JOURNAL OF MULTIVARIATE ANALYSIS 12, 450–455 (1982)

The Fréchet Distance between Multivariate Normal Distributions

D. C. DOWSON AND B. V. LANDAU

University of Salford, England

Communicated by P. R. Krishnaiah

The Fréchet distance between two multivariate normal distributions having means μ_X, μ_Y and covariance matrices Σ_X, Σ_Y is shown to be given by $d^2 = |\mu_X - \mu_Y|^2 + \operatorname{tr}(\Sigma_X + \Sigma_Y - 2(\Sigma_X \Sigma_Y)^{1/2})$. The quantity d_0 given by $d_0^2 = \operatorname{tr}(\Sigma_X + \Sigma_Y - 2(\Sigma_X \Sigma_Y)^{1/2})$ is a natural metric on the space of real covariance matrices of given order.

1. INTRODUCTION

In [1], M. Fréchet introduced a metric on the space of probability distributions on R having first and second moments. The Fréchet distance $d(F, G)$ between two distributions F and G is defined by

$$d^2(F, G) = \min_{X,Y} E\,|X - Y|^2 \tag{1}$$

where the minimization is taken over all random variables X and Y having distributions F and G, respectively. The bivariate distribution H which minimizes the right-hand side of (1) is well-known [1, 2] to be the singular distribution with distribution function

$$H(x, y) = \min[F(x), G(y)], \tag{2}$$

where F and G are the distribution functions of F and G, respectively. In the particular case when F and G belong to a family of distributions which is closed with respect to changes of location and scale, the Fréchet distance takes the simple form

$$d^2(F, G) = (\mu_X - \mu_Y)^2 + (\sigma_X - \sigma_Y)^2 \tag{3}$$

Received July 11, 1979.

AMS 1980 subject classification: Primary 62E10, 62H05.

Key words and phrases: Fréchet distance, multivariate normal distributions, covariance matrices.

0047-259X/82/030450–06$02.00/0

where μ_X, μ_Y and σ_X, σ_Y are the respective means and standard deviations of F and G.

Definition [1] generalises in an obvious way to define a metric on the space of probability distributions on R^n having second moments. The solution (2) does not apply in the case when X and Y are vectors and the evaluation of the Fréchet distance is extremely difficult in general. Again, however, the distance $d(F, G)$ is easy to determine when F and G belong to a family of n-dimensional distributions which is closed with respect to linear transformations of the random vector. We shall prove that in this case, the Fréchet distance is given by

$$d^2(F, G) = |\mu_X - \mu_Y|^2 + \operatorname{tr}[\Sigma_X + \Sigma_Y - 2(\Sigma_X \Sigma_Y)^{1/2}] \tag{4}$$

where μ_X, μ_Y and Σ_X, Σ_Y are the respective means and covariance matrices of F and G, and the positive square root is taken. The above formula holds in particular when F and G are normal distributions on R^n. Additionally it will be seen that

$$d_0^2(\Sigma_X, \Sigma_Y) = \operatorname{tr}[\Sigma_X + \Sigma_Y - 2(\Sigma_X \Sigma_Y)^{1/2}] \tag{5}$$

defines a metric on the space of all covariance matrices of order n.

2. Main Result

Since the results (4) and (5) generalise easily to the complex case, it will be convenient to take X and Y as complex-valued random vectors in C^n having, initially, zero means. We prove the following theorem.

THEOREM. *Let X, Y be random vectors taking values in C^n and having zero means and covariance matrices Σ_X, Σ_Y respectively. Then*

$$\operatorname{tr}[\Sigma_X + \Sigma_Y - 2(\Sigma_X \Sigma_Y)^{1/2}] \leqslant E\,|X - Y|^2 \leqslant \operatorname{tr}[\Sigma_X + \Sigma_Y + 2(\Sigma_X \Sigma_Y)^{1/2}] \tag{6}$$

where the square roots are the positive roots.

The bounds in (6) *are attained when $X - Y$ has covariance matrix*

$$\Sigma_{X-Y} = \Sigma_X + \Sigma_Y \pm [(\Sigma_X \Sigma_Y)^{1/2} + (\Sigma_Y \Sigma_X)^{1/2}], \tag{7}$$

and, for non-singular Σ_X, this occurs when X and Y are related by

$$Y = \pm \Sigma_X^{-1} (\Sigma_X \Sigma_Y)^{1/2} X. \tag{8}$$

Proof. Let W be the random vector taking values in C^{2n} defined by $W = [{}^X_Y]$, and denote the covariance matrix of W by

$$\Sigma_W = \begin{bmatrix} \Sigma_X & V \\ V^* & \Sigma_Y \end{bmatrix}$$

where $*$ denotes conjugate transpose. Clearly,

$$E\,|X - Y|^2 = \mathrm{tr}[\Sigma_X + \Sigma_Y - V - V^*]$$

so that we require extreme values of $\mathrm{tr}(V + V^*)$ subject to the condition that Σ_W is a covariance matrix. We note that $\mathrm{tr}(V + V^*)$ is a linear functional defined on a convex region of the Σ_W space and hence we can apply the method of Lagrange multipliers. Let Q denote the Hermitian form $w^* \Sigma_W w$ so that

$$Q = x^*\Sigma_X x + x^* V y + y^* V^* x + y^* \Sigma_Y y.$$

Since Q is non-negative definite, it must be capable of being written

$$Q = \sum_{i=1}^{m} |a_i^* x + b_i^* y|^2,$$

where $a_i, b_i \in C^n$ for $i = 1,\ldots, m$ and $m \leqslant 2n$.

