DISTANCE MEASURES AND RELATED CRITERIA

HISASHI KOBAYASHI AND JOHN B. THOMAS
Princeton University, Princeton, New Jersey

ABSTRACT

Distance measures and related criteria in estimation and detection are discussed, and the Bhattacharyya distance and J-divergence are studied in detail. One of the results is that the ratio of the distance measures of two test statistics is equivalent to asymptotic relative efficiency. Two-input systems are discussed as examples.

1. SOME DISTANCE MEASURES AND THEIR COMMON PROPERTIES

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. In this treatment distance measures and related criteria in statistical communication theory are considered.

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.,

$$L(x) = \frac{p_2(x)}{p_1(x)} \tag{1.1}$$

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

$$f\{E_1[\phi(L)]\} \tag{1.2}$$

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:

(a) Kullback-Leibler numbers [5]

$$I(1,2) = \int p_1(x) \ln \frac{p_1(x)}{p_2(x)} dx = E_1[-\ln L] \tag{1.3}$$

$$I(2,1) = \int p_2(x) \ln \frac{p_2(x)}{p_1(x)} dx = E_1[L \ln L] \tag{1.4}$$

(b) J-divergence [5]

$$J = I(1,2) + I(2,1) = E_1[(L-1) \ln L] \tag{1.5}$$

(c) Matsushita's measure of distance [6]

$$d = [\int (\sqrt{p_2(x)} - \sqrt{p_1(x)})^2 dx]^{1/2} = [E_1(\sqrt{L}-1)^2]^{1/2} \tag{1.6}$$

(d) Kolmogorov's measure of variational distance

$$K = \frac{1}{2} \int |p_2(x) - p_1(x)| dx = \frac{1}{2} E_1|L - 1| \tag{1.7}$$

(e) Bhattacharyya distance (B-distance) [2]

$$B = - \ln \rho \tag{1.8}$$

where ρ is defined by

$$\rho = \int \sqrt{p_1(x)\ p_2(x)}\ dx = E_1[\sqrt{L}\ ] \qquad (1.9)$$

and is called the Bhattacharyya coefficient [2], or Hellinger integral [3], or affinity [6].

An important theorem can now be derived which relates the Bayes risk to quantities of the form of Eq. (1.2) and which has direct application to signal selection problems: Let $\pi = (\pi_1, \pi_2)$ be the prior probability set where π_i is the prior probability of the hypothesis H_i, $i = 1,2$. Let C_{ij} be the cost assigned if H_i is chosen when H_j is true. Let X_i be a subspace of X such that $X_i = \{x | H_i$ is chosen when x observed$\}$, $i = 1,2$. If the criterion is minimization of average risk, then by the well-known Neyman-Pearson lemma, X_i should be chosen as follows:

$$X_1 = \{x | L(x) < \frac{\pi_1(C_{12} - C_{11})}{\pi_2(C_{21} - C_{22})}\} \text{ and } X_2 = X \sim X_1$$

Let α and β be two signal sets (or communication systems, in general) which are to be compared under the criterion of the Bayes risk. The Bayes risk when the signal α is adopted, is given by

$$R(\alpha,\pi) = \pi_1 C_{12} + \pi_2 C_{22} + \int_{X_{\alpha 1}} [\pi_2 C_2' L_\alpha(x) + \pi_1 C_1'] p_{\alpha 1}(x)\, dx \qquad (1.10)$$

where $C_1' = C_{11} - C_{12}$, $C_2' = C_{21} - C_{22}$ and where $X_{\alpha 1}$, $L_\alpha(x)$, etc. are self-explanatory. On defining $R(\beta,\pi)$ in exactly the same way, the following theorem can be stated.

Theorem 1

$R(\alpha,\pi) \le R(\beta,\pi)$ for all sets of prior probability $\pi = (\pi_1, \pi_2)$ if, and only if

$$E_1[\phi(L_\alpha)] \ge E_1[\phi(L_\beta)] \qquad (1.11)$$

for all continuous convex functions $\phi(\cdot)$.

Proof

The proof will be found elsewhere [8] and can be obtained by generalizing the argument given by Karlin and Bradt [7].//

On letting $P_e(\alpha,\pi)$ and $P_e(\beta,\pi)$ be the error probability under the signal sets α and β, respectively, the following statement can be derived from Theorem 1:

Corollary 1

$P_e(\alpha,\pi) \le P_e(\beta,\pi)$ for all sets of prior probability $\pi = (\pi_1, \pi_2)$ if and only if

$$E_1[\phi(L_\alpha)] \ge E_2[\phi(L_\beta)]$$

for all continuous convex functions $\phi(\cdot)$.

