The Annals of Mathematical Statistics
1968, Vol. 39, No. 6, 1844–1848

ON DOMINATING AN AVERAGE ASSOCIATED WITH DEPENDENT GAUSSIAN VECTORS[1]

BY S. C. SCHWARTZ AND W. L. ROOT

Princeton University and University of Michigan

0. Abstract. Mehler's identity is used to obtain a bound for the integral of the absolute difference between the bivariate gaussian density function and its corresponding marginal densities. In a sense, this integral measures the contribution of the dependency between the gaussian random variables to an expectation. It is shown that the integral is dominated by $|\rho|/(1 - |\rho|)$, where ρ is the correlation coefficient between the random variables. Using the Hotelling canonical decomposition of a variance-covariance matrix, the result is extended to the case of dependent gaussian vectors with the bound now given in terms of the canonical correlations, i.e., the roots of a characteristic equation related to the variance-covariance matrix of the vectors. As an application of the results, a bound is obtained for the variance of the function $F(X_1, \cdots, X_n) = n^{-1} \sum_{i=1}^{n} f(X_i)$. The $\{X_i\}$ denote a sequence of dependent, non-stationary, gaussian random variables (or vectors), and $f(x)$ is any bounded measurable function. For the stationary case, the rate of convergence of the variance is easily expressed in terms of the summability properties of the correlation coefficients. The paper concludes with some comments on extending the results to the class of φ^2-bounded bivariate density functions.

1. Preliminaries. The following formulas concerning the Hermite polynomials can be found in Cramér [3], page 133. The polynomials are defined by

$$H_j(x) = (-1)^j e^{x^2/2} (d/dx)^j e^{-x^2/2} \tag{1}$$

and satisfy the orthogonality relation

$$(2\pi)^{-\frac{1}{2}} \int_{-\infty}^{+\infty} e^{-x^2/2} H_n(x) H_m(x)\, dx = n!, \qquad m = n, \tag{2}$$
$$= 0, \qquad m \neq n.$$

A useful generating function is

$$(1 - \rho^2)^{-\frac{1}{2}} \exp - \{(\rho^2 x^2 + \rho^2 y^2 - 2\rho xy)/2(1 - \rho^2)\} = \sum_0^\infty \rho^j (j!)^{-1} H_j(x) H_j(y), \tag{3}$$

which holds under the condition that $|\rho| < 1$. Equation (3), in a slightly different form, is sometimes called Mehler's formula (see [5], page 194).

Let $g(x; \sigma)$ designate the gaussian density function with standard deviation

Received 12 June 1967; revised 6 March 1968.
[1] This work was supported in part by the National Aeronautics and Space Administration under Research Grant NsG-2-59 and by the National Science Foundation under Grant NSF-GK-1532.

σ, and let $g_2(x, y; \sigma_1, \sigma_2, \rho)$ be the bivariate gaussian density. The gaussian random variables X (with $E\{X\} = 0$, $E\{X^2\} = \sigma_1^2$) and Y (with $E\{Y\} = 0$, $E\{Y^2\} = \sigma_2^2$) are correlated with a correlation coefficient equal to ρ. (There is no difficulty in extending the results to non-zero mean variables.) An expansion for the bivariate density is obtained by multiplying (3) by $(2\pi\sigma_1 \sigma_2)^{-1} \exp\{-(x^2 + y^2)/2\}$ and then substituting $x = x/\sigma_1$ and $y = y/\sigma_2$. The result is,

$$(4) \quad g_2(x, y; \sigma_1, \sigma_2, \rho) = g(x; \sigma_1)g(y; \sigma_2) \sum_{j=0}^{\infty} \rho^j (j!)^{-1} H_j(x/\sigma_1) H_j(y/\sigma_2).$$

Note that $H_0(x) = 1$ and that the first term of the series is just the product of the univariate gaussian densities.

