$k=1,2,\cdots,m$, such that

$$\sum_{i=1}^{\infty} \Pr\left\{A_i^k|H_k\right\}=\infty,$$

$$\sum_{i=1}^{\infty} \Pr\left\{A_i^k|H_j\right\}<\infty, \qquad j=1,2,\cdots,m,\ j\neq k.$$

Let the observations be divided into blocks $B_1, B_2, \cdots$, where B_i itself consists of the blocks

$$B_i=S_i^{12}, S_i^{13}, \cdots, S_i^{lm}, S_i^{21}, S_i^{23}, \cdots, S_i^{kj}, \cdots, S_i^{m(m-1)},$$

$$k=1,2,\cdots,m,\ j=1,2,\cdots,m,\ j\neq k.$$

Each of the blocks S_i^{kj}, $j=1,2,\cdots,m$, $j\neq k$, tests for the event A_i^k. If A_i^k occurs during an S_i^{kj} block and the currently favored hypothesis is H_j, then the favored hypothesis becomes H_k. Otherwise the favored hypothesis does not change. The algorithm can be implemented with $m+1$ states by letting states $T_n \in \{1,2,\cdots,m\}$ denote the currently favored hypothesis, while $T_n=m+1$ denotes that during an S_i^{jk} block the currently favored hypothesis is H_j and the observations are matching with the event A_i^k. Thus if each sequence $\{A_i^k\}$ satisfies (5), then $m+1$ states can resolve the m-hypothesis test.

V. Remarks

Results on hypothesis testing with a finite memory have been extended to a class of problems with dependent observations. These extensions can be viewed as a set of sufficient conditions on the probability measure P defined on the observation space for which $m+1$ states can resolve the m-hypothesis test. It would be of interest to investigate problems for which the finite memory constraint precludes resolution of the true hypothesis. The proof of Theorem 2 given here requires discrete-valued random variables. It is not known if the result holds in general for the continuous case.

Consider also the following example. Let x_1, x_2 be a pair of random variables with joint probability measure P, and consider the hypothesis test H_0: $P=P_0$ versus H_1: $P=P_1$. Suppose the measures are such that $\Pr\{x_1+x_2=0|H_0\}=1$ and $\Pr\{x_1+x_2=1|H_1\}=1$. While the hypothesis test can be resolved without a memory constraint, it can be shown that a finite memory cannot resolve the test with probability one. The probability of error can be made arbitrarily small only when the number of memory states becomes arbitrarily large.

This raises the question of finding necessary conditions for a finite memory to resolve a hypothesis test. Another question of interest is to find the class of problems that can be resolved with a finite memory, but require more than $m+1$ states.

References

[1] T. M. Cover, "Hypothesis testing with finite statistics," *Ann. Math. Statist.*, vol. 40, pp. 828–835, 1968.

[2] J. Koplowitz, "Necessary and sufficient memory size for m-hypothesis testing," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 44–47, 1975.

[3] S. Yakowitz, "Multiple hypothesis testing by finite-memory algorithms," *Ann. Math. Statist.*, vol. 2, pp. 323–336, 1974.

[4] T. J. Wagner, "Estimation of the mean with a time-vary finite memory," *IEEE Trans. Inform. Theory*, vol. IT-18, 523–525, 1972.

[5] P. R. Hirschler, "Finite-memory algorithms for testing Bernoulli random variables," *Inform. Contr.*, vol. 24, pp. 11–19, 1974.

[6] T. M. Cover, "A note on the two-armed bandit problem with finite memory," *Inform. Contr.*, vol. 12, pp. 371–377, 1968.

[7] C. T. Mullis and R. A. Roberts, "Memory limitation and multistage decision processes." *IEEE Trans. on Sys.. Sci. and Cyber.*, vol. SSC-4 pp. 307–316, 1968.

[8] S. M. Samuels, "Randomized rules for the two-armed bandit problem with finite memory," *Ann. Math. Statist.*, vol. 39, pp. 2103–2107, 1968.

[9] R. W. Muise and R. R. Boorstyn, "Detection with time-varying finite memory receivers," Abstracts of Papers, *IEEE Int. Sym. on Information Theory*, pp. 35–36, 1972.

[10] P. F. Lynn and R. R. Boorstyn, "Bounds on finite memory detectors," Abstracts of Papers, *IEEE Int. Sym. on Information Theory*, 1972.

[11] M. E. Hellman and T. M. Cover, "Learning with finite memory," *Ann. Math. Statist.*, vol. 41, pp. 765–782, 1970.

[12] T. M. Cover and M. E. Hellman, "The two-armed bandit problem with time invariant finite memory," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 185–195, 1970.

[13] M. E. Hellman and T. M. Cover, "On memory saved by randomization," *Ann. Math. Statist.*, vol. 42, pp. 1075–1078, 1971.

[14] J. Koplowitz and R. A. Roberts, "Sequential estimation with a finite-statistic," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 631–635, 1973.

[15] R. A. Roberts and J. R. Tooley, "Estimation with finite memory," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 685–691, 1970.

[16] F. J. Samaniego, "Estimating a binomial parameter with finite memory," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 636–643, 1973.

[17] B. O. Shubert and C. M. Anderson, "Testing a simple symmetric hypothesis by a finite-memory deterministic algorithm." *IEEE Trans. on Information Theory*, vol. IT-19, pp. 644–647, 1973.

[18] F. S. Spitzer, *Principles of Random Walk*, pp. 317–320, Van Nostrand, Princeton, N.J., 1964.

[19] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I, Third Ed., 1968.