Proof

See [7,8].

By contrapositive relations, Theorem 1 and Corollary 1 are equivalent to the following:

Corollary 2

$R(\alpha,\pi) > R(\beta,\pi)$ or, equivalently, $P_e(\alpha,\pi) > P_e(\beta,\pi)$ for some set of prior probability $\pi = (\pi_1, \pi_2)$ if and only if

$$E_1[\phi(L_\alpha)] < E_1[\phi(L_\beta)]$$

for some continuous convex function $\phi(\cdot)$.

Grettenberg [1] and Kailath [2] applied Corollary 2 to the cases of the J-divergence and the B-distance, respectively and discussed the signal selection problem.

2. THE BHATTACHARYYA COEFFICIENT AND THE BAYES RISK

It will be shown in this section that an upper and lower bound of the Bayes risk is given in terms of the Bhattacharyya coefficient ρ, which is a generalization of the inequality relation for the probability of error given by Kailath [2]. The Bayes Risk under the probability set π is

$$R(\pi) = \pi_1 C_{12} + \pi_2 C_{22} + \int_{X_1} [\pi_2 C_2' p_2(x) - \pi_1 C_1' p_1(x)]dx \quad (2.1)$$

Consider the following quantity which may be regarded as generization of Kolmogorov's measure of variational distance:

$$K(\pi) = \frac{1}{2}\int_X |\pi_1 C_1' p_1(x) - \pi_2 C_2' p_2(x)|dx$$
$$= \frac{1}{2}\pi_1(C_{11}+C_{12}) + \frac{1}{2}\pi_2(C_{22}+C_{21}) - R(\pi) \quad (2.2)$$

An upper bound of $K(\pi)$ is obtained from the Cauchy-Schwartz inequality as follows:

$$K(\pi) \le \frac{1}{2}[(\pi_1 C_1' + \pi_2 C_2')^2 - 4\pi_1\pi_2 C_1' C_2' \rho^2]^{1/2} \quad (2.3)$$

A lower bound of $K(\pi)$ is

$$K(\pi) \ge \frac{1}{2}[\pi_1 C_1' + \pi_2 C_2' - 2\sqrt{\pi_1\pi_2 C_1' C_2'}\ \rho] \quad (2.4)$$

From Eqs. (2.2) - (2.4) an upper and lower-bound of the Bayes risk is given by

$$R_0(\pi) + \frac{2\pi_1\pi_2 C_1' C_2' \rho^2}{\pi_1 C_1' + \pi_2 C_2' + \sqrt{(\pi_1 C_1' + \pi_2 C_2')^2 - 4\pi_1\pi_2 C_1' C_2' \rho^2}} \le R(\pi) \le R_0(\pi) + \sqrt{\pi_1\pi_2 C_1' C_2'}\ \rho \quad (2.5)$$

where $R_0(\pi) = \pi_1 C_{11} + \pi_2 C_{22}$.

Let us adopt a simpler but less strict lower bound:

$$R_0(\pi) + \frac{\pi_1\pi_2 C_1' C_2'}{\pi_1 C_1' + \pi_2 C_2'}\cdot\rho^2 \le R(\pi) \le R_0(\pi) + \sqrt{\pi_1\pi_2 C_1' C_2'}\ \rho \quad (2.6)$$

When $C_{11} = C_{22} = 0$, $C_{12} = C_{21} = 1$, the average loss is equal to $P_e(\pi)$, the error probability, and Eqs. (2.5) and (2.6) become

$$\frac{2\pi_1\pi_2\rho^2}{1 - \sqrt{1 - 4\pi_1\pi_2\rho^2}} \le P_e(\pi) \le \sqrt{\pi_1\pi_2}\cdot\rho \quad (2.8)$$

and

$$\pi_1\pi_2\rho^2 \le P_e(\pi) \le \sqrt{\pi_1\pi_2}\cdot\rho \quad (2.8)$$

Thus we see that Eq. (2.5) - (2.9) make the minimization of ρ a reasonable criterion of optimality. Minimization of ρ is equivalent to maximization of the B-distance defined by Eq. (1.8).