2. Dominating the integral. It is relatively simple to dominate the integral (all integrals are taken from $-\infty$ to $+\infty$)

$$(5) \qquad I = \int\int |g_2(x, y; \sigma_1, \sigma_2, \rho) - g(x; \sigma_1)g(y; \sigma_2)| \, dx \, dy$$

which provides a measure of the contribution of the dependency between the random variables x and y to an expectation. Substituting the expansion for the bivariate density and cancelling the first term of the series gives

$$(6) \qquad I = \int\int |g(x;\sigma_1)g(y;\sigma_2) \sum_{j=1}^{\infty} \rho^j (j!)^{-1} H_j(x/\sigma_1) H_j(y/\sigma_2)| \, dx \, dy.$$

The magnitude sign is brought inside the summation and the integration and summation operations are interchanged (by the Lebesgue monotone convergence theorem) to obtain

$$(7) \quad I \leqq \sum_{j=1}^{\infty} |\rho|^j (j!)^{-1} \int\int g(x; \sigma_1)g(y; \sigma_2)|H_j(x/\sigma_1)||H_j(y/\sigma_2)| \, dx \, dy.$$

Using the Schwarz inequality and the orthogonality relation (equation (2)) in the form

$$\int g(x; \sigma_1)|H_j(x/\sigma_1)| \, dx \leqq \{\int g(x; \sigma_1) \, dx \int g(x; \sigma_1) H_j^2(x/\sigma_1) \, dx\}^{\frac{1}{2}} \leqq (j!)^{\frac{1}{2}},$$

the double integral is majorized by

$$(8) \qquad I \leqq \sum_{j=1}^{\infty} |\rho|^j = |\rho|/(1 - |\rho|).$$

Again, the bound is valid for $|\rho| < 1$.

3. The multivariate case. Let $\mathbf{x}$ and $\mathbf{y}$ be k-dimensional zero-mean gaussian vectors with covariance matrices $E\{\mathbf{x}\mathbf{x}'\} = A_1$, $E\{\mathbf{y}\mathbf{y}'\} = A_2$ and a cross-covariance matrix $E\{\mathbf{x}\mathbf{y}'\} = B$. (Vector quantities are denoted by $\mathbf{x}$ and the transpose by a prime.) Define the covariance matrices

$$(9) \qquad M = E\left[\begin{Bmatrix}\mathbf{x}\\ \mathbf{y}\end{Bmatrix} [\mathbf{x}' \ \mathbf{y}']\right] = \begin{bmatrix} A_1 & B \\ B' & A_2 \end{bmatrix},$$

$$N = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}.$$

We now consider the multivariate version of (5)

$$I_{2k} = \int \cdots \int^{2k} |g_{2k}(\mathbf{x}, \mathbf{y}; M) - g_{2k}(\mathbf{x}, \mathbf{y}; N)| \, d\mathbf{x} \, d\mathbf{y}. \tag{10}$$

Equation (10) is transformed into k double integrals of the form (5) through the use of the Hotelling canonical decomposition of the variance-covariance matrix ([6], Section 3: [1], Section 12.2). In analogy to the previously imposed condition ($|\rho| < 1$), the covariance matrix M is assumed to be strictly positive definite.

The canonical decomposition of Hotelling (see [1], [6] for the details) can be viewed as a simultaneous transformation of the variance and covariance matrices (9) into congruent matrices of the form

$$M_1 = TMT' = \begin{bmatrix} I & R \\ R & I \end{bmatrix}, \qquad N_1 = TNT' = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}. \tag{11}$$

I is the identity matrix and R is a diagonal matrix (the matrix of canonical correlation) with elements $r_1 \cdots r_k$. The r_i are the positive values of $(r_i^2)^{\frac{1}{2}}$, $i = 1, \cdots, k$, which are the solutions of the kth order characteristic equation

$$|r^2 A_1 - B A_2^{-1} B'| = 0. \tag{12}$$

In addition, it can be shown that $|r_i| < 1$, $i = 1, \cdots, k$. In the context of our problem we have

THEOREM. (Hotelling) *With M strictly positive definite, the change of variables*

$$\begin{Bmatrix} \mathbf{u} \\ \mathbf{v} \end{Bmatrix} = T \begin{Bmatrix} \mathbf{x} \\ \mathbf{y} \end{Bmatrix} \tag{13}$$

takes the integral (10) into

$$\begin{aligned} I_{2k} &= \int \cdots \int^{2k} |g_{2k}(\mathbf{u}, \mathbf{v}; M_1) - g_{2k}(\mathbf{u}, \mathbf{v}; N_1)| \, d\mathbf{u} \, d\mathbf{v} \\ &= \int \cdots \int^{2k} |\textstyle\prod_{i=1}^{k} g_2(u_i, v_i; 1, 1, r_i) \\ &\quad - \textstyle\prod_{i=1}^{k} g(u_i; 1) g(v_i; 1)| \, du_1 \cdots du_k \, dv_1 \cdots dv_k. \end{aligned} \tag{14}$$