Single-Threshold Detection of a Random Signal in Noise with Multiple Independent Observations, Part 2: Continuous Case

PAUL R. PRUCNAL, MEMBER, IEEE, AND MALVIN CARL TEICH, SENIOR MEMBER, IEEE

***Abstract*—A single-threshold detector is derived for a wide class of classical binary decision problems involving the likelihood-ratio detection of a signal embedded in noise. The class of problems considered encompasses the case of multiple independent (but not necessarily identically distributed) observations of a nonnegative (or nonpositive) signal embedded in additive and independent noise, where the range of the signal and noise is continuous. It is shown that a comparison of the sum of the observations with a unique threshold comprises an optimum detector if a weak condition on the noise is satisfied independent of the signal. Examples of noise densities that satisfy and that violate this condition are presented. A sufficient condition on the likelihood ratio which implies that the sum of the observations is also a sufficient statistic is considered.**

I. Introduction

The likelihood-ratio detection of a signal embedded in noise is an important class of classical binary decision problems that has found widespread application in the synthesis and analysis of many types of systems [1], [2]. For complex signal and noise statistics, however, it is sometimes difficult or impossible to express the likelihood ratio in closed form. Even for simple signal and noise statistics, direct implementation of the likelihood ratio as an optimum detector may be difficult. Sometimes it is possible to reduce the likelihood ratio to a simpler but equivalent detector by using various ad hoc geometric arguments or lengthy algebraic manipulations.

It is the purpose of this correspondence to derive a remarkably simple detector that is optimum for a broad range of classical binary decision problems involving the likelihood-ratio detection of a signal embedded in noise. The class of problems we consider encompasses the case of N independent (but not necessarily identically distributed) observations of a nonnegative (or nonpositive) signal random variable embedded in an additive and independent noise random variable, where the range of the signal and noise is continuous. We show that a comparison of the sum of the N observations with a unique threshold comprises

Manuscript received April 10, 1978; revised July 28, 1978. This work was supported by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DAAG29-77-C-0019.

The authors are with the Columbia Radiation Laboratory, Department of Electrical Engineering and Computer Science, Columbia University, New York, NY 10027.

0018-9448/79/0300-0213$00.75

an optimum detector, provided that the logarithm of the noise probability density does not contain a point of inflection. This condition on the noise probability density is sufficient to imply our single-threshold detector and does not depend on the signal probability density. We show by example that in many cases it is not difficult to test the logarithm of the noise density for a point of inflection analytically. In more difficult cases, a graphical representation of the noise density with a logarithmic ordinate scale may be useful in revealing a point of inflection. Finally, we develop a restriction on the form of the likelihood ratio that renders the sum of the N observations a sufficient statistic. We note that our results relate particularly to diversity combining, for which a large literature exists [1], [9].

We have previously [2] derived a limited version of the discrete case for a single observation (the case $N=1$) of a nonnegative signal embedded in noise, where the logarithm of the noise density was concave downward. A more detailed treatment of single-threshold detection for the discrete case of N observations has been presented elsewhere since it differs substantially from the continuous case presented here [3].

We consider the following general classical binary detection problem. Each of two source outputs corresponds to a hypothesis H_0 or H_1. To decide which hypothesis is true, based on the Bayes or Neyman–Pearson criterion, the optimal processing of the observation vector $\boldsymbol{x}$ is the well-known likelihood-ratio test [1, p. 26]

$$\Lambda(\boldsymbol{x})=\frac{p(\boldsymbol{x}|H_1)}{p(\boldsymbol{x}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \tag{1}$$

where $\Lambda(\boldsymbol{x})$ represents the likelihood ratio, $p(\boldsymbol{x}|H_i)$ is the probability density of $\boldsymbol{x}$ given that H_i is true, and λ is a constant depending on the choice of decision criterion. The observation vector $\boldsymbol{x}=(x_1,\cdots,x_N)$ consists of N independent observations.

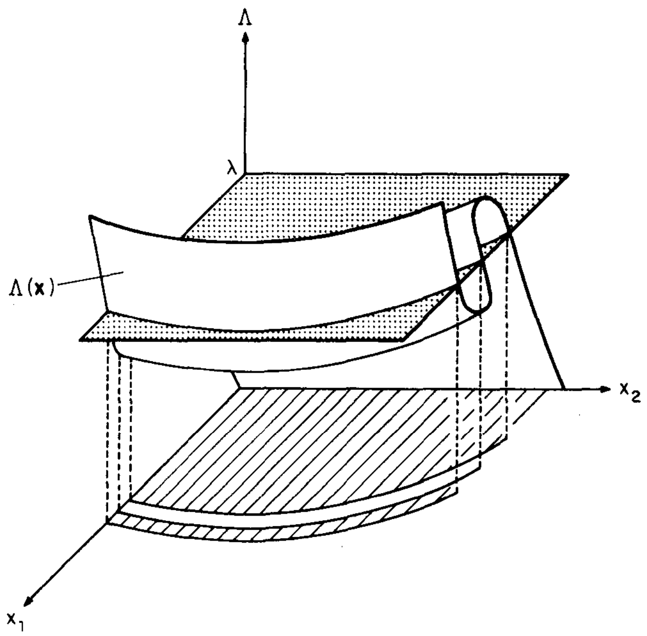
In the simplest situation $N=1$, corresponding to a single observation x_1. In this case $\Lambda(x_1)$ may be graphically represented by a curve in a two-dimensional Cartesian coordinate system. In Section II we derive a condition on the noise density which implies that $\Lambda(x_1)$ is monotonic with respect to x_1. The monotonicity of $\Lambda(x_1)$ implies, in turn, that (1) is equivalent to the *single-threshold* detector

$$x_1 \underset{H_0}{\overset{H_1}{\gtrless}} \lambda' \tag{2}$$

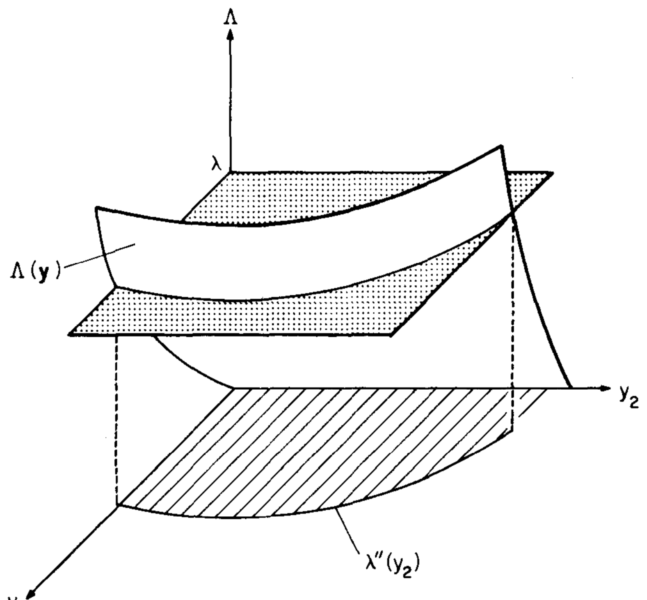
with threshold λ'. Equation (2) completely specifies the optimum processing of x_1.