Another distance measure frequently used is the J-divergence of Eq. (1.5). Unlike the B-distance, no upper bound on $P_e(\pi)$ in terms of J has been given[1], although a lower bound is

$$P_e(\pi) \geq \pi_1\pi_2 \ \rho^2 > \pi_1\pi_2 \ \exp(-J/2) \tag{2.9}$$

Because of this fact, Kailath [2] claims that the B-distance is a better criterion for signal selection problems than J-divergence. Another superiority of the B-distance is shown in cases where the average with respect to unknown parameters is required, such as in non-coherent reception problems. The next section is intended to demonstrate this fact.

3. APPLICATION OF THE B-DISTANCE TO COMMUNICATION PROBLEMS

Example 3.1 Detection of a Signal with Unknown Phase and Amplitude

Let the observable be the continuous signal x(t) such that

$$\text{under } H_1 \ : \ x(t) = n(t) \tag{3.1}$$

$$\text{under } H_2 \ : \ x(t) = \alpha s(t) + n(t) \tag{3.2}$$

where s(t) is the complex envelope of a known signal and α is a complex Gaussian variable [9] with probability distribution

$$p(\alpha)\,d\alpha d\alpha^* = \frac{1}{2\pi D_\alpha^2} \ \exp(-\frac{\alpha\ \alpha^*}{D_\alpha^2})\,d\alpha d\alpha^* \tag{3.3}$$

The process n(t) is the complex envelope of a stationary Gaussian noise process, with zero mean and covariance

$$E[n(s)n^*(t)] = R(s-t) \tag{3.4}$$

The B-coefficient between H_1 and H_2, given α, can be calculated as

$$\rho(\alpha) = \exp\ (-\frac{1}{4}\|\, s(t)\,\|^2_{H(R)} \ \alpha\ \alpha^*) \tag{3.5}$$

where $\|\, s(t)\,\|^2_{H(R)}$ is the norm square of s(t) on the reproducing kernel Hilbert space generated by the covariance function R(s-t). On taking the expectation of Eq. (3.5) with respect to α, we have

$$E_\alpha\{\rho(\alpha)\} = \frac{1}{1+\frac{1}{4}\mu} \tag{3.6}$$

where

$$\mu = \|s(t)\|^2_{H(R)} \cdot D_\alpha^2$$

Here the quantity μ may be regarded as the S/N ratio. Then the B-distance is

$$B = \ln(1+\frac{1}{4}\mu) \tag{3.7}$$

This result can be applied readily to a signal selection problem in the same way as Grettenberg [1] applied the J-divergence to the detection of a completely known signal [8].

---

[1] For the particular case of selection between two Gaussian probability measures, we have [3]

$$\sqrt{\pi_1\pi_2}\ \exp\{(J/2)^{-1/4}\} > P_e(\pi) \geq \pi_1\pi_2\ \exp(-J/4)$$

Example 3.2 Radar Parameters Estimation

The following problem was originally studied by Grettenberg[1] by the J-divergence method. Let τ and λ be the time delay and doppler-shift frequency of the returned echo. The τ-λ plane can be divided into a set of resolution cells, which Grettenberg calls a message space M:

$$M = \{m = (\tau_m, \lambda_m)\} \tag{3.8}$$

Let w(t) be the complex envelope of the transmitted signal. Assuming that the amplitude is known and is common to all cells m in the range of interest, the observed data under the hypothesis H_m is given by

$$x(t) = e^{j\theta_m} w(t-\tau_m) e^{j2\pi\lambda_m t} + n(t) = e^{j\theta_m} w_m(t) + n(t) \tag{3.9}$$

The phase θ_m is not measured by the observer and hence is assumed to be a random variable, uniformly distributed between 0 and 2π.

The B-coefficient between H_m and H_q ($m,q \in M$) is given after some manipulation [8] as follows:

$$\rho(H_m, H_q) = \exp\left[-\frac{1}{4} \|w_m(t) - w_q(t)\|^2_{H(R)}\right] \tag{3.10}$$

The reproducing kernel inner product which appears in Eq. (3.10) can be written in terms of the generalized ambiguity function as follows:

$$\begin{aligned}[w_m(t), w_q(t)]_{H(R)} &= [w(t-\tau_m)e^{j2\pi\lambda_m t}, w(t-\tau_q)e^{j2\pi\lambda_q t}]_{H(R)} \cdot e^{-j(\theta_m-\theta_q)} \\ &= \chi_w(\tau_m-\tau_q, \lambda_m-\lambda_q)\, e^{-j\theta_{mq}}\end{aligned} \tag{3.11}$$

where

$$\theta_{mq} = \theta_m - \theta_q \tag{3.12}$$

Then Eq. (3.10) can be written as

$$\rho(H_m, H_q) = \exp\{-\frac{1}{2}[\chi_w(0,0) - |\chi_w(\tau_m-\tau_q, \lambda_m-\lambda_q)|\cos\theta]\} \tag{3.13}$$

where $\theta = \theta_{mq} - \arg.\ \chi_w(\tau_m-\tau_q, \lambda_m-\lambda_q)$ (3.14)

