Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (2): 201-206, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.201.206

Two-stage Multiple Hypotheses LAO Test of Distributed Detection System for Many Families of Distributions

Farshin Hormozi Nejad
Department of Statistics and Mathematics, Islamic Azad University, Ahvaz Branch, Iran

Abstract: Problem of multiple hypotheses of distributed detection forming many families of possible probability distributions is considered. Reliabilities matrix and Logarithmically Asymptotically Optimal (LAO) hypotheses testing for a pair of stages are studied. The fusion center detects one of some known probability distributions of the source inside many disjoint families of probability distributions. In first stage the first fusion center identifies one family of probability distributions on the base of a messages sample, then in second stage the second fusion center denotes exact distribution in the mentioned family on the base of another messages sample and the result of the first fusion center. The results show the compatibility conditions to happen of characteristics of LAO hypotheses testing of distributed detection.

Key words: Distributed detection, information theory, logarithmically asymptotically optimal test, two-stage test, reliability

INTRODUCTION

There is a considerable literature on the problems of distributed detection and decision in engineering contexts such as Kreidl *et al*. (2011), Tsitsiklis (1988) and Tsitsiklis and Athans (1985). The problem is important because the components of a distributed detection system may amass more data than they can transmit to a fusion node and must summarize that data by choice of a message drawn from a small set. The decentralized or distributed detection problem was first formulated and studied by Tenney and Sandell (1981) which considers a "parallel configuration" whereby each sensor makes an observation and sends a quantized version of that observation to a fusion center. The goal is to make a decision on the possible hypotheses, based on the messages received at the fusion center.

The most common architecture in distributed detection is the parallel system depicted in Fig. 1. It consists of N geographically dispersed sensors, one-way communication links and a fusion center. Each sensor makes an observation denoted by $X_i, i=\overline{1,N}$ of a random source, quantizes X_i into an M-ary message $U_i = g_i(X_i)$ and then transmits $U_i, i=\overline{1,N}$ to the fusion center. Upon receipt of $U_1, U_1, \ldots, U_N$, the fusion center makes a global decision $U_0 = D(U_1, U_2, \ldots, U_N)$ about the nature of the random source.

The optimal design of system, entails choosing quantizers $g_1, g_2, \ldots, g_N$ and a global decision rule D so as to optimize the reliabilities. The messages $U_1, U_2, \ldots, U_N$ are all transmitted to the fusion center which declares hypothesis $H_i, i=\overline{1,S}$ to be true, applying a decision rule D.

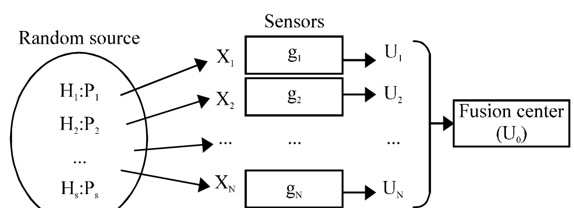

Fig. 1: Multiple hypotheses parallel distributed detection system, H: Hypothesis, P: Probability distribution, X: Observation, g: Quantizer, U: Message

As was shown by Tsitsiklis (1988), when N tends to infinity, the error exponents of the absolutely optimal system coincide with those achieved by the best identical quantizer system. Now, in an N-sensor identical quantizer system, the quantizer outputs U = g(X) are clearly i.i.d. The optimal error exponents are then obtained by choosing the mapping g so as to maximize the appropriate functional such as reliabilities matrix. In case of two hypotheses both reliabilities corresponding to two possible error probabilities could not be increased simultaneously, it is an accepted way to fix the value of one of the reliabilities and try to make the tests sequence get the greatest value of the remaining reliability.

The need of testing of more than two hypotheses in many scientific and applied fields has essentially increased recently. The models of multiple hypotheses optimal testing are studied in some direct such as Ahlswede and Haroutunian (2006), Hoeffding (1965), Haroutunian (1990) and Tusnady (1977). The models of the two-stage LAO testing in multiple hypotheses for a pair of families and many families of Probability Distributions (PDs) are investigated by Hormozi Nejad and Haroutunian (2012a) and Hormozi Nejad *et al.* (2011) and the model of two-stage LAO test of distributed detection for a pair of family of PDs is investigated by Hormozi Nejad and Haroutunian (2012b). In this paper the problem of distributed detection of two-stage multiple hypotheses LAO testing to detect between hypotheses consisting of many families of PDs is studied. The matrices of optimal asymptotic interdependencies of all pairs of the error probability exponents are studied.

