[22] J. L. Doob, *Stochastic Processes.* New York: Wiley, 1953.
[23] T. Kailath and P. Frost, "Mathematical modeling of stochastic processes," (to be published).
[24] R. L. Stratonovich, "A new representation for stochastic integrals," Vestnik Mosk. Univ., ser. I, *Mat. Mekh.*, vol. 1, pp. 3–12 (transl. in *J. SIAM Control*, vol. 4, pp. 362–371, 1966).
[25] F. C. Schweppe, "Evaluation of likelihood functions for Gaussian signals," *IEEE Trans. Information Theory*, vol. IT-11, pp. 61–70, January 1965.
[26] R. L. Stratonovich and Yu. G. Sosulin, "Optimal detection of a Markov process in noise," *Engrg. Cybernet.*, vol. 6, pp. 7–19, October 1964.
[27] T. E. Duncan, "Probability densities for diffusion processes with applications to nonlinear filtering theory and detection theory," Ph.D. dissertation, Dept. of Elec. Engrg., Stanford University, Stanford, Calif., May 1967.
[28] J. Hájek, "Linear statistical problems," *Czech. Math. J.*, vol. 87, pp. 404–444, December 1962.
[29] A. M. Yaglom, "On the equivalence and perpendicularity of two Gaussian probability measures in function space," *Proc. Symp. on Time-Series Analysis*, M. Rosenblatt, Ed. New York: Wiley, 1963, ch. 22.
[30] D. E. Varberg, "On equivalence of Gaussian measures," *Pac. J. Math.*, vol. 11, pp. 751–762, 1961.
[31] T. Kailath and R. Geesey, "Covariance factorization—An explication via examples," *Proc. 2nd Asilomar Conf. on Systems and Circuits* (Monterey, Calif., November 1968).
[32] R. Geesey and T. Kailath, "Application of the canonical representation of second-order processes to estimation and detection in colored noise," *Proc. Symp. on Computer Processing in Communications* (Polytechnic Institute of Brooklyn, Brooklyn, N. Y., April 1969).
[33] T. Kailath, "An RKHS approach to detection and estimation problems, Part II—Simultaneous diagonalization and Gaussian detection," *IEEE Trans. Information Theory* (to be published).
[34] Yu. A. Rozanov, "Infinite-dimensional Gaussian distributions," *Proc. Math. Inst. Acad. Sci.* (Moscow), vol. CV-3, 1968.
[35] T. Kadota and D. Slepian, "Four integral equations of detection theory," *J. SIAM* (to be published).
[36] I. Gohberg and M. G. Krein, "Theory of Volterra operators in Hilbert space," *Izdat. Nauk.*, 1967.
[37] N. Wiener, "The homogeneous chaos," *Am. J. Math.*, vol. 60, pp. 897–936, 1938.
[38] ——, *Nonlinear Problems in the Theory of Random Processes.* Cambridge, Mass.: M.I.T. Technology Press, 1958.
[39] N. Dunford and J. T. Schwarz, *Linear Operators*, vol. 2. New York: Wiley, 1963.
[40] I. Gohberg and M. G. Krein, "Theory of linear non-self-adjoint operators in Hilbert space," *Izdat. Nauk.*, 1965.

Distance Measures and Asymptotic Relative Efficiency

HISASHI KOBAYASHI, MEMBER, IEEE

***Abstract*—The relationship between distance measures and asymptotic relative efficiency is discussed. It is shown that the ratio of the Bhattacharyya distance or J divergences of two test statistics is equivalent to asymptotic relative efficiency. Two-input systems are discussed as examples, and the performances of the polarity coincidence correlator (PCC) and the correlator are discussed in terms of the distance measures of reduced data.**

THE NOTION of a *distance measure* between two probability measures is widely used in statistics. Grettenberg [1], Kailath [2], and Kadota and Shepp [3] discuss the application of some of these measures to communication problems.