Now the averaged B-coefficient is given by

$$E_\theta\{\rho(H_m, H_q)\} = \exp\{-\frac{1}{2}\chi_w(0,0)\} I_0[\frac{1}{2}|\chi_w(\tau_m-\tau_q, \lambda_m-\lambda_q)|] \tag{3.15}$$

Therefore the B-distance is

$$B(H_m, H_q) = \frac{1}{2}\chi_w(0,0) - \ln I_0[\frac{1}{2}|\chi_w(\tau_m-\tau_q, \lambda_m-\lambda_q)|] \tag{3.16}$$

The maximization of the B-distance is equivalent to the minimization of the absolute value of the ambiguity function; this latter approach is well known in radar signal design.

4. DISTANCE MEASURES AND ASYMPTOTIC RELATIVE EFFICIENCY

Thus far we have considered the distance between two probability measures defined on the space of the observed data x and the relation of this distance to performance criterion such as the Bayes risk and error probability. Let us turn our attention from the observed raw data to condensed data, i.e., to a statistic T(x). Estimation and detection procedures can be regarded as data reduction processing and can enable us to discriminate between alternative situations.

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 or J:

$$B_T = -\ln\rho_T \quad , \quad \text{where } \rho_T = \int \sqrt{p(T/H_1)\,p(T/H_2)}\; dT \qquad (4.1)$$

and

$$J_T = E_1[\ln \frac{p(T/H_1)}{p(T/H_2)}] + E_2[\ln \frac{p(T/H_2)}{p(T/H_1)}] \qquad (4.2)$$

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., $B_T \le B_x$ and $J_T \le J_x$. 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: $\theta = \theta_o$ against H_2: $\theta > \theta_o$ (or $\theta < \theta_o$). 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 A.R.E. (Asymptotic Relative Efficiency) defined [10] by

$$A_{\beta\alpha} = \lim_{\theta\to\theta_o} \frac{n_\alpha}{n_\beta} \qquad (4.5)$$

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

The smaller the distance, the more difficult it will be to discriminate 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 A.R.E. and 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 $D^2_\alpha(\theta)$, $\theta \ge \theta_o$. Assume that the first (m-1)st order derivatives of $M_\alpha(\theta)$ are zero at $\theta = \theta_o$, i.e., $M_\alpha^{(i)}(\theta_o) = 0$ for $i = 1,2,\ldots,m-1$ and $M_\alpha^{(m)}(\theta_o) \neq 0$. Then it can be shown [8] after some manipulation that the B-coefficient between $\theta = \theta_o$ and $\theta = \theta_o + \delta\theta$ is given by

$$\rho_{T_\alpha}(\theta_o, \theta_o+\delta\theta) = 1 - \frac{1}{8}(\frac{\delta\theta^m}{m!})^2 \cdot I_{\alpha,m} + O(\delta\theta^{3m}) \qquad (4.6)$$

where

$$I_{\alpha,m} = E(\frac{\partial^m \ln p(T_\alpha/\theta_o)}{\partial\theta^m})^2 = [\frac{M_\alpha^{(m)}(\theta_o)}{D_\alpha(\theta_o)}]^2 \qquad (4.7)$$

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

$$B_{T_\alpha}(\theta_o, \theta_o+\delta\theta) = \frac{1}{8}\left(\frac{\delta\theta^m}{m!}\right)^2 \cdot I_{\alpha,m} + O(\delta\theta^{3m}) \qquad (4.8)$$

For the case m = 1, the quantity $I_{\alpha,m}$ is equal to Fisher's information measure; 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

$$B_{T_\beta} / B_{T_\alpha} = \left[\frac{M_\beta^{(m)}(\theta_o)}{D_\beta(\theta_o)}\right]^2 \Big/ \left[\frac{M_\alpha^{(m)}(\theta_o)}{D_\alpha(\theta_o)}\right]^2 \qquad (4.9)$$