For the case of multiple observations ($N>1$), we visualize $\Lambda(\boldsymbol{x})$ as an N-dimensional surface in $N+1$ space. An N-dimensional hyperplane, orthogonal to the Λ axis at λ, cuts through the surface $\Lambda(\boldsymbol{x})$. This is illustrated in Fig. 1(a) for $N=2$. Given an observation $\hat{\boldsymbol{x}}$, the test given by (1) is equivalent to determining whether $\Lambda(\hat{\boldsymbol{x}})$ is located above or below the hyperplane: if it is above, H_1 is chosen; if it is below, H_0 is chosen. The projections of the intersections of the hyperplane and $\Lambda(\boldsymbol{x})$ partition the remaining N coordinates into N-dimensional decision regions R_1 and R_0, corresponding to the regions where $\Lambda(\boldsymbol{x})$ is above the hyperplane (H_1 is chosen) and below the hyperplane (H_0 is chosen), respectively. The decision is then based upon the region in which the tip of the observation vector $\hat{\boldsymbol{x}}$ falls. In Fig. 1(a), R_0 is represented by the crosshatched region and R_1 by the unshaded region. If there are multiple intersections of the surface and the hyperplane, as in Fig. 1(a), then multiple boundaries divide the decision regions R_0 and R_1.

In Section II we prove that if the same condition on the noise density considered for $N=1$ applies to each component of the N-dimensional noise density, then $\Lambda(\boldsymbol{y})$ is monotonic with respect to $y_1=\sum_{i=1}^{N}x_i$, as illustrated in Fig. 1(b). (Here the likelihood ratio has now been transformed to the coordinate system $y_1,\cdots,y_N$.) This implies that the decision regions R_0 and R_1 are partitioned by a boundary λ'' which is a single-valued function



(a)



(b)

Fig. 1. (a) Likelihood ratio $\Lambda(\boldsymbol{x})$ versus the observations x_i for case $N=2$. Solution $\Lambda(\boldsymbol{x})=\lambda$ is represented by multiple curved intersections of $\Lambda(\boldsymbol{x})$ with dotted plane. Decision regions R_0 are crosshatched and represent coordinates (x_1,x_2) for which $\Lambda(\boldsymbol{x})<\lambda$. Decision regions R_1 are unshaded and represent coordinates (x_1,x_2) for which $\Lambda(\boldsymbol{x})\geqslant\lambda$. This case exhibits multiple curved decision boundaries. (b) Transformed likelihood ratio $\Lambda(\boldsymbol{y})$ for case $N=2$, where $\Lambda(\boldsymbol{y})$ is monotonic with respect to y_1. Solution $\Lambda(\boldsymbol{y})=\lambda$ is represented by single curved intersection of $\Lambda(\boldsymbol{y})$ with dotted plane. Region R_0 is crosshatched and represents coordinates (y_1,y_2) for which $\Lambda(\boldsymbol{y})<\lambda$. Decision region R_1 is unshaded and represents coordinates (y_1,y_2) for which $\Lambda(\boldsymbol{y})\geqslant\lambda$. This case exhibits a single-valued decision boundary $\lambda''(y_2)$ and therefore single-threshold detection.

of $y_2,\cdots,y_N$, as in Fig. 1(b). In this case, therefore, (1) is equivalent to the single-threshold detector

$$y_1 \underset{H_0}{\overset{H_1}{\gtrless}} \lambda''(y_2,\cdots,y_N)=\lambda''(\boldsymbol{x}) \tag{3}$$

where $\lambda''(\boldsymbol{x})$ is single valued. This single-threshold detector does not completely specify the optimal processing, as the single-threshold detector does in the case $N=1$, since λ'' is now a function of $\boldsymbol{x}$. However, (3) does assure the uniqueness of the