Let $p_1(x)$ and $p_2(x)$ be density functions of probability measures $P_1(x)$ and $P_2(x)$ defined over X, a space of observations x, under the hypotheses H_1 and H_2, respectively. Let $L(x)$ be the Radon–Nikodym derivative of P_2 with respect to P_1, i.e., the likelihood ratio

$$L(x) = p_2(x)/p_1(x). \tag{1}$$

Then many of the distance measures currently used can be written in the form [4]

Manuscript received June 11, 1969; revised October 10, 1969. This work was supported by the National Science Foundation under Grants GK-187 and GK-1439 and by the U.S. Army Research Office, Durham, N. C. under Contract DA-31-124-ARO-D-292.

The author is on sabbatical leave from IBM Thomas J. Watson Research Center, Yorktown Heights, N. Y. He is presently with the School of Engineering and Applied Science, University of California, Los Angeles, Calif. 90024.

$$f\{E_1[\phi(L)]\} \tag{2}$$

where $\phi(\cdot)$ is a continuous convex function on $(0, \infty)$, $f(\cdot)$ is an increasing real-valued function of a real variable, and $E_1[\cdot]$ is the expectation under the probability measure P_1. Typical examples are the following.

J Divergence [5]

$$J_x = E_1[(L(x) - 1) \ln L(x)]. \tag{3}$$

Bhattacharyya Distance (*B Distance*) [2], [6]

$$B_x = -\ln \rho_x \tag{4}$$

where ρ_x is defined by

$$\rho_x = \int (p_1(x)p_2(x))^{1/2}\, dx$$
$$= E_1[(L(x))^{1/2}] \tag{5}$$

and is called the Bhattacharyya coefficient [2], or Hellinger integral [3], or affinity [7].

In many situations, it is desirable to consider condensed data rather than the original data, i.e., to deal with a statistic $T(x)$ rather than with x. For a given statistic $T(x)$, the amount of discrimination between H_1 and H_2 provided by $T(x)$ will be defined by the distance B_T or J_T:

$$B_T = -\ln \rho_T \qquad \rho_T = \int \sqrt{p(T \mid H_1)p(T \mid H_2)}\, dT \tag{6}$$

and

$$J_T = E_1\left[\ln \frac{p(T \mid H_1)}{p(T \mid H_2)}\right] + E_2\left[\ln \frac{p(T \mid H_2)}{p(T \mid H_1)}\right]. \qquad (7)$$

It can be shown [4] that each of these measures is no larger than the corresponding expression when T is replaced by the original data x, i.e., $B_T \leqq B_x$ and $J_T \leqq J_x$. The equality in these relations holds if and only if $T(x)$ is a sufficient statistic. Therefore, the ratio of B_T to B_x may represent the efficiency of $T(x)$, and the same thing can be said about the J divergence.

Let T_α and T_β be any two consistent test statistics for testing $H_1 : \theta = \theta_0$ against $H_2 : \theta > \theta_0$ (or $\theta < \theta_0$). Let the data of T_α and T_β consist of n_α and n_β independent observations, respectively. The most frequently used criterion for comparing the efficiency of two statistics is the ARE (asymptotic relative efficiency) defined [8] by

$$A_{\beta\alpha} = \lim_{\theta\to\theta_0} n_\alpha / n_\beta \qquad (8)$$

where the limiting operation is taken under the constraint that the errors of types I and II of T_α and T_β are each kept fixed.

The smaller the distance, the more difficult it will be to descriminate H_2 from H_1, and hence the larger the amount of data that will be required to achieve the assigned performance. Therefore, it is clear that some relationship exists between ARE and the ratio of the distance measures. Assume that T_α is asymptotically normally distributed whatever the value of θ may be, with mean $M_\alpha(\theta)$ and variance

$$D_\alpha^2(\theta) \qquad \theta \geqq \theta_0.$$

Assume that the first $(m-1)$th order derivatives of $M_\alpha(\theta)$ are zero at $\theta = \theta_0$, i.e.,