Since $|r_i| < 1, i = 1, \cdots, k$, the bivariate gaussian density functions appearing in (14) can be expanded using (4). Then, if one proceeds in a manner completely analogous to the previous development (Equation (6)–(8)), one can dominate the integral by

$$\int \cdots \int |g_{2k}(\mathbf{x}, \mathbf{y}; M) - g_{2k}(\mathbf{x}, \mathbf{y}; N)| \, d\mathbf{x} \, d\mathbf{y} \leqq -1 + \textstyle\prod_{i=1}^{k} (1 - r_i)^{-1}. \tag{15}$$

This is the desired result.

4. An application. Let $\{X_i\}, i = 1, 2, \cdots$, represent a sequence of gaussian random variables with $E\{X_i\} = 0$, $E\{X_i^2\} = \sigma_i^2$, $E\{X_i X_j\} = \rho_{ij}\sigma_i\sigma_j$. As an application of our results, we consider dominating the variance of the average of any bounded function of the observations.

$$F = F(X_1, \cdots, X_n) = n^{-1} \textstyle\sum_{i=1}^{n} f(X_i), \tag{16}$$

where $f(x)$ is any bounded measurable function $|f(x)| < c$. The second moment is

$$E\{F^2\} = n^{-2}\{\textstyle\sum_{i=1}^{n} E\{f^2(X_i)\} + 2\sum_{i=1}^{n} \sum_{j=i+1}^{n} E\{f(X_i)f(X_j)\}\}. \tag{17}$$

In order to use the results of the previous section, add

$$n^{-2}\left(\sum_{i=1}^{n} E\{f(X_i)\}\right)^2 - n^{-2}\sum_{i=1}^{n} (E\{f(X_i)\})^2$$

to (17) and subtract its equivalent

$$2n^{-2}\sum_{i=1}^{n}\sum_{j=i+1}^{n} E\{f(X_i)\}E\{f(X_j)\}.$$

Form the variance and regroup terms to obtain:

$$(18)\quad E\{(F - E\{F\})^2\} = n^{-2}\{\textstyle\sum_{i=1}^{n} [E\{f^2(X_i)\} - E\{f(X_i)\}^2]\} + 2n^{-2}\{\textstyle\sum_{i=1}^{n}\sum_{j=i+1}^{n} [E\{f(X_i)f(X_j)\} - E\{f(X_i)\}E\{f(X_j)\}]\}.$$

The second term of (18) is the part due to the dependency between the random variables. Using (8) and the fact that $f(x)$ is bounded, the second term of (18) is dominated by:

$$(19)\quad \begin{aligned} &E\{f(X_i)f(X_j)\} - E\{f(X_i)\}E\{f(X_j)\} \\ &\quad = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x_i)f(x_j)(g_2(x_i, x_j; \sigma_i, \sigma_j, \rho_{ij}) \\ &\qquad - g(x_i; \sigma_i)g(x_j; \sigma_j))\, dx_i\, dx_j \leqq c^2 |\rho_{ij}|(1 - |\rho_{ij}|)^{-1}. \end{aligned}$$

Consequently, the variance is bounded by

$$(20)\quad E\{(F - E\{F\})^2\} \leqq n^{-2}\textstyle\sum_{i=1}^{n} E\{(f(X_i) - E\{f(X_i)\})^2\} + 2c^2n^{-2}\textstyle\sum_{i=1}^{n}\sum_{j=i+1}^{n} |\rho_{ij}|(1 - |\rho_{ij}|)^{-1}.$$

To simplify the discussion, assume the sequence $\{X_i\}$ is stationary. Then, $\rho_{ij} = \rho_{i-j}$ and letting $\tau = j - i$, (20) reduces to

$$(21)\quad E\{(F - E\{F\})^2\} \leqq n^{-1}E\{(f(X) - E\{(f(X)\})^2\} + 2c^2n^{-2}\textstyle\sum_{\tau=1}^{n} (n - \tau)|\rho_\tau|(1 - |\rho_\tau|)^{-1}.$$

Recall that by assumption, $\rho_* = \max_\tau |\rho_\tau| < 1$. If the correlation coefficients satisfy $\sum_1^n |\rho_\tau| = O(n^\delta)$, the second term is bounded by $2c^2(1 - \rho_*)^{-1}O(n^{\delta-1})$. Hence, if the correlation coefficients are absolutely summable, $\delta = 0$ and the second term of (21) is of order $O(1/n)$—the same order as if the sequence of random variables $\{X_i\}$ were independent. Similar results can be obtained for the vector case. (See [10], Chapter 3.)