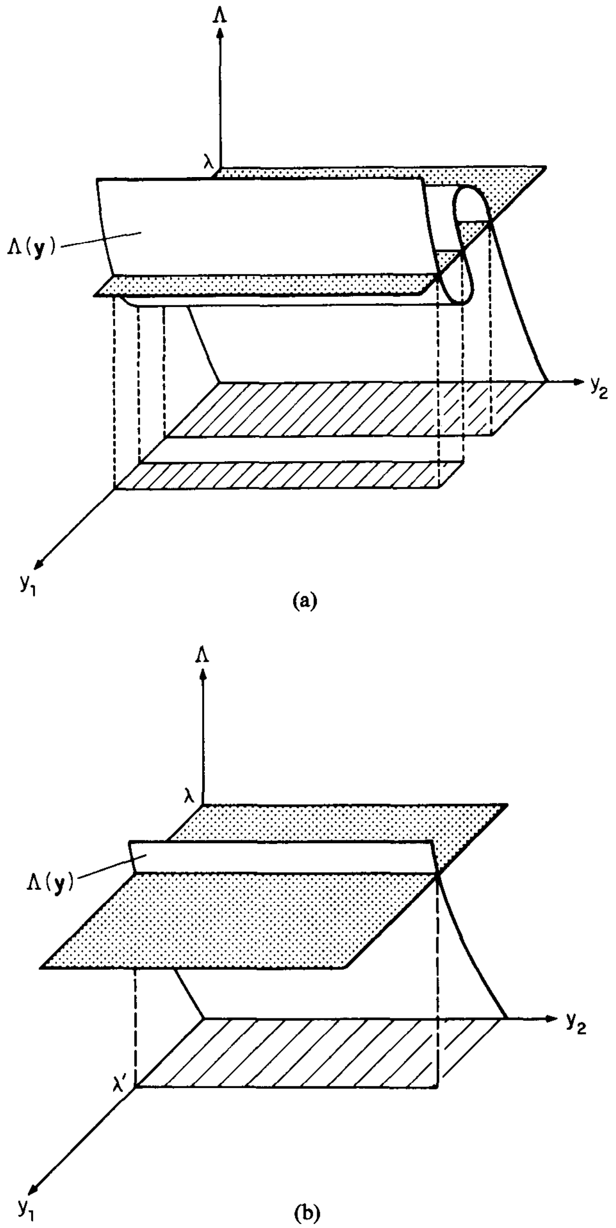


Fig. 2. (a) Transformed likelihood ratio $\Lambda(y)$ for case $N=2$, where $\Lambda(y)$ depends only on coordinate y_1. Solution $\Lambda(y)=\lambda$ is represented by multiple straight-line intersections of $\Lambda(y)$ with dotted plane. Decision regions R_0 are crosshatched and represent coordinates (y_1,y_2) for which $\Lambda(y)<\lambda$. The decision regions R_1 are unshaded and represent coordinates (y_1,y_2) for which $\Lambda(y)\geqslant\lambda$. This case exhibits multiple straight decision boundaries, so that y_1 is a sufficient statistic. (b) Transformed likelihood ratio $\Lambda(y)$ for case $N=2$, where $\Lambda(y)$ is monotonic with respect to y_1 and depends only on y_1. Solution $\Lambda(y)=\lambda$ is represented by single constant intersection of $\Lambda(y)$ with dotted plane. Decision region R_0 is crosshatched and represents coordinates (y_1,y_2) for which $\Lambda(y)<\lambda$. Decision region R_1 is unshaded and represents coordinates (y_1,y_2) for which $\Lambda(y)\geqslant\lambda$. This case exhibits a single constant boundary λ', so that single-threshold detection completely specifies optimal processing.

threshold, in contrast to the nonmonotonic case of Fig. 1(a). Note that if $N=1$, (3) reduces to (2). In Section III we examine a number of noise densities to determine whether single-threshold detection is optimal.

The transformed likelihood ratio $\Lambda(y)$ may depend explicitly only on the coordinate y_1, in which case the decision boundaries in three-dimensional space would be straight lines, as illustrated in Fig. 2(a). The quantity y_1 then contains all of the information necessary to make a decision and is therefore a sufficient statistic. In Section IV we develop a sufficient condition on $\Lambda(x)$ that renders $y_1=\sum_{i=1}^{N}x_i$ a sufficient statistic. If, in addition, the conditions discussed in Section II are satisfied, then R_0 and R_1 are partitioned by a single constant boundary λ', as illustrated in Fig. 2(b). In this case optimal detection is represented by the comparison

$$\sum_{i=1}^{N} x_i \underset{H_0}{\overset{H_1}{\gtrless}} \lambda', \tag{4}$$

which completely specifies the optimal processing.

The extension of the results presented here from two to M hypotheses does not appear to be straightforward.

II. Single-Threshold Detection for Continuous Distributions with N Observations

Let H_1 represent the presence of a signal with probability density $p_{Si}(s_i)$ embedded in noise with probability density $p_{Ni}(n_i)$, and let H_0 represent the absence of a signal (noise alone). The noise is within the continuous range $a\leqslant n_i\leqslant b$, and the signal is within the continuous range $c\leqslant s_i\leqslant d$. We assume that the signal and noise random variables are additive and independent. The probability density of $x_i=s_i+n_i$ under each hypothesis is then

$$p_i(x_i|H_1)=\int_{u_0}^{u_1} p_{Ni}(x_i-\xi_i)p_{Si}(\xi_i)\,d\xi_i \tag{5}$$

and

$$p_i(x_i|H_0)=p_{Ni}(x_i) \tag{6}$$

where $u_0=\max\,(x_i-b,c)$ and $u_1=\min\,(x_i-a,d)$. We further assume that the x_i are statistically independent, though not necessarily identically distributed, so that the likelihood-ratio test in (1) becomes

$$\Lambda(x)=\prod_{i=1}^{N}\Lambda_i(x_i) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \tag{7}$$

with

$$\Lambda_i(x_i)=p_i(x_i|H_1)/p_i(x_i|H_0). \tag{8}$$

Substituting (5) and (6) into (8), we obtain

$$\Lambda_i(x_i)=\left[\int_{u_0}^{u_1} p_{Ni}(x_i-\xi_i)p_{Si}(\xi_i)\,d\xi_i\right]\Big/p_{Ni}(x_i). \tag{9}$$

We now prove that if the noise distribution satisfies either the simple condition

$$d^2[\log p_{Ni}(n_i)]/dn_i^2\leqslant 0, \qquad \text{for all } n_i \text{ and all } i \tag{10}$$

or

$$d^2[\log p_{Ni}(n_i)]/dn_i^2\geqslant 0, \qquad \text{for all } n_i \text{ and all } i, \tag{11}$$

then the test

$$\sum_{i=1}^{N} x_i \underset{H_0}{\overset{H_1}{\gtrless}} \lambda''(x) \tag{12}$$

is optimal. Thus if the logarithm of the noise distribution does not contain a point of inflection, single-threshold detection is optimal.

Equation (10) implies that the function $\log p_{Ni}(n_i)$ is concave downward, or equivalently,

$$d[\log p_{Ni}(x_i-\xi_i)]/dx_i-d[\log p_{Ni}(x_i)]/dx_i \gtrless 0,$$

$$\cdot\text{for all } \xi_i \gtrless 0, \text{ all } x_i, \text{ and all } i \tag{13}$$

where the left side of the equation is $\geqslant 0$ if $\xi_i\geqslant 0$, $\leqslant 0$ if $\xi_i\leqslant 0$ and where $p_{Ni}(n_i)$ has been evaluated at x_i. Computing the derivatives in (13) and cross multiplying yields

$$p_{Ni}(x_i)\frac{dp_{Ni}(x_i-\xi_i)}{dx_i}-p_{Ni}(x_i-\xi_i)\frac{dp_{Ni}(x_i)}{dx_i}\gtrless 0,$$

$$\text{for all } \xi_i \gtrless 0, \text{ all } x_i, \text{ and all } i. \quad (14)$$

Equation (11) leads to an expression that is identical to (14) with one set of inequalities reversed (e.g., $\xi_i \lessgtr 0$).