$$M_\alpha^{(i)}(\theta_0) = 0 \qquad i = 1, 2, \cdots, m-1$$

and $M_\alpha^{(m)}(\theta_0) \neq 0$. Then it can be shown [6], [9] after some manipulation that the B coefficient between $\theta = \theta_0$ and $\theta = \theta_0 + \delta\theta$ is given by

$$\rho_{T_\alpha}(\theta_0, \theta_0 + \delta\theta) = 1 - \frac{1}{8}\left(\frac{\delta\theta^m}{m}\right)^2 I_{\alpha,m} + O(\delta\theta^{3m}) \qquad (9)$$

where

$$I_{\alpha,m} = E\left[\frac{\partial^m \ln p(T_\alpha \mid \theta_0)}{\partial \theta^m}\right]^2 = \left[\frac{M_\alpha^{(m)}(\theta_0)}{D_\alpha(\theta_0)}\right]^2. \qquad (10)$$

Therefore, the B distance provided by the statistic $T_\alpha(x)$ is

$$B_{T_\alpha}(\theta_0, \theta_0 + \delta\theta) = \frac{1}{8}\left(\frac{\delta\theta^m}{m}\right)^2 \cdot I_{\alpha,m} + O(\delta\theta^{3m}). \qquad (11)$$

For $m = 1$, the quantity $I_{\alpha,m}$ is equal to Fisher's information measure [10]; hence, $I_{\alpha,m}$ may be called a generalized Fisher's information measure. On assuming that the asymptotic distribution of T_β behaves essentially in the same fashion as T_α, it follows that

$$B_{T_\beta}/B_{T_\alpha} = \left[\frac{M_\beta^{(m)}(\theta_0)}{D_\beta(\theta_0)}\right]^2 \Big/ \left[\frac{M_\alpha^{(m)}(\theta_0)}{D_\alpha(\theta_0)}\right]^2. \qquad (12)$$

Let

$$\frac{M_\alpha^{(m)}(\theta_0)}{D_\alpha(\theta_0)} \sim c_\alpha \cdot n_\alpha^{m\cdot r} \qquad n_\alpha \to \infty.$$

This relation defines the constants c_α and r. We define the *efficacy* of T_α by

$$E_\alpha = \lim_{n_\alpha\to\infty}\left[\frac{M_\alpha^{(m)}(\theta_0)}{n_\alpha^{mr}\, m!\, D_\alpha(\theta_0)}\right]^{1/m\cdot r} < \infty. \qquad (13)$$

The efficacy defined here is a generalization of the term given by Capon [11]. Then, keeping $n_\alpha = n_\beta$, we have the following asymptotic relation:

$$\lim_{\delta\theta\to 0} \frac{B_{T_\beta}(\theta_0, \theta_0 + \delta\theta)}{B_{T_\alpha}(\theta_0, \theta_0 + \delta\theta)} = (E_\beta/E_\alpha)^{2mr} = A_{\beta\alpha}^{2m\cdot r}. \qquad (14)$$

If $m = 1$ and $r = \frac{1}{2}$, as for many problems, then B_{T_α} approaches $A_{\beta\alpha}$ as $\delta\theta$ approaches zero.

Similar results can be obtained for the J divergence, i.e.,

$$J_{T_\alpha}(\theta_0, \theta_0 + \delta\theta) = \left(\frac{\delta\theta^m}{m}\right)^2 \cdot I_{\alpha,m} + O(\delta\theta^{3m}) \qquad (15)$$

and hence

$$J_{T_\alpha}(\theta_0, \theta_0 + \delta\theta) = 8B_{T_\alpha}(\theta_0, \theta_0 + \delta\theta) \qquad \delta\theta \to 0. \qquad (16)$$

Therefore, it follows that

$$B_{T_\beta}/B_{T_\alpha} = J_{T_\beta}/J_{T_\alpha} \to A_{\beta\alpha}^{2m\cdot r} \qquad \delta\theta \to 0. \qquad (17)$$