Let $\frac{M_\alpha^{(m)}(\theta_o)}{D_\alpha(\theta_o)} \sim c_\alpha \cdot n_\alpha^{m\cdot r}$ as $n_\alpha \to \infty$. This relation defines the constants c_α and r. We define the <u>efficacy</u> of T_α by

$$E_\alpha = \lim_{n_\alpha \to \infty} \left[\frac{M_\alpha^{(m)}(\theta_o)}{n_\alpha^{mr}\ m!\ D_\alpha(\theta_o)}\right]^{\frac{1}{m\cdot r}} < \infty \qquad (4.10)$$

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:

$$\lim_{\delta\theta\to 0} \frac{B_{T_\beta}(\theta_o, \theta_o+\delta\theta)}{B_{T_\alpha}(\theta_o, \theta_o+\delta\theta)} = \frac{E_\beta}{E_\alpha}^{2m\cdot r} = A_{\beta\alpha}^{2m\cdot r} \qquad (4.11)$$

If m = 1 and $r = \frac{1}{2}$, as is the case for many problems, then $\frac{B_{T_\beta}}{B_{T_\alpha}} \to A_{\beta\alpha}$. Similar results can be obtained for the J-divergence, i.e.,

$$J_{T_\alpha}(\theta_o, \theta_o+\delta\theta) = \left(\frac{\delta\theta^m}{m!}\right)^2 \cdot I_{\alpha,m} + O(\delta\theta^{3m}) \qquad (4.12)$$

and hence

$$J_{T_\alpha}(\theta_o, \theta_o+\delta\theta) = 8\ B_{T_\alpha}(\theta_o, \theta_o+\delta\theta) \quad \text{as } \delta\theta\to 0 \qquad (4.13)$$

Therefore it follows that

$$B_{T_\beta} / B_{T_\alpha} = J_{T_\beta}/J_{T_\alpha} \to A_{\beta\alpha}^{2m\cdot r} \qquad (4.14)$$

5. THE DETECTION PROBLEM IN TWO-INPUT SYSTEMS

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

$$\begin{aligned} x_a(i) &= \sqrt{\theta}\ s(t) + n_a(t) \\ x_b(i) &= \sqrt{\theta}\ s(t) + n_b(t) \end{aligned} \quad , \quad i = 1,2,\ldots,N \qquad (5.1)$$

where s(i), $n_a(i)$ and $n_b(i)$ are real Gaussian random processes, independent of each other and with zero mean and correlation $E[s(i)s(j)] = P_s(i)\ \delta_{ij}$, $E[n_a(i)n_a(j)] = E[n_b(i)n_b(j)] = P_n(i)\delta_{ij}$

Here θ is the signal-to-noise ratio parameter. Consider the testing hypothesis problem: $\theta = 0$ against $\theta > 0$. The B-coefficient between the two hypotheses, based on the observed data $x_a(i)$ and $x_b(i)$, $i = 1,2,\ldots,N$ is given by

$$\rho_x(0,\theta) = \prod_{i=1}^{N} [1 + 2\theta\cdot\mu(i)]^{1/4}\ [1 + \theta\cdot\mu(i)]^{-1/2} \qquad (5.2)$$

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

$$B_x(0,\theta) = \frac{1}{2}\sum_{i=1}^{N} [\ln(1+\theta\mu(i)) - \frac{1}{2}\ln(1+2\theta\mu(i))] \qquad (5.3)$$

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

$$y_a(i) = \operatorname{sgn} x_a(i) \quad \text{and} \quad y_b(i) = \operatorname{sgn} x_b(i)\ ,\quad i = 1,2,\ldots,N$$

The B-coefficient and the B-distance provided by $y_a(i)$ and $y_b(i)$ are given, after some manipulation [8], by

$$\rho_y(0,\theta) = \prod_{i=1}^{N}\frac{1}{2}\left\{\sqrt{1+\frac{2}{\pi}\sin^{-1}\frac{\theta\cdot\mu(i)}{1+\theta\cdot\mu(i)}} + \sqrt{1-\frac{2}{\pi}\sin^{-1}\frac{\theta\cdot\mu(i)}{1+\theta\cdot\mu(i)}}\right\} \qquad (5.4)$$