Forming the derivative of the likelihood ratio in (9), we obtain

$$\frac{d\Lambda_i(x_i)}{dx_i}=\Bigg\{p_{Ni}(x_i)\Bigg[p_{Ni}(x_i-u_1)p_{Si}(u_1)\frac{du_1}{dx_i} - p_{Ni}(x_i-u_0)p_{Si}(u_0)\frac{du_0}{dx_i} + \int_{u_0}^{u_1}\frac{dp_{Ni}(x_i-\xi_i)}{dx_i}p_{Si}(\xi_i)\,d\xi_i\Bigg] - \frac{dp_{Ni}(x_i)}{dx_i}\Bigg[\int_{u_0}^{u_1}p_{Ni}(x_i-\xi_i)p_{Si}(\xi_i)\,d\xi_i\Bigg]\Bigg\}\Big/p_{Ni}^2(x_i). \quad (15)$$

Rearranging terms yields

$$\frac{d\Lambda_i(x_i)}{dx_i}=\Bigg\{-p_{Ni}(x_i)p_{Ni}(x_i-u_0)p_{Si}(u_0)\frac{du_0}{dx_i} + p_{Ni}(x_i)p_{Ni}(x_i-u_1)p_{Si}(u_1)\frac{du_1}{dx_i} + \int_{u_0}^{u_1}p_{Si}(\xi_i)\Bigg[p_{Ni}(x_i)\frac{dp_{Ni}(x_i-\xi_i)}{dx_i} - p_{Ni}(x_i-\xi_i)\frac{dp_{Ni}(x_i)}{dx_i}\Bigg]d\xi_i\Bigg\}\Big/p_{Ni}^2(x_i). \quad (16)$$

The first term on the right-hand side (RHS) of (16) disappears if $x_i<a$, $x_i>b$, or $x_i\leqslant b+c$ (so that $u_0=c$). Therefore, the first term on the RHS of (16) disappears for all x_i if $c\geqslant 0$ or $b=\infty$. The second term on the RHS of (16) disappears if $x_i<a$, $x_i>b$, or $x_i\geqslant a+d$ (so that $u_1=d$). Therefore, the second term on the RHS of (16) disappears for all x_i if $d\leqslant 0$ or $a=-\infty$. According to (14) the integral on the RHS of (16) is nonnegative if (10) is satisfied and $\xi_i\geqslant 0$ or if (11) is satisfied and $\xi_i\leqslant 0$ and is nonpositive if (10) is satisfied and $\xi_i\leqslant 0$ or if (11) is satisfied and $\xi_i\geqslant 0$. The restriction $\xi_i\geqslant 0$ requires that $u_0\geqslant 0$ (this is true if $c\geqslant 0$), and the restriction $\xi_i\leqslant 0$ requires that $u_1\leqslant 0$ (this is true if $d\leqslant 0$).

Combining the above requirements, we find that

$$\frac{d\Lambda_i(x_i)}{dx_i}\geqslant 0, \qquad \text{for all } x_i \text{ and all } i \quad (17)$$

if (10) is satisfied and the signal is nonnegative, or if (11) is satisfied, the signal is nonpositive, and the upper limit of the noise is infinity. If the upper limit of the noise $b\neq\infty$, (17) holds for all $x_i<b+c$ and all $x_i>b+d$. Similarly we find that

$$\frac{d\Lambda_i(x_i)}{dx_i}\leqslant 0, \qquad \text{for all } x_i \text{ and all } i \quad (18)$$

if (10) is satisfied and the signal is nonpositive, or if (11) is satisfied, the signal is nonnegative, and the lower limit of the noise is minus infinity. If the lower limit of the noise $a\neq-\infty$, (18) holds for all $x_i<a+c$ and all $x_i>a+d$. In accordance with the discussion preceding (2), (17) and (18) directly demonstrate that single-threshold detection is optimal for the case $N=1$ under the conditions specified above.

Let us now consider an N-dimensional coordinate system defined by the orthonormal basis $\boldsymbol{E}_1,\cdots,\boldsymbol{E}_N$, where $\boldsymbol{E}_1=N^{-1/2}(1,\cdots,1)$ and $\boldsymbol{E}_i=(e_{i1},\cdots,e_{iN})$. The basis vectors $\boldsymbol{E}_2,\cdots,\boldsymbol{E}_N$ are selected by an orthonormalization procedure [4, p. 165]. The proof is carried out for a general set of such basis vectors since the result is independent of the particular choice of the orthonormal basis. An arbitrary vector $\boldsymbol{y}$ in this coordinate system is a linear combination of the $\boldsymbol{E}_i$. The transformation of the x-coordinate system into the y-coordinate system is given by

$$\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x} \quad (19)$$

where

$$\boldsymbol{A}=\begin{bmatrix} N^{-1/2} & N^{-1/2} & \cdots & N^{-1/2} \\ e_{21} & e_{22} & \cdots & e_{2N} \\ \vdots & \vdots & & \vdots \\ e_{N1} & e_{N2} & \cdots & e_{NN} \end{bmatrix}. \quad (20)$$

In particular,

$$y_1=N^{-1/2}\sum_{i=1}^{N}x_i. \quad (21)$$

Since the rows of $\boldsymbol{A}$ form a basis, they are linearly independent, and $\boldsymbol{A}^{-1}$ exists. The orthonormality of the rows of $\boldsymbol{A}$ insures that $\boldsymbol{A}^{-1}=\boldsymbol{A}^T$, so that the inverse transformation is

$$\boldsymbol{x}(\boldsymbol{y})=\boldsymbol{A}^T\boldsymbol{y} \quad (22)$$

with components

$$x_i(\boldsymbol{y})=\sum_{k=1}^{N}e_{ki}y_k. \quad (23)$$

Using (7) for the likelihood ratio in conjunction with the chain rule for differentiation, we form the derivative

$$\frac{\partial\Lambda(\boldsymbol{x}(\boldsymbol{y}))}{\partial y_j}=\sum_{i=1}^{N}\Bigg[\frac{\partial x_i(\boldsymbol{y})}{\partial y_j}\frac{\partial\Lambda_i(x_i(\boldsymbol{y}))}{\partial x_i}\prod_{\substack{k=1\\k\neq i}}^{N}\Lambda_k(x_k(\boldsymbol{y}))\Bigg]. \quad (24)$$