Let us consider a detection problem in a two-input system, where the stochastic signal $s(t)$ is common to both channels, say "channel a" and "channel b." Noises are additive and independent, i.e.,

$$\begin{aligned} x_a(t) &= \theta^{1/2} s(t) + n_a(t) \\ x_b(t) &= \theta^{1/2} s(t) + n_b(t) \end{aligned} \qquad t = 1, 2, \cdots, N \qquad (18)$$

where $s(t)$, $n_a(t)$, and $n_b(t)$ are real Gaussian discrete processes, independent of each other and with zero mean and correlation

$$E[s(t)s(t')] = P_s(t)\, \delta_{t,t'}$$

$$E[n_a(t)n_a(t')] = E[n_b(t)n_b(t')] = P_n(t)\, \delta_{t,t'}.$$

Here θ is the signal-to-noise ratio parameter. Consider the hypothesis testing problem: $\theta = 0$ against $\theta > 0$. The B coefficient between the two hypotheses based on the observed data

$$x_a(t) \text{ and } x_b(t) \qquad t = 1, 2, \cdots, N$$

is given by

$$\rho_x(0, \theta) = \prod_{t=1}^{N} [1 + 2\theta \cdot \mu(t)]^{1/4} [1 + \theta \cdot \mu(t)]^{-1/2} \qquad (19)$$

where $\mu(t) = P_s(t)/P_n(t)$. Therefore, the B distance is given by

$$B_x(0, \theta) = \frac{1}{2} \sum_{t=1}^{N} [\ln (1 + \theta \cdot \mu(t)) - \tfrac{1}{2} \ln (1 + 2\theta \cdot \mu(t))]. \tag{20}$$

Now consider the case where the data $x_a(t)$ and $x_b(t)$ are passed through hard limiters, the outputs of which will be denoted by $y_a(t)$ and $y_b(t)$:

$$y_a(t) = \operatorname{sgn} x_a(t)$$

and

$$y_b(t) = \operatorname{sgn} x_b(t) \qquad t = 1, 2, \cdots, N.$$

The B coefficient and the B distance provided by $y_a(t)$ and $y_b(t)$ are given, after some manipulation [9], by

$$\rho_y(0, \theta) = \prod_{t=1}^{N} \frac{1}{2} \left\{ \left[1 + \frac{2}{\pi} \sin^{-1} \frac{\theta \cdot \mu(t)}{1 + \theta \cdot \mu(t)} \right]^{1/2} + \left[1 - \frac{2}{\pi} \sin^{-1} \frac{\theta \cdot \mu(t)}{1 + \theta \cdot \mu(t)} \right]^{1/2} \right\} \tag{21}$$

and

$$B_y(0, \theta) = - \ln \rho_y(0, \theta). \tag{22}$$

Now consider another reduction of the data obtained by defining a new random process $z(t)$ as the product of the outputs of the two channels, i.e.,

$$z(t) = x_a(t) \cdot x_b(t) \qquad t = 1, 2, \cdots, N. \tag{23}$$

The distribution density of $z(t)$ is known [12] and is given by

$$p(z(t) \mid \theta) = \frac{1}{\pi} \sqrt{|W(t)|}\, e^{-w_{ab}(t) z(t)} \cdot K_0(|z(t)| \sqrt{w_{aa}(t) w_{bb}(t)}). \tag{24}$$

Here $W(t)$ is a 2-by-2 matrix function and is the inverse of the covariance matrix function of the vector process

$$\begin{bmatrix} x_a(t) \\ x_b(t) \end{bmatrix},$$

i.e.,

$$W(t) = \begin{bmatrix} w_{aa}(t) & w_{ab}(t) \\ w_{ba}(t) & w_{bb}(t) \end{bmatrix} = M^{-1}(t) \tag{25}$$

where

$$M(t) = E \begin{bmatrix} x_a(t) \\ x_b(t) \end{bmatrix} \begin{bmatrix} x_a(t) \\ x_b(t) \end{bmatrix}^T = P_n(t) \begin{bmatrix} \theta \cdot \mu(t) + 1 & \theta \cdot \mu(t) \\ \theta \cdot \mu(t) & \theta \cdot \mu(t) + 1 \end{bmatrix}. \tag{26}$$

