}, \boldsymbol{m}_y)] = E\{R^n(\boldsymbol{x}, \boldsymbol{m}_y)\} = E\left\{\left[\sum_{i=1}^{d} w_i/\gamma(i)\right]^n\right\}. \tag{47}$$

The coefficients of a sum raised to the nth power are given by the multinomial theorem [13]; that is,

$$\alpha_n[R(\boldsymbol{x}, \boldsymbol{m}_y)] = E\left\{\sum \frac{n!}{e_1!e_2! \cdots e_d!} \prod_{i=1}^{d} [w_i/\gamma(i)]^{e_i}\right\}, \tag{48}$$

where the summation is taken over all nonnegative values of $e_1, e_2, \cdots, e_d$ such that $e_1 + e_2 + \cdots + e_d = n$.

In view of the independence of the w_i's, it follows that

$$\alpha_n[R(\boldsymbol{x}, \boldsymbol{m}_y)] = \sum\left(\frac{n!}{e_1!e_2! \cdots e_d!}\right) \prod_{i=1}^{d} \left[\alpha_{e_i}(w_i)/(\gamma(i))^{e_i}\right] \tag{49}$$

where the $\alpha_{e_i}(w_i)$, $i = 1, 2, \cdots, d$, are given by (24), using values of λ_i given by (19).

Since

$$R(\boldsymbol{y}, \boldsymbol{m}_x) = \sum_{i=1}^{d} (s_i - m_{ri})^2 = \sum_{i=1}^{d} w_i \gamma(i), \tag{50}$$

where w_i is given by (21), it follows using a similar development that

$$\alpha_n[R(\boldsymbol{y}, \boldsymbol{m}_x)] = \sum\left(\frac{n!}{e_1!e_2! \cdots e_d!}\right) \prod_{i=1}^{d} \left[\alpha_{e_i}(w_i)(\gamma(i))^{e_i}\right], \tag{51}$$

where the $\alpha_{e_i}(w_i)$, $i = 1, 2, \cdots, d$, are given by (24) using values of λ_i as given in (22).

IV. Special Cases

In this section we consider special cases involving various arrangements of mean vectors and covariance matrices of two pattern populations.

Equal Covariance Matrices

When $\boldsymbol{C}_x = \boldsymbol{C}_y = \boldsymbol{C}$, it follows from (13) and (14) that $\boldsymbol{C}_r = \boldsymbol{C}_s = \boldsymbol{I}$. Consequently, $\gamma(i) = 1$, $i = 1, 2, \cdots, d$, and both forms of λ_i ((19) and (2.22)) become the same. This leads to equal moments via (49) and (51).

Equal Mean Vectors

When $\boldsymbol{m}_x = \boldsymbol{m}_y$, it follows from (7) and (8) that $\boldsymbol{m}_r = \boldsymbol{m}_s$ and, consequently, $\lambda_i = 0$ in (19) and (22). Then from (24) and (25)

$$\alpha_n(w_i) = c(n, 0) = \prod_{s=1}^{n} (2s - 1) \tag{52}$$

for both populations. Substitution of (52) into (49) and (51) yields

$$\alpha_n[R(\boldsymbol{x}, \boldsymbol{m}_y)] = \sum\left(\frac{n!}{e_1!e_2! \cdots e_d!}\right) \prod_{i=1}^{d} \left[c(e_i, 0)/(\gamma(i))^{e_i}\right] \tag{53}$$

and

$$\alpha_n[R(\boldsymbol{y}, \boldsymbol{m}_x)] = \sum\left(\frac{n!}{e_1!e_2! \cdots e_d!}\right) \prod_{i=1}^{d} \left[(c(e_i, 0)\gamma(i)^{e_i}\right] \tag{54}$$

Equal Mean Vectors and Covariance Matrices (Intraclass Mahalanobis Distance)