Multiplying and dividing by $\Lambda(\boldsymbol{x})$, (24) becomes

$$\frac{\partial\Lambda(\boldsymbol{x}(\boldsymbol{y}))}{\partial y_j}=\Lambda(\boldsymbol{x})\sum_{i=1}^{N}\frac{\partial x_i(\boldsymbol{y})}{\partial y_j}\frac{\partial\Lambda_i(x_i(\boldsymbol{y}))/\partial x_i}{\Lambda_i(x_i(\boldsymbol{y}))}. \quad (25)$$

Using (23) and the orthogonality of the y_j,

$$\frac{\partial\Lambda(\boldsymbol{x})}{\partial y_j}=\Lambda(\boldsymbol{x})\sum_{i=1}^{N}e_{ji}\frac{\partial\Lambda_i(x_i(\boldsymbol{y}))/\partial x_i}{\Lambda_i(x_i(\boldsymbol{y}))}. \quad (26)$$

Setting $j=1$, substituting for y_1 using (21), and substituting $e_{1i}=N^{-1/2}$, we obtain

$$\frac{\partial\Lambda(\boldsymbol{x})}{\partial\left(\sum_{i=1}^{N}x_i\right)}=\frac{1}{N}\Lambda(\boldsymbol{x})\sum_{i=1}^{N}\frac{\partial\Lambda_i(x_i)/\partial x_i}{\Lambda_i(x_i)}. \quad (27)$$

From (17) each of the terms $\partial\Lambda_i(x_i)/\partial x_i$ in (27) is nonnegative for all x_i and all i; since N and $\Lambda_i(x_i)$ are also nonnegative, it follows that

$$\frac{\partial\Lambda(\boldsymbol{x})}{\partial\left(\sum_{i=1}^{N}x_i\right)}\geqslant 0, \qquad \text{for all } x_i. \quad (28)$$

If (11) and (18) are satisfied instead of (10) and (17), each of the terms $\partial\Lambda_i(x_i)/\partial x_i$ in (27) is nonpositive for all x_i and all i. Since N and $\Lambda_i(x_i)$ are nonnegative, it follows that in this case

$$\frac{\partial\Lambda(\boldsymbol{x})}{\partial\left(\sum_{i=1}^{N}x_i\right)}\leqslant 0, \qquad \text{for all } x_i. \quad (29)$$

Equations (28) and (29) indicate that $\Lambda(x)$ is either entirely monotonically nondecreasing or entirely monotonically nonincreasing, with respect to $\Sigma_{i=1}^{N} x_i$, so that the test

$$\sum_{i=1}^{N} x_i \underset{H_0}{\overset{H_1}{\gtrless}} \lambda''(x) \tag{30}$$

is optimal in accordance with the discussion preceding (3). Therefore, if (10) or (11) is satisfied and if the conditions stated in Section II are adhered to, single-threshold detection is optimal.

III. Discussion

In this section we consider the optimal processing of the observation vector x for a number of different noise densities. If $p_{Ni}(n_i)$ satisfies (10) or (11) and the conditions stated in Section II are satisfied, then single-threshold detection is optimal. As indicated in Section II, the N components of the noise density need not be identically distributed. Though single-threshold detection is optimal for most of the usual noise densities encountered, we also cite counterexamples for which our single-threshold detector does not necessarily apply. For convenience we use natural logarithms in (10) and (11), though logarithms to an arbitrary base may be used.

The *Gaussian* noise density [5, vol. I, p. 174; 6, p. 372] (with mean $\langle a\rangle$ and variance $\langle(\Delta a)^2\rangle$) is

$$p_{Ni}(n_i) = (2\pi\langle(\Delta a)^2\rangle)^{-1/2} \exp\left[-(n_i - \langle a\rangle)^2/2\langle(\Delta a)^2\rangle\right], \tag{31}$$

from which

$$d^2[\ln p_{Ni}(n_i)]/dn_i^2 = -\langle(\Delta a)^2\rangle^{-1} < 0, \qquad \text{for all } n_i. \tag{32}$$

Equation (32) satisfies (10), so single-threshold detection is optimal.

The *Rayleigh* noise density [7, p. 366] with unique mode $\langle(\Delta a)^2\rangle$ is

$$p_{Ni}(n_i) = \mathrm{u}(n_i) n_i\left[\exp\left(-n_i^2/2\langle(\Delta a)^2\rangle\right)\right]/\langle(\Delta a)^2\rangle \tag{33}$$

where the Heaviside unit step function $\mathrm{u}(n_i)$ is defined by

$$\mathrm{u}(n_i) = \begin{cases} 1, & \text{for } n_i \geqslant 0 \\ 0, & \text{for } n_i < 0. \end{cases} \tag{34}$$

From (33),

$$d^2[\ln p_{Ni}(n_i)]/dn_i^2 = -1/n_i^2 - 1/\langle(\Delta a)^2\rangle < 0, \qquad \text{for all } n_i \tag{35}$$

which satisfies (10), so single-threshold detection is optimal.

The *gamma* noise density [6, p. 124] with parameters $M>0$ and $\beta>0$ is

$$p_{Ni}(n_i) = \mathrm{u}(n_i)\beta^M n_i^{M-1} e^{-\beta n_i}/\Gamma(M). \tag{36}$$

This distribution (also called Erlangian and Pearson type III) can assume a variety of shapes for different values of M and β. The gamma density is the continuous analog of the negative binomial density [5, vol. II, p. 176]. Observe that if $M=1$, the gamma density reduces to the *exponential* density [6, p. 124], which is the continuous analog of the geometric (Bose–Einstein) density [5, vol. II, p. 8]. The exponential and Bose–Einstein densities share the property of lack of memory. Finally, observe that if $\beta=\frac{1}{2}$ and $M=k/2$, where k is a positive integer, the gamma density is the (central) *chi-square* density with k degrees of freedom [6, p. 124]. Using the gamma noise density given in (36), we obtain

$$d^2[\ln p_{Ni}(n_i)]/dn_i^2 = (1-M)/n_i^2, \qquad \text{for all } x_i. \tag{37}$$

For $0<M<1$, (37) obeys (11); for $M=1$, (37) obeys both (10) and (11); and for $M>1$, (37) obeys (10). Thus single-threshold detection is optimal for any value of M, including the gamma, exponential, and chi-square densities.