It follows that Σ_X, Σ_Y and V must have simultaneous representations of the form

$$\Sigma_X = \sum_{i=1}^{m} a_i a_i^*, \qquad \Sigma_Y = \sum_{i=1}^{m} b_i b_i^*, \qquad V = \sum_{i=1}^{m} a_i b_i^*. \tag{9}$$

We seek to minimize $\mathrm{tr} \sum_{i=1}^{m} (a_i b_i^* + b_i a_i^*)$ subject to the matrices $\sum_{i=1}^{m} a_i a_i^*$ and $\sum_{i=1}^{m} b_i b_i^*$ having given values Σ_X and Σ_Y, respectively. Introducing Hermitian matrices Λ and M of Lagrange multipliers we seek unconstrained extreme values of the real quantity

$$S = \mathrm{tr} \sum_{i=1}^{m} (a_i b_i^* + b_i a_i^*) + \mathrm{tr}\left(\sum_{i=1}^{m} a_i a_i^*\right)\Lambda + \mathrm{tr}\left(\sum_{i=1}^{m} b_i b_i^*\right) M.$$

It is easily seen that extreme values satisfy the conditions

$$\begin{aligned} b_i &= \Lambda a_i \\ a_i &= M b_i \end{aligned} \qquad (i = 1,\ldots, m).$$

Thus, Λ is a Hermitian matrix which by virtue of conditions (9) must satisfy the equations

$$\Lambda \Sigma_X \Lambda = \Sigma_Y \tag{10}$$

and

$$V = \Sigma_X \Lambda, \tag{11}$$

which together imply $V^2 = \Sigma_X \Sigma_Y$.

For non-singular Σ_X we may put

$$\Lambda = \Sigma_X^{-1/2} R \Sigma_X^{-1/2} \tag{12}$$

so that R is Hermitian and, because of (10), R^2 simplifies to the non-negative definite $\Sigma_X^{1/2} \Sigma_Y \Sigma_X^{1/2}$. If the latter has eigenvalues λ_i and eigenvectors s_i with $s_i^* s_j = \delta_{ij}$, then all its Hermitian square roots are of the form $\sum_{i=1}^m \varepsilon_i \lambda_i^{1/2} s_i s_i^*$ where, for each i, $\varepsilon_i^2 = 1$. From (11) and (12), $V = \Sigma_X^{1/2} R \Sigma_X^{-1/2}$ has the same eigenvalues as R, so the maximum and minimum of tr V, viz., $\pm \sum_{i=1}^m \lambda_i^{1/2}$, are attained when

$$V = \pm\, \Sigma_X^{1/2} (\Sigma_X^{1/2} \Sigma_Y \Sigma_X^{1/2})^{1/2}\, \Sigma_X^{-1/2}.$$

The latter product is a way of expressing $(\Sigma_X \Sigma_Y)^{1/2}$, the positive square root of $\Sigma_X \Sigma_Y$, when Σ_X is non-singular. (More generally $\Sigma_X \Sigma_Y$ has the same non-negative eigenvalues λ_i as $\Sigma_X^{1/2} \Sigma_Y \Sigma_X^{1/2}$ above, and its positive square root is $\sum_{i=1}^m \lambda_i^{1/2} u_i t_i^*$ where t_i, u_i are left and right eigenvectors of $\Sigma_X \Sigma_Y$ with $t_i^* u_j = \delta_{ij}$.) Thus, $\mathrm{tr}(V + V^*)$ is maximised or minimised by taking

$$V = \pm (\Sigma_X \Sigma_Y)^{1/2} \tag{13}$$

That the covariance matrix of $X - Y$ is given by (7) and that relation (8) between X and Y yields the upper and lower bounds is now routine. Moreover, by continuity considerations (13) carries over to the case where Σ_X is singular; consequently (6) and (7) hold generally and the theorem is proved.

For singular Σ_X, (8) obviously no longer applies; but if the null space of Σ_X is contained in that of Σ_Y, it holds with Σ_X^{-1} replaced by the pseudo-inverse of Σ_X.

COROLLARY. *Let $\mathscr{C}_n(\Sigma)$ be the set of all covariance (non-negative, Hermitian) matrices of order n. Then*

$$d_0(\Sigma_X, \Sigma_Y) = [\mathrm{tr}(\Sigma_X + \Sigma_Y - 2(\Sigma_X \Sigma_Y)^{1/2})]^{1/2} \tag{14}$$

defines a metric on $\mathscr{C}_n(\Sigma)$.

