The Annals of Mathematical Statistics
1969, Vol. 40, No. 6, 2180–2182

A BOUND FOR THE VARIATION OF GAUSSIAN DENSITIES[1]

By S. Kullback

The George Washington University

0. Abstract. Schwartz and Root [5] used Mehler's identity to obtain a bound for the integral of the absolute difference between the bivariate Gaussian density function and the product of its corresponding marginal densities. The result was also extended to the case of two dependent Gaussian vectors. The bounds were given in terms of the correlation coefficient in the bivariate case and canonical correlations in the two vector case. In this note an information-theoretic inequality is applied to derive a better bound than reached in [5] and to extend the result to the case of $m > 2$ dependent gaussian vectors. No series expansion is required as in [5].

1. Preliminaries. Let X be a space of points x, S a sigma-algebra of sets of X, P_1 and P_2 probability measures on S. We suppose that P_1 is absolutely continuous with respect to P_2, $P_1 \ll P_2$, so that there exists a probability measure P on S such that $P_1 \ll P$, $P_2 \ll P$. Let us denote the Radon-Nikodym derivatives by

$$f_1(x) = dP_1/dP, \qquad f_2(x) = dP_2/dP \tag{1.1}$$

then the discrimination information is ([3], page 5)

$$I(P_1 ; P_2) = \int_X f_1(x) \log [f_1(x)/f_2(x)] \, dP \tag{1.2}$$

where natural logarithms are used, and the variation is

$$V(P_1, P_2) = \int_X |f_1(x) - f_2(x)| \, dP. \tag{1.3}$$

It was shown in [4] that

$$I(P_1 ; P_2) \geqq V^2(P_1, P_2)/2 + V^4(P_1, P_2)/12 \tag{1.4}$$

but for the purposes of this note we shall use

$$V^2(P_1, P_2) \leqq 2I(P_1 ; P_2). \tag{1.5}$$

If $P_2 \ll P_1$, then it follows as in [4] that

$$V^2(P_1, P_2) \leqq 2I(P_2 ; P_1) \tag{1.6}$$

hence

$$V^2(P_1, P_2) \leqq I(P_1 ; P_2) + I(P_2 ; P_1) = J(P_1, P_2) \tag{1.7}$$

Received 9 January 1969.

[1] Supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under Grant AFOSR-68-1513.

where ([3], page 6)

$$J(P_1, P_2) = \int_X (f_1(x) - f_2(x)) \log [f_1(x)/f_2(x)]\, dP. \tag{1.8}$$

2. Application. Let $\mathbf{x}_1', \mathbf{x}_2', \cdots, \mathbf{x}_m'$ be a set of m dependent Gaussian vectors, where $\mathbf{x}_i'$ is $n_i \times 1$, the means are zero, the covariance matrix of the joint distribution is

$$\mathbf{\Sigma}_1 = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \cdots & \mathbf{\Sigma}_{1m} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \cdots & \mathbf{\Sigma}_{2m} \\ \cdots & & & \cdots \\ \mathbf{\Sigma}_{m1} & \mathbf{\Sigma}_{m2} & & \mathbf{\Sigma}_{mm} \end{pmatrix} \tag{2.1}$$

with $\mathbf{\Sigma}_{ij} = E(\mathbf{x}_i\mathbf{x}_j')\ n_i \times n_j$, $n = n_1 + n_2 + \cdots + n_m$, and the covariance matrix of the product of the marginal distributions, that is, assuming the vectors are independent, is

$$\mathbf{\Sigma}_2 = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{22} & \cdots & \mathbf{0} \\ \cdots & & & \cdots \\ \mathbf{0} & \mathbf{0} & & \mathbf{\Sigma}_{mm} \end{pmatrix}. \tag{2.2}$$

It may be shown that for this case ([3], page 208)

$$\begin{aligned} 2I(P_1 ; P_2) &= \log |\mathbf{\Sigma}_{11}|\cdot|\mathbf{\Sigma}_{22}| \cdots |\mathbf{\Sigma}_{mm}|/|\mathbf{\Sigma}_1| \\ &= \log |\mathbf{R}_{11}|\cdot|\mathbf{R}_{22}| \cdots |\mathbf{R}_{mm}|/|\mathbf{R}_1| \end{aligned} \tag{2.3}$$

and ([3], page 190)