Using the same method it is easy to show that single-threshold detection is optimal for the *Maxwell* noise density. The *Rician* noise density, however, contains a modified Bessel function and does not permit easy calculation of $d^2[\ln p_{Ni}(n_i)]/dn_i^2$, so that our usual method is not convenient. An alternate method is to plot the Rician noise density with a logarithmic ordinate and to inspect it visually for a point of inflection. Of course, this must be done on a case-by-case basis. For the case of unit variance and unit specular component [7, eq. 5-184], such a plot has no point of inflection, indicating that single-threshold detection is optimal. In the limits of very small and very large specular components, the Rician density reduces to the Rayleigh and Gaussian densities, for which we have already shown that single-threshold detection is optimal.

The *beta* noise density [6, p. 124] with free parameters $\alpha>0$ and $\beta>0$ is

$$p_{Ni}(n_i) = [\mathrm{u}(n_1) - \mathrm{u}(n_i - 1)][\Gamma(\alpha+\beta)/\Gamma(\alpha)\Gamma(\beta)]n_i^{\alpha-1}(1-n_i)^{\beta-1} \tag{38}$$

from which

$$d^2[\ln p_{Ni}(n_i)]/dn_i^2 = \left[(1-\alpha)/n_i^2\right] + \left[(1-\beta)/(1-n_i)^2\right], \qquad \text{for all } n_i. \tag{39}$$

For different choices of α and β, the beta density can assume a wide variety of shapes. If $\alpha \leqslant 1$ and $\beta \leqslant 1$, the graph of $p_{Ni}(n_i)$ is U-shaped, approaching ∞ at the extremes, and (39) satisfies (11), so that single-threshold detection is optimal. If $\alpha \geqslant 1$ and $\beta \geqslant 1$, the graph of $p_{Ni}(n_i)$ is bell-shaped, and (39) satisfies (10), so that single-threshold detection is again optimal. However, if $\alpha<1$ and $\beta>1$, or $\alpha>1$ and $\beta<1$, the inflection points $(1 \pm \sqrt{-\gamma})/(1+\gamma)$ with $\gamma=(1-\beta)/(1-\alpha)$ are real, so that our single-threshold detector does not necessarily apply. (For example, assume that the N observations are identically distributed, that $p_{Si}(s_i) = \delta(s_i - \frac{1}{2})$, where δ is the Dirac delta, and that $p_{Ni}(n_i)$ is the beta density with $\alpha=2$ and $\beta=\frac{1}{2}$. Using (7) and (9), we find by direct calculation that

$$\Lambda(x) = \prod_{i=1}^{N}\left(x_i - \frac{1}{2}\right)x_i^{-1}(1-x_i)^{1/2}\left(1 - x_i + \frac{1}{2}\right)^{-1/2} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

specifies the optimal processing. A graphical representation shows that $\Lambda_i(x_i)$ is nonmonotonic, so that single-threshold detection is in fact not optimal.) Observe that in both the U-shaped and bell-shaped case, the curve flattens as α and β approach unity. In the limit where $\alpha=1$ and $\beta=1$, the beta density reduces to the *uniform* density on the interval (0, 1).

The *Cauchy* noise density [6, pp. 121, 132] with scale parameter σ and location parameter μ is

$$p_{Ni}(n_i) = \sigma/\pi\left[\sigma^2 + (n_i - \mu)^2\right], \tag{40}$$

from which

$$d^2[\ln p_{Ni}(n_i)]/dn_i^2 = 2\left[(n_i-\mu)^2 - \sigma^2\right]/\left[(n_i-\mu)^2 + \sigma^2\right]^2. \tag{41}$$

The Cauchy density resembles the Gaussian density in shape, but its tails approach the axis so slowly that a mean does not exist. In contrast to the Gaussian density, we see from (41) that the logarithm of the Cauchy density has inflection points at $\mu \pm \sigma$, so that our single-threshold detector does not necessarily apply. (For example, assume that the N observations are identically distributed, that $p_{Si}(s_i) = \delta(s_i - 2)$, and that $p_{Ni}(n_i)$ is the Cauchy density with $\mu=0$ and $\sigma=1$. Using (7) and (9), we find by direct calculation that

$$\Lambda(x) = \prod_{i=1}^{N}(1+x_i^2)\left[1+(x_i-2)^2\right]^{-1} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

specifies the optimal processing. Calculation of $d\Lambda_i(x_i)/dx_i$ indicates that $\Lambda_i(x_i)$ has a peak and is nonmonotonic, so that single-threshold detection is indeed not optimal.)

In both cases where the logarithm of the noise density contained a point of inflection (i.e., Cauchy, and beta with $\alpha>1$, $\beta<1$ or $\alpha<1$, $\beta>1$), we have shown by a specific example that single-threshold detection is indeed not optimal. This suggests that the sufficient conditions given in (10) and (11) may also be necessary. It is easy to show that (10) and (11) are in fact necessary conditions for any class of signal distributions that includes $p_{S_i}(s_i)=\delta(s_i-1)$.

For noise densities that cannot be expressed in closed form, it may be impossible to test for a point of inflection analytically. This would be the case, for example, if the noise density were expressed as a sum. In that case it may be possible to inspect a plot of the noise density with logarithmic ordinate for a point of inflection.

Finally, we note that our result is more general than it may first appear because convolutions of distributions that are IFR (increasing failure rate) remain IFR [10].