In the particular case when the real covariance matrices Σ_X and Σ_Y have the same principal axes, the metric has the particularly simple form

$$d_0^2(\Sigma_X, \Sigma_Y) = \sum_{i=1}^n (\sigma_i - \rho_i)^2 \tag{15}$$

where σ_i, ρ_i are the standard deviations of the (principal) components of X and Y, respectively, along the ith principal axis.

3. Fréchet Distance between Multinormal Distributions

Since multinormal distributions are determined completely by their means and covariance matrices, it follows immediately that the Fréchet distance between normal distributions F and G on R^n with means μ_X, μ_Y and (real) covariance matrices Σ_X, Σ_Y is given by

$$d^2(F, G) = |\mu_X - \mu_Y|^2 + \operatorname{tr}(\Sigma_X + \Sigma_Y - 2(\Sigma_X \Sigma_Y)^{1/2}). \tag{16}$$

Indeed, the above result holds for any two distributions from a family of real, elliptically contoured distributions having finite second-order moments and with density function of the form

$$p(x; \mu, A) = (\text{const}) \times f((x-\mu)^* A(x-\mu)), \tag{17}$$

where f is a non-negative function on the positive real axis such that $0 < \int_0^\infty r^{n/2-1} f(r)\, dr < \infty$, and A is a positive-definite, symmetric matrix. In particular, by taking

$$\begin{aligned} f(u) &= \text{const}, \qquad 0 \leqslant u \leqslant c \\ &= 0 \qquad\qquad \text{elsewhere}, \end{aligned} \tag{18}$$

we see that Eq. (16) also determines the Fréchet distance between uniform distributions on two ellipsoids centered at μ_X, μ_Y and having covariance matrices Σ_X, Σ_Y. (This implies that the ellipsoids have the same shape and orientation as the principal ellipsoids of Σ_X, Σ_Y.) It is worth emphasizing that for non-singular Σ_X the minimizing transformation (8) between X and Y is (a) linear and (b) Hermitian. This latter property has the interpretation that the transformation is a pure strain and does not involve a rotation. This non-rotational property holds in a certain sense even if F and G have arbitrary continuous density functions in R^n provided that the minimum of $E|X-Y|^2$ is achieved by a differentiable transformation between X and Y. By consideration of nearly uniform distributions over small ellipsoids, one may show the non-rotational property of the minimizing transformation to hold locally. This result can also be proved by a standard calculus of variations argument in which $E|X-Y|^2$ is minimized subject to the constraint that the Jacobian $\partial(y)/\partial(x)$ satisfies the equation

$$\frac{\partial(y)}{\partial(x)} = \frac{p_X(x)}{p_Y(y)} \tag{19}$$

where $p_X(x)$, $p_Y(y)$ are the given density functions of the vectors X and Y, respectively.

4. Two Trace Inequalities for Real Covariance Matrices

The result

$$\min_{X,Y} E\,|X-Y|^2 = \operatorname{tr}[\Sigma_X + \Sigma_Y - 2(\Sigma_X \Sigma_Y)^{1/2}] \tag{20}$$

when X, Y have zero means and covariance matrices Σ_X, Σ_Y has two simple consequences, First,

$$\operatorname{tr}(\Sigma_X \Sigma_Y)^{1/2} \leqslant \operatorname{tr}\left(\frac{\Sigma_X + \Sigma_Y}{2}\right) \tag{21}$$

with equality iff $\Sigma_X = \Sigma_Y$. This well-known result is a generalisation of the fact that the geometric mean of two positive numbers is less than or equal to the arithmetic mean. Second, by replacing Y in (20) by tY where t is any complex number we see that

$$\min_{X,Y} E\,|X - tY|^2 = (\operatorname{tr}\Sigma_Y)\,|t|^2 - 2\operatorname{tr}(\Sigma_X \Sigma_Y)^{1/2} \cdot |t| + \operatorname{tr}\Sigma_X \geqslant 0$$

which implies

$$\operatorname{tr}(\Sigma_X \Sigma_Y)^{1/2} \leqslant [\operatorname{tr}\Sigma_X \cdot \operatorname{tr}\Sigma_Y]^{1/2} \tag{22}$$

where equality holds iff $\Sigma_X = |t|^2\,\Sigma_Y$ for some complex scalar t. Inequality (22) is a generalisation of the Schwarz inequality. The quantity

$$\rho = \operatorname{tr}(\Sigma_X \Sigma_Y)^{1/2}\,[\operatorname{tr}\Sigma_X \cdot \operatorname{tr}\Sigma_Y]^{-1/2} \tag{23}$$

is the largest correlation coefficient possible between two random vectors X, Y having prescribed covariance matrices Σ_X, Σ_Y.

References

[1] Fréchet, M. (1957). Sur la distance de deux lois de probabilité. *C. R. Acad. Sci. Paris*, **244** 689–692.

[2] Cambanis, S., Simons, G., and Stout, W. (1976). Inequalities for $Ek(X, Y)$ when the marginals are fixed. *Z. Wahrsch. Verw. Gebiete*, **36** 285–294.