$$\begin{aligned} J(P_1, P_2) &= \tfrac{1}{2} \operatorname{tr} \mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1} + \tfrac{1}{2} \operatorname{tr} \mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1} - n \\ &= \tfrac{1}{2} \operatorname{tr} (\mathbf{\Sigma}_{11}\mathbf{\Sigma}^{11} + \mathbf{\Sigma}_{22}\mathbf{\Sigma}^{22} + \cdots + \mathbf{\Sigma}_{mm}\mathbf{\Sigma}^{mm}) - \tfrac{1}{2}n \\ &= \tfrac{1}{2} \operatorname{tr} (\mathbf{R}_{11}\mathbf{R}^{11} + \mathbf{R}_{22}\mathbf{R}^{22} + \cdots + \mathbf{R}_{mm}\mathbf{R}^{mm}) - \tfrac{1}{2}n \end{aligned} \tag{2.4}$$

where the $\mathbf{R}_{ij}$ are the corresponding matrices of correlation coefficients and the superscript represents the element in the inverse of the joint matrix. In particular, for the bivariate case, $m = 1$, $n_1 = 2$, ([3], page 203)

$$2I(P_1 ; P_2) = -\log (1 - \rho^2), \qquad J(P_1, P_2) = \rho^2/(1 - \rho^2) \tag{2.5}$$

where ρ is the correlation coefficient; and for the case of two vectors $m = 2$, $n_1 + n_2 = n$, $n_2 \leqq n_1$ ([3], page 203)

$$\begin{aligned} 2I(P_1 ; P_2) &= -\log (1 - \rho_1^2)(1 - \rho_2^2) \cdots (1 - \rho_{n_2}^2) \\ J(P_1, P_2) &= \rho_1^2/(1 - \rho_1^2) + \rho_2^2/(1 - \rho_2^2) + \cdots + \rho_{n_2}^2/(1 - \rho_{n_2}^2) \end{aligned} \tag{2.6}$$

where the ρ_i are Hotelling's [1] canonical correlations.

Using convexity properties it may be verified that

$$(-\log (1 - \rho^2))^{\frac{1}{2}} \leqq |\rho|/(1 - |\rho|), \qquad (\rho^2/(1 - \rho^2))^{\frac{1}{2}} \leqq |\rho|/(1 - |\rho|) \tag{2.7}$$

where $|\rho|/(1 - |\rho|)$ is the bound given in [5] for the bivariate case, and that

$$(-\log (1 - \rho_1^2)(1 - \rho_2^2) \cdots (1 - \rho_{n_2}^2))^{\frac{1}{2}} \leqq -1 + \prod_{i=1}^{n_2} (1 - \rho_i)^{-1}$$

$$(\rho_1^2/(1 - \rho_1^2) + \rho_2^2/(1 - \rho_2^2) + \cdots + \rho_{n_2}^2/(1 - \rho_{n_2}^2))^{\frac{1}{2}} \leqq -1 + \prod_{i=1}^{n_2} (1 - \rho_i)^{-1} \tag{2.8}$$

where $-1 + \prod_{i=1}^{n_2} (1 - \rho_i)^{-1}$ is the bound given in [5] for the two vector case and there $n_1 = n_2$.

3. Remarks. Ikeda [2] studied the variation (1.3) between a joint density and products of the marginal densities for properties of asymptotic independence. He used the bound

$$V^2(P_1, P_2) \leqq 4I(P_1 ; P_2) \tag{3.1}$$

instead of (1.5).

It might also be noted that in using the bounds (1.5), (1.7) we are in fact applying the logarithm of the generating function given by the left-hand side of (3) in [5] rather than the expansion itself as was done in [5].

The results (1.5), (1.6) could also be expressed as

$$V^2(P_1, P_2) \leqq \min (2I(P_1 ; P_2), 2I(P_2 ; P_1)). \tag{3.2}$$

REFERENCES

[1] HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika* **28** 321–377.
[2] IKEDA, S. (1963). Asymptotic equivalence of probability distributions with application to some problems of asymptotic independence. *Ann. Inst. Statist. Math.* **15** 87–116.
[3] KULLBACK, S. (1959). *Information Theory and Statistics.* Wiley, New York. 1968 edition, Dover, New York.
[4] KULLBACK, S. (1967). A lower bound for discrimination information in terms of variation. *IEEE Trans. Information Theory* **IT-13** 126–127.
[5] SCHWARTZ, S. C. and ROOT, W. L. (1968). On dominating an average associated with dependent Gaussian vectors. *Ann. Math. Statist.* **39** 1844–1848.