IV. Sufficient Statistic

It is apparent that our method is most powerful for the single observation case ($N=1$). Here we have shown that the optimal receiver structure is completely specified by the single-threshold detector (see (2)), if the logarithm of the noise does not contain an inflection point. For example, if the Neyman–Pearson criterion [1, p. 33] is used, then λ' is the unique fixed solution to

$$P_F=\int_{\lambda'}^{\infty} p_{N1}(x_1)\,dx_1 \leqslant \alpha \tag{42}$$

where the false-alarm rate P_F is constrained to be less than the constant α. Since λ' is a fixed solution to (42), (2) completely specifies the optimal processing.

For $N>1$, however, λ' is not necessarily fixed, so that (3) does not completely specify the optimal processing. Equation (42) becomes

$$P_F=\int_{h}^{\infty} p_N(y)\,dy \leqslant \alpha \tag{43}$$

where $h=\{y: y_1 \geqslant \lambda''(x(y))\}$ and λ'' is now dependent on the observation x. Therefore, although λ'' is a unique (single-threshold) solution to (43), it is not necessarily fixed for different values of x, so that y_1 is not necessarily a sufficient statistic. (In practice, however, the threshold might be set at some average level, ignoring the detailed dependence of λ'' on x.) As an example, we consider the exponential noise density with parameter β, which was shown to satisfy the condition for single-threshold detection (see (37)). Embedded in this noise is an exponential signal density with parameter $\alpha\neq\beta$. After calculating and simplifying $\Lambda(x)$, it is easy to see that $\sum_{i=1}^N x_i$ is not a sufficient statistic. Therefore $\sum_{i=1}^N x_i$ alone does not contain all of the information necessary to make a decision, even though single-threshold detection is optimal.

We may, however, obtain a rather restrictive sufficient condition on $\Lambda(x)$ for which $\sum_{i=1}^N x_i$ is a sufficient statistic. We begin by choosing the basis vectors $E_2,\cdots,E_N$ in (20) using the Helmert orthonormalization procedure [8, p. 563]. The components of E_i, for $i>1$, are then

$$e_{ij}=\begin{cases}[i(i-1)]^{-1/2}, & \text{for } j<i\\ -(i-1)[i(i-1)]^{-1/2}, & \text{for } j=i\\ 0, & \text{for } j>i.\end{cases} \tag{44}$$

We observe that

$$\sum_{j=1}^{N} e_{ij}=0, \qquad \text{for all } i>1. \tag{45}$$

Now assume that the $\Lambda_i(x_i)$ are *identically distributed* with the form

$$\Lambda_i(x_i)=ae^{bx_i}, \qquad \text{for all } i \tag{46}$$

where a and b are real. Taking the derivative, we find

$$\partial\Lambda_i(x_i)/\partial x_i=b\Lambda_i(x_i), \qquad \text{for all } i, \tag{47}$$

and substituting in (26),

$$\partial\Lambda(x)/\partial y_j=\Lambda(x)b\sum_{i=1}^{N} e_{ji}. \tag{48}$$

Using (45), it is apparent that

$$\partial\Lambda(x)/\partial y_j=0, \qquad \text{for all } j>1. \tag{49}$$

Therefore, if $\Lambda_i(x_i)$ has the form given by (46), $\Lambda(x)$ is independent of y_j, for all $j>1$, so that $y_1=\sum_{i=1}^N x_i$ is a sufficient statistic in accordance with the discussion of Fig. 2(a).

Equation (46) is satisfied, for example, by a constant signal $\langle a\rangle$ embedded in additive independent exponential noise, implying that $\sum_{i=1}^N x_i$ is a sufficient statistic. Equation (46) is also satisfied by a constant signal $\langle a\rangle$ embedded in additive independent Gaussian noise, so that $p_i(x_i|H_0)=\mathrm{N}(0,\langle(\Delta a)^2\rangle)$ and $p(x_i|H_1)=\mathrm{N}(\langle a\rangle,\langle(\Delta a)^2\rangle)$ [1, p. 27], and $\sum_{i=1}^N x_i$ is a sufficient statistic where N represents the normal density.

If both (46) and the condition on the noise in (10) or (11) is satisfied, then processing is completely specified by (4), again provided that the conditions of additivity, independence, and positivity (or negativity) of the signal are obeyed. This is indeed the case for the above examples, which may therefore be represented by Fig. 2(b).

Acknowledgment

It is a pleasure to thank G. Lachs and D. R. Smith for valuable suggestions.

References

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part 1.* New York: Wiley, 1968.
[2] M. C. Teich, P. R. Prucnal, and G. Vannucci, "Optimum photon detection with a simple counting processor," *Opt. Lett.*, vol. 1, pp. 208–210, Dec. 1977.
[3] P. R. Prucnal and M. C. Teich, "Single-threshold detection of a random signal in noise with multiple independent observations. 1: Discrete case with application to optical communications," *Appl. Opt.*, vol. 17, pp. 3576–3583, Nov. 15, 1978.
[4] D. C. Murdoch, *Linear Algebra.* New York: Wiley, 1970.
[5] W. Feller, *An Introduction to Probability Theory and Its Applications.* New York: Wiley, vol. I, 3rd ed., 1968; vol. II, 2nd ed., 1971.
[6] M. Woodroofe, *Probability with Applications.* New York: McGraw-Hill, 1975.
[7] M. Schwartz, *Information Transmission, Modulation, and Noise.* New York: McGraw-Hill, 2nd ed., 1970.
[8] S. Zacks, *The Theory of Statistical Inference.* New York: McGraw-Hill, 1971.
[9] R. S. Kennedy, *Fading Dispersive Communication Channels.* New York: Wiley, 1969.
[10] R. E. Barlow and F. Proschan, *Mathematical Theory of Reliability.* New York: Wiley, 1965.

Moments of the Truncated Noncentral Chi-Squared Distribution

ROGER J. WEBSTER

***Abstract*—The truncated form of the noncentral chi-squared distribution arises in connection with dynamic range-limited spectra. A method of evaluating the moments of this distribution is presented.**

Manuscript received October 18, 1977; revised June 29, 1978.
The author is with Computing Devices Company, MS 1-10D, P.O. Box 8508, Ottawa, ON, K1G 3M9, Canada.

0018-9448/79/0300-0218$00.75