When $\boldsymbol{m}_x = \boldsymbol{m}_y = \boldsymbol{m}$ and $\boldsymbol{C}_x = \boldsymbol{C}_y = \boldsymbol{C}$, we have only one population and the problem reduces to computing the moments of the intraclass Mahalanobis distance. It follows from (52) and (53) and the fact that each $\gamma(i) = 1$ (see the remarks above on equal covariance matrices) that

$$\alpha_n[R(\boldsymbol{x}, \boldsymbol{m}_x)] = \sum \frac{n!}{e_1!, e_2!, \cdots, e_d!} \cdot \prod_{i=1}^{e_1} (2i - 1) \cdots \prod_{i=1}^{e_d} (2i - 1) = \prod_{j=0}^{n-1} (d + 2j), \tag{55}$$

where the summation in (54) is over all $e_i \geqslant 0$ such that $e_1 + e_2 + \cdots + e_d = n$, and (55) follows from (54) by the following argument. By the extended binomial theorem

$$(1 - 2x)^{-1/2} = \sum_{e=0}^{\infty} \binom{-1/2}{e} (-2x)^e = \sum_{e=0}^{\infty} \left(\prod_{e=0}^{e} (2i - 1)\right) \frac{x^e}{e!}. \tag{56}$$

Raising (56) to the dth power yields

$$(1 - 2x)^{-d/2} = \sum_{n=0}^{\infty} \left(\sum\left(\prod_{i=1}^{e_1} (2i - 1) \cdots \prod_{i=1}^{e_d} (2i - 1)\right) \Big/ e_1! \cdots e_d!\right) x^n. \tag{57}$$

On the other hand, the extended binomial theorem yields

$$(1 - 2x)^{-d/2} = \sum_{n=0}^{\infty} \binom{-d/2}{n} (-2x)^n = \sum_{n=0}^{\infty} \frac{(d)(d+2) \cdots (d + 2n - 2)}{n!} x^n. \tag{58}$$

Comparison of the coefficients of x^n in (57) and (58) yields the desired result.

One observes from (55) that the moment in question depends only on the order of the moment and the dimension of the pattern vectors.

V. Conclusion

The expressions given in (24), (25), (26), (49), and (51) lead to a straightforward algorithm for computing the moments about zero of the interclass Mahalanobis distance. If desired, the central moments can be obtained from these results by means of a well-known transformation [10].

The importance of these results is that they allow direct computation of the moments without having to resort to the intermediate step of obtaining the pdf which, as indicated in Section I, is not a trivial problem in the case of unequal covariance matrices.

The expressions for the moments were considerably simplified in the special cases discussed in Section IV. In particular, the intraclass Mahalanobis distance was shown to lead to expressions which depend only on the order of the moments and the dimension of the vector populations.

Acknowledgment

The authors express their appreciation to Professor Balram Rajput, Department of Mathematics, University of Tennessee for many helpful discussions.

References

[1] R. C. Gonzalez and L. C. Howington, "Machine recognition of abnormal behavior in nuclear reactors," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-7, no. 10, pp. 717–728.

[2] S. Y. Hus, "The Mahalanobis classifier with the generalized inverse approach for automated analysis of imagery texture data," *Comput. Graph. Image Proc.*, vol. 9, no. 2, pp. 117–134, 1979.

[3] R. E. Pogue, "Some investigations of multivariate discrimination procedures with applications to diagnosis clinical electrocardiography," Ph.D. dissertation, Univ. Minnesota, Minneapolis, 1966.

[4] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*. Reading, MA: Addison-Wesley, 1974.

[5] G. T. Toussaint, "Bibliography on estimation of misclassification," *IEEE Trans. Inform. Theory*, vol. IT-20, no. 4, pp. 472–479, 1974.

[6] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1962.

[7] K. Fukunaga and T. F. Krile, "Calculation of Bayes' recognition error for two multivariate gaussian distributions," *IEEE Trans. Comput.*, vol. C-18, no. 3, pp. 220–229, 1969.

[8] R. Bellman, *Introduction to Matrix Analysis*. New York: McGraw-Hill, 1970.

[9] F. A. Graybill, *An Introduction to Linear Statistical Models*, Vol. 1. New York: McGraw-Hill, 1961.

[10] H. Cramer, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton University, 1946.

[11] C. Jordan, *Calculus of Finite Differences*. New York: Chelsea Publishing, 1947.

[12] C. Berge, *Principles of Combinatorics*. New York: Academic, 1971.

[13] G. Berman and K. D. Fryer, *Introduction to Combinatorics*. New York: Academic, 1972.

[14] B. V. Gnedenko, *The Theory of Probability*. New York: Chelsea Publishing, 1967.

[15] D. M. Young and R. T. Gregory, *Survey of Numerical Mathematics*, Vol. I. Reading, MA: Addison-Wesley, 1972.

[16] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic, 1972.