and

$$B_y(0,\theta) = -\ln \rho_y(0,\theta) \qquad (5.5)$$

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

$$z(i) = x_a(i)\cdot x_b(i)\quad ,\quad i = 1,2,\ldots,N \qquad (5.6)$$

The distribution density of $z(i)$ is given, after some manipulation [8], by

$$p(z(i)/\theta) = \frac{1}{\pi P_n(i)\sqrt{1+2\theta\mu(i)}}\ \exp\left\{\frac{\theta\mu(i)z(i)}{(1+2\theta\mu(i))\cdot P_n(i)}\right\} \times \qquad (5.7)$$

$$K_o\left(\frac{(1+\theta\mu(i))\,|z(i)|}{(1+2\theta\mu(i))\cdot P_n(i)}\right)$$

where $K_0(\cdot)$ is the modified Bessel function of the second kind and of order zero. The B-coefficient based on $z(i)$, $i = 1,2,\ldots,N$ can be calculated for the weak signal case, i.e., for $\theta \ll 1$, as follows:

$$\rho_z(0,\theta) = \sum_{i=1}^{N}(1+2\theta\mu(i))^{-1/4}\left[\frac{1+\theta\mu(i)}{1+2\theta\mu(i)} - \left\{\frac{\theta\mu(i)}{2(1+2\theta\mu(i))}\right\}^2\right]^{-1/2} \qquad (5.8)$$

and

$$B_z(0,\theta) = -\ln \rho_z(0,\theta) \tag{5.9}$$

Let $\theta \ll 1$ in Eqs. (5.3), (5.5) and (5.9); then we will have

$$B_x(0,\theta) = \frac{\theta^2}{4} \sum_{i=1}^{N} \mu^2(i) \tag{5.10}$$

$$B_y(0,\theta) = \frac{\theta^2}{2\pi^2} \sum_{i=1}^{N} \mu^2(i) \tag{5.11}$$

$$B_z(0,\theta) = \frac{\theta^2}{8} \sum_{i=1}^{N} \mu^2(i) \tag{5.12}$$

Taking the ratio of these distances, we obtain $B_y/B_x = \frac{1}{2}(\frac{2}{\pi})^2$ and $B_z/B_x = \frac{1}{2}$. These quantities are equal, respectively, to the A.R.E.'s of the Polarity Coincidence Correlator (P.C.C.) and the Correlator, with respect to the optimum detector[12]. 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

$$T_{opt} = \sum_{i=1}^{N} \mu(i) \cdot \{x_a(i) + x_b(i)\}^2 \tag{5.13}$$

$$T_{pcc} = \sum_{i=1}^{N} \mu(i) \cdot y_a(i) \cdot y_b(i) \tag{5.14}$$

$$T_c = \sum_{i=1}^{N} \mu(i) \cdot z(i) \tag{5.15}$$

ACKNOWLEDGMENT

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, under Contract DA-31-124-ARO-D-292.

REFERENCES

1. T. L. Grettenberg, "Signal Selection in Communication and Radar Systems", IEEE Trans. on Information Theory, Vol. IT-9, pp. 265-275; October 1963.
2. T. Kailath, "The Divergence and Bhattacharyya Distance Measures in Signal Selection", IEEE Trans. on 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. on 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. Royal Stat., Vol. 28, pp. 131-142; 1966.
5. S. Kullback, <u>Information Theory and Statistics</u>, New York,

Wiley; 1959.

6. K. Matsushita, "A Distance and Related Statistics in Multivariate Analysis", in P. R. Krishnaiah, ed., Multi-variate Analysis, pp. 187-200, Academic Press; 1966.
7. S. Karlin and R. N. Bradt, "On the Design and Comparison of Dichotomous Experiments", Ann. Math. Stat., Vol 27, pp. 390-409; 1956.
8. H. Kobayashi, "Distance Measures and Related Criteria", Chapter V of Representations of Complex-Valued Vector Processes and Their Application to Estimation and Detection, Ph.D. Thesis, Princeton University, N.J.; August 1967.
9. Ibid., Chapter I, "A Representation of the Complex Gaussian Distribution and Its Application to Complex Gaussian Processes".
10. M. G. Kendall and A Stuart, The Advanced Theory of Statistics, Vol. 2, Inference and Relationship, Hafner Publishing Co., New York.
11. J. Capon, "On the Asymptotic Efficiency of Locally Optimum Detectors", IRE Trans. on Information Theory, Vol. IT-7, pp. 67-71; April 1961.
12. J. W. Carlyle and J. B. Thomas, "On Nonparametric Signal Detectors", IEEE Trans. on Information Theory, Vol. IT-10, pp. 146-152; 1964.