5. Concluding comments. In conclusion, we note that the technique used in deriving the bound for the bivariate case (equation (8)) is applicable to a class of bivariate densities, of which the gaussian is the most prominent member. Specifically, the technique is applicable to any bivariate density function which can be expanded in the form

$$(22)\quad p_2(x, y) = p_a(x)p_b(y)\sum_{i,j} a_{ij}\theta_i^{(a)}(x)\theta_j^{(b)}(y)$$

with

$$a_{ij} = \int\int p_2(x, y)\theta_i^{(a)}(x)\theta_j^{(b)}(y)\, dx\, dy$$

and where p_a and p_b are the corresponding marginal density functions. (Equation (22) has been called the Barrett-Lampard expansion [2], and has found other uses in noise theory.) The functions $\theta_i^{(a)}(x)$ and $\theta_i^{(b)}(y)$ are sets of (complete) orthogonal functions with respect to the weights $p_a(x)$ and $p_b(y)$. Observe that, in general, $\theta_0^{(a)}(x) = \theta_0^{(b)}(y) = a_{00} = 1$, and that the first term of (22) is the product of the marginal densities. Then, the remainder of the series (assuming it is summable), when integrated as in (7), is dominated by $\sum_{i,j} |a_{ij}| - a_{00}$.

Sufficient conditions for the expansion (22) to hold have been given by Lancaster ([7], [8]). He shows that if the bivariate density function is φ^2-bounded,

$$\int\int (p_2^2(x, y)/p_a(x)p_b(y))\, dx\, dy - 1 = \varphi^2 < \infty, \tag{23}$$

the series converges in mean square to $p_2(x, y)$. Furthermore, the functions $\{\theta_i^{(a)}(x)\}$, $\{\theta_i^{(b)}(y)\}$ can be chosen as "canonical variables" so that they obey an additional orthogonality condition ([7], Theorem 2)

$$\int\int \theta_i^{(a)}(x)\theta_j^{(b)}(y)p_2(x, y)\, dx\, dy = \delta_{ij}. \tag{24}$$

It then follows ([7], Theorem 3) that the expansion (22) is diagonal, i.e., $a_{ij} = 0$, $i \neq j$. The remaining coefficients, a_{jj}, can be shown to be maximal correlations, and are called canonical correlations. For the gaussian case discussed above, canonical correlations are the corresponding powers of $|\rho|$, and the canonical variables are the Hermite polynomials. Other expansions for φ^2-bounded bivariate densities can be found in the literature ([4], [9], [11]).

Acknowledgment. We acknowledge the assistance of the referee who pointed out a number of references with which the authors were not familiar.

REFERENCES

[1] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[2] BARRETT, J. F. and LAMPARD, D. G. (1955). An expansion for some second-order probability distributions and its applications to noise problems. *IRE Trans. PGIT* **1** 10–15.

[3] CRAMÉR, H. (1947). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[4] EAGLESON, G. K. (1964). Polynomial expansions of bivariate distributions. *Ann. Math. Statist.* **35** 1208–1215.

[5] ERDELYI, A., MAGNUS, W., OBERHETTINGER, F. and TRICOMI, F. G. (1953). *Higher Transcendental Functions*, **2.** Bateman Manuscript Project, McGraw-Hill, New York.

[6] HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika* **28** 321–377.

[7] LANCASTER, H. O. (1958). The structure of bivariate distributions. *Ann. Math. Statist.* **29** 719–736.

[8] LANCASTER, H. O. (1963). Correlations and canonical forms of bivariate distributions. *Ann. Math. Statist.* **34** 532–538.

[9] SARMANOV, O. V. (1963). Investigation of stationary Markov processes by the method of eigenfunction expansion. *Select. Trans. Math. Statist. Prob.* **4** 245–269.

[10] SCHWARTZ, S. (1966). An empirical Bayes technique in communication theory. Tech. Report 02905-16-T, Univ. of Michigan.

[11] WONG, E. and THOMAS, J. B. (1962). On polynomial expansions of second-order distributions. *SIAM* **10** 507–516.