The function $K_0(\cdot)$ in (24) is the modified Bessel function of the second kind and of order zero. The B coefficient based on

$$z(t) \qquad t = 1, 2, \cdots, N,$$

can be calculated for weak signals, i.e., for $\theta \ll 1$, as follows:

$$\rho_z(0, \theta) = \sum_{t=1}^{N} (1 + 2\theta\mu(t))^{-1/4} \cdot \left[\frac{1 + \theta\mu(t)}{1 + 2\theta\mu(t)} - \left\{ \frac{\theta\mu(t)}{2(1 + 2\theta\mu(t))} \right\}^2 \right]^{-1/2} \tag{27}$$

and

$$B_z(0, \theta) = - \ln \rho_z(0, \theta). \tag{28}$$

Let $\theta \ll 1$ in (20), (22), and (26); then

$$B_x(0, \theta) = \frac{\theta^2}{4} \sum_{t=1}^{N} \mu^2(t), \tag{29}$$

$$B_y(0, \theta) = \frac{\theta^2}{2\pi^2} \sum_{t=1}^{N} \mu^2(t), \tag{30}$$

$$B_z(0, \theta) = \frac{\theta^2}{8} \sum_{t=1}^{N} \mu^2(t). \tag{31}$$

Taking the ratio of these distances, we obtain $B_y/B_x = \frac{1}{2}(2/\pi)^2$ and $B_z/B_x = \frac{1}{2}$. These quantities are equal, respectively, to the ARE of the PCC and the correlator, with respect to the optimum detector [13]. This is the case since the optimum detector T_{opt}, the polarity coincidence correlator T_{PCC}, and the correlator T_c are sufficient test statistics (for $\theta \ll 1$) of x, y, and z, respectively, for distinguishing between the two hypotheses. These systems are given by

$$T_{\text{opt}} = \sum_{t=1}^{n} \mu(t) \cdot \{x_a(t) + x_b(t)\}^2, \tag{32}$$

$$T_{\text{PCC}} = \sum_{t=1}^{n} \mu(t) \cdot y_a(t) \cdot y_b(t), \tag{33}$$

$$T_c = \sum_{t=1}^{n} \mu(t) \cdot z(t). \tag{34}$$

Acknowledgment

The author would like to thank Prof. J. B. Thomas of Princeton University, Princeton, N. J. for his guidance and encouragement during the course of this work.

References

[1] T. L. Grettenberg, "Signal selection in communication and radar systems," *IEEE Trans. Information Theory*, vol. IT-9, pp. 265–275, October 1963.

[2] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Communication Technology*, vol. COM-15, pp. 52–60, February 1967.

[3] T. T. Kadota and L. A. Shepp, "On the best finite set of linear observables for discriminating two Gaussian signals," *IEEE Trans. Information Theory*, vol. IT-13, pp. 278–284, April 1967.

[4] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. Roy. Stat. Soc.*, vol. 28, pp. 131–142, 1966.

[5] S. Kullback, *Information Theory and Statistics.* New York: Wiley, 1959, p. 6.

[6] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, vol. 35, pp. 99–109, 1943.

[7] K. Matsushita, "A distance and related statistics in multivariate analysis," in *Multivariate Analysis*, P. R. Krishnaiah, Ed. New York: Academic Press, 1966, pp. 187–200.

[8] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics. Inference and Relationship*, vol. 2. New York: Hafner, 1961.

[9] H. Kobayashi, "Distance measures and related criteria," in

Representations of Complex-Valued Vector Processes and Their Application to Estimation and Detection, Ph.D. dissertation, Princeton University, Princeton, N. J., August 1967.
[10] C. R. Rao, *Linear Statistical Inference and Its Applications*. New York: Wiley, 1965.
[11] J. Capon, "On the asymptotic efficiency of locally optimum detectors," *IRE Trans. Information Theory*, vol. IT-7, pp. 67–71, April 1961.
[12] K. S. Miller, *Multidimensional Gaussian Distributions*. New York: Wiley, 1964, p. 45.
[13] J. W. Carlyle and J. B. Thomas, "On nonparametric signal detectors," *IEEE Trans. Information Theory*, vol. IT-10, pp. 146–152, April 1964.
[14] H. Kobayashi and J. B. Thomas, "Distance measures and related criteria," presented at the 5th Ann. Allerton Conf. Circuits and Systems, University of Illinois, Urbana, 1967.

Nonsingular Detection and Likelihood Ratio for Random Signals in White Gaussian Noise

T. T. KADOTA

***Abstract*—This paper is concerned with the mathematical aspect of a detection problem (a random signal in white Gaussian noise). Specifically, we obtain a sufficient condition for nonsingular detection and derive a likelihood-ratio expression in terms of least-mean-square estimates. The problem itself is old, and the likelihood-ratio expression is also well known. The contribution of this paper is a relatively elementary and self-contained derivation of the likelihood-ratio expression as well as the nonsingularity condition.**

I. Introduction

Consider a problem of optimally detecting a random signal in white Gaussian noise. One mathematical treatment of such a problem is to interpret it in terms of the integrated signal and noise. Thus, the signal portion is a time integral of the given signal process, and the noise a standard Wiener process. We regard the detection problem as one of discrimination between the signal-plus-noise and the noise processes. When optimality of this discrimination is defined in the sense of the Neyman–Pearson hypothesis test, the solution of the problem consists of obtaining a sufficient condition for nonsingularity of the two processes and expressing their likelihood ratio in terms of the observable.

It was conjectured that when the signal-plus-noise measure is absolutely continuous with respect to the noise measure, the likelihood ratio takes the same form as the one in the sure signal-in-noise problem, except for the fact that the signal is replaced by its least-mean-square estimate. Inasmuch as our interest is in the mathematical proof, we refer to Kailath [1] for the historical account and physical interpretation of this likelihood-ratio expression. In a special case where the signal is a diffusion process, the conjecture was proved by Duncan [2]. Kailath [1] and Wong [3] gave more general proofs under the condition that the signal has finite average energy and is independent of the noise. Kailath [4] later replaced the independence part of the condition by a weaker condition that the signal-plus-noise and the noise measures are equivalent. In applications, finiteness of average signal energy is not restrictive, but mutual independence between the signal and the noise is a serious restriction. For example, in communication systems with feedback the signal is a function of the past noise; thus, it necessarily depends on the noise. Unfortunately, this dependence makes the two measures no longer equivalent. The purpose of this paper is to establish the likelihood-ratio expression under a much weaker independence condition.

We prove that if the signal has finite energy with probability 1 and if it is independent of future increments of a delayed version of the noise (delay can be arbitrarily small), then the signal-plus-noise measure is absolutely continuous with respect to the noise measure; and if, in addition the expectation of the signal energy is finite, the likelihood ratio is given effectively by the same expression. We remark that in general neither of the first two conditions is necessary for absolute continuity. For example, if the signal is Gaussian, there is Shepp's necessary and sufficient condition [5], which is weaker than our two. In fact, we explicitly show that our first two conditions imply his, and our third coincides with the first in the case of Gaussian signals. In applications, however, it is inconceivable that a signal should have infinite average energy in a finite time interval. Also, dependence of a signal on additive noise is typically through feedback, which necessarily introduces some delay. Thus, though mathematically restrictive, these conditions seem physically acceptable.

Since this paper was submitted, our second condition has been relaxed by eliminating the delay [6]. Furthermore, assuming absolute continuity rather than equivalence, Kailath and Zakai (private communication) have verified

Manuscript received May 21, 1969; revised October 13, 1969.
The author is with Bell Telephone Laboratories, Inc., Murray Hill, N. J. 07971.