Some preliminaries are in coming each sensor observation x takes values in the set X. A deterministic M-ary quantizer is a measurable mapping g from the observation space X to the message space U = {1,2,...,M}.

Random Variable (RV) X characterizing the studied object takes values in the set X and P(X) is the space of all distributions on X. The random source has S hypothetical PDs that are divided in K disjoint families of PDs. The first family includes R_1 PDs $P_1, P_2,..., P_{R_1}$, the second family consists of R_2 PDs $P_{R_1+1}, P_{R_1+2},..., P_{R_1+R_2}$ and etc., the K-th family have R_K PDs. The distributions of X under hypotheses $H_i,\ i=\overline{1,S}$ are denoted by $P_i,\ i=\overline{1,S}$. The distribution of the messages produced by g are denoted by $P_{i(g)}$ and it is obtainable from P_i and g.

Let N-sample $x = (x_1, x_2,..., x_N)$ be a vector of results of N independent observations of the RV X and $u = (u_1, u_2,..., u_N)$ be a vector of results of N transmitted messages to the fusion center. The purpose of the test is using sample u to detect the actual distribution from given list. The divergence (Kullback-Leibler distance) of PDs P and Q, is defined by Cover and Thomas (1991) and Haroutunian *et al.* (2008) as follows:

$$D(Q \| P) \triangleq \sum_{u \in U} Q(u) \log \frac{Q(u)}{P(u)}$$

ONE-STAGE MULTIPLE HYPOTHESES LAO TEST OF DISTRIBUTED DETECTION

The procedure of making decision on the base of N-sample is called by the test ϕ^N. The statistician must detect one among S hypotheses. An answer must be defined using vector of results of N-sample $u = (u_1, u_2,..., u_N)$.

The probabilities of the erroneous acceptance of hypothesis H_l provided that H_s is true, are defined as follows:

$$\alpha_{l|s}\left(\phi^N\right) \triangleq P_s^N\left(U_0 = l\right),\ l,s=\overline{1,S},\ l \neq s \qquad (1)$$

If the hypothesis H_s is true but it is not accepted, then the probability of error is:

$$\alpha_{s|s}\left(\phi^N\right) \triangleq P_s^N\left(U_0 \neq s\right) = \sum_{l \neq s} \alpha_{l|s}\left(\phi^N\right),\ s=\overline{1,S} \qquad (2)$$

Corresponding "reliabilities", are defined as follows for infinite sequence of tests ϕ:

$$E_{l|s}(\phi) \triangleq \limsup_{N\to\infty}\left\{-\frac{1}{N}\log \alpha_{l|s}\left(\phi^N\right)\right\},\ l,s=\overline{1,S} \qquad (3)$$

It follows from Eq. 1-3 that for every test sequence ϕ:

$$E_{s|s}(\phi) = \min_{l \neq s} E_{l|s}(\phi),\ s=\overline{1,S}$$

The following theorem contains the solution of problem of LAO test ϕ^* construction and existence conditions of the test of elements of matrix $E(\phi^*)$ of which are positive. For construction of the necessary LAO test for preliminarily given positive values $E_{1|1}, E_{2|2},..., E_{S-|S-1}$, the following subsets of distributions are defined:

$$R_s \triangleq \left\{Q : D\left(Q \| P_{s(g)}\right) \leq E_{s|s}\right\},\ s=\overline{1,S-1}$$

$$R_S \triangleq \left\{Q : D\left(Q \| P_{s(g)}\right) > E_{s|s},\ s=\overline{1,S-1}\right\}$$

$$E_{s|s}^* \triangleq E_{s|s},\ s=\overline{1,S-1} \qquad (4)$$

$$E_{l|s}^* \triangleq \inf_{Q:\ D(Q\|P_{l(g)}) \leq E_{l|l}^*} D\left(Q \| P_{s(g)}\right),\ l,s=\overline{1,S},\ l \neq s \qquad (5)$$

$$E_{S|S}^* \triangleq \min_{l \neq S} E_{l|S}^* \qquad (6)$$

Theorem 1: Hormozi Nejad and Haroutunian (2012b): If all distributions $P_s = s = \overline{1,S}$ are different and the positive values $E_{1|1}, E_{2|2},..., E_{S-|S-1}$ are such that the following inequalities hold:

$$E_{1|1} < \min_{l=\overline{2,S}} D\left(P_{l(g)} \| P_{1(g)}\right) \qquad (7)$$

$$E_{s|s} < \min\left[\min_{l=\overline{1,s-1}} E_{l|s}^*,\ \min_{l=\overline{s+1,S}} D\left(P_{l(g)} \| P_{s(g)}\right)\right],\ s=\overline{2,S-1} \qquad (8)$$

then there exists a LAO sequence of tests, all elements of the reliabilities matrix $E^* = \{E^*_{l|s}\}$ of which are positive and are defined in Eq. 4-6.

When one of the inequalities Eq. 7-8 is violated, then at least one element of the matrix E^* is equal to zero.

TWO-STAGE MULTIPLE HYPOTHESES LAO TEST OF DISTRIBUTED DETECTION

Now another version of testing will be discussed by supposing $N = N_1+N_2$ such that:

$$N_1 = \lceil\psi\rceil,\ N_2 = [(1-\psi)N],\quad 0<\psi<1$$

and so vectors of messages are as follows:

$$u = (u_1, u_2),\qquad u \in u^N,\qquad u^N = u^{N_1}\times u^{N_2}$$

The two-stage test on the base of N-sample denoted by $\Phi^N = (\varphi_1^{N_1},\varphi_2^{N_2})$ is the system depicted in Fig. 2. The first stage is to choice of a family of PDs, it is executed by a non-randomized test $\varphi_1^{N_1}(u)$ using the first messages sample u_1. The next stage of test is a non-randomized test $\varphi_2^{N_2}(u_2,U')$ based on another messages sample u_2 and the outcome of test $\varphi_1^{N_1}(u_1)$ that is the first fusion center U'.

First stage of two-stage test of distributed detection: The first stage of decision making consists of using the first messages sample u_1 for selection of one family of Pds and it is shown by a test $\varphi_1^{N_1}(u_1)$.

Consider for convenience the cumulative numbers:

$$C_k = \sum_{i=1}^{k} R_i$$

and the sets of indexes:

$$D_1 = \{\overline{1,C_1}\},\ D_2 = \{\overline{C_1+1,C_2}\},\ldots,\ D_K = \{\overline{C_{K-1}+1,S}\}$$

Therefore, suppose there are K disjoint families of PDs $P_1, P_2,\ldots, P_K$ such that:

$$P_k = \{P_s,\ s\in D_k\},\ k=\overline{1,K}$$

Let $\alpha'_{m|k}(\varphi_1^{N_1}),\ m,k=\overline{1,K},\ m\neq k,$ be the probability of the erroneous acceptance of the m-th family of PDs provided that the k-th family of PDs is true (that is the correct PD is in the k-th family):

$$\alpha'_{m|k}\left(\varphi_1^{N_1}\right) = \max_{s\in D_k} P_s^{N_1}\left(U'=m\right),\ m,k=\overline{1,K},\ m\neq k$$

The reliabilities of the infinite sequence of tests φ_1 are considered as follows:

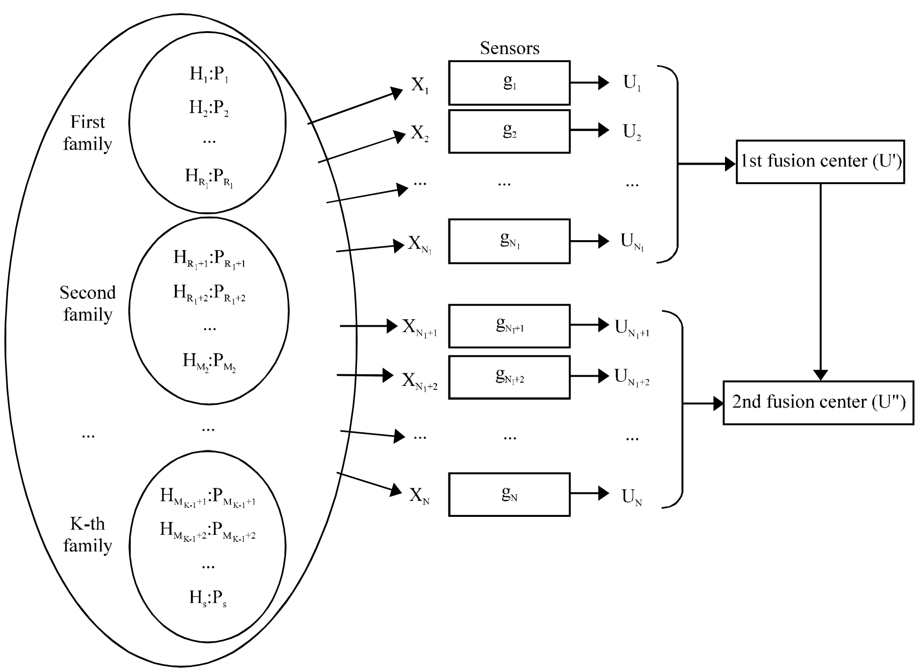

Fig. 2: Two-stage multiple hypotheses test of distributed detection system, H: Hypothesis, P: Probability distribution, X: Observation, g: Quantizer, U: Message, U', U": Decision rules

$$E'_{m|k}(\varphi_1) \triangleq \limsup_{N_1\to\infty}\left\{-\frac{1}{N_1}\log\alpha'_{m|k}(\varphi_1^{N_1})\right\},\ m,k=\overline{1,K}$$

For preliminarily given positive values $E'_{1|1}$, $E'_{2|2}$,..., $E'_{S-|S-1}$, the following subsets of distributions are defined:

$$R'_k \triangleq \left\{Q:\min_{s\in D_k} D(Q\,||\,P_{s(g)}) \le E'_{k|k}\right\},\ k=\overline{1,K-1}$$

$$R'_K \triangleq \left\{Q:\min_{s\in D_k} D(Q\,||\,P_{s(g)}) > E'_{k|k},\ k=\overline{1,K-1}\right\}$$

$$E'^{*}_{k|k} \triangleq E'_{k|k},\ k=\overline{1,K-1} \quad (9)$$

$$E'^{*}_{m|k} \triangleq \min_{s\in D_k}\ \inf_{Q:\ \min_{l\in D_m} D(Q||P_{l(g)})\le E'^{*}_{m|m}} D\left(Q\,||\,P_{s(g)}\right),\ m,k=\overline{1,K},\ m\ne k \quad (10)$$

$$E'^{*}_{K|K} \triangleq \min_{l\ne K} E'^{*}_{l|K} \quad (11)$$

Theorem 2: Hormozi Nejad and Haroutunian (2012a): If all distributions P_s, $s=\overline{1,S}$ are different and the positive values $E'_{1|1}$, $E'_{2|2}$,..., $E'_{S-|S-1}$ are such that the following inequalities hold:

$$E'_{1|1} < \min_{l\notin D_1,\ s\in D_1} D\left(P_{l(g)}\,||\,P_{s(g)}\right) \quad (12)$$

$$E'_{k|k} < \min\left[\min_{m=1,k-1} E'^{*}_{m|k},\ \min_{m=k+1,K,\ l\in D_m,\ s\in D_k} D\left(P_{l(g)}\,||\,P_{s(g)}\right)\right],\ k=\overline{2,K-1} \quad (13)$$

then there exists a LAO sequence of tests, all elements of the reliabilities matrix $E'^{*} = \{E'^{*}_{m|k}\}$ of which are positive and are defined in Eq. 9-11.

When one of the inequalities Eq. 12-13 is violated, then at least one element of the matrix E'^{*} is equal to zero.

Second stage of the two-stage test of distributed detection: The test $\varphi_2^{N_2}(u_2,U')$ can be defined by using result of the first fusion center U' and the second messages sample u_2.

The probability of the fallacious acceptance of PD P_l at the second stage of test, when P_s is correct and k-th family of PDs is accepted, is:

$$\alpha''_{l|s}\left(\varphi_2^{N_2}\right) \triangleq P_s^{N_2}\left(U''=l\,|\,U'=k\right),\ l\ne s,\ l\in D_k,\ k=\overline{1,K}$$

The probability to reject P_s when it is true and k-th family of PDs is accepted, is:

$$\alpha''_{s|s}\left(\varphi_2^{N_2}\right) \triangleq P_s^{N_2}\left(U''\ne s\,|\,U'=k\right) = \sum_{l\ne s}\alpha''_{l|s}\left(\varphi_2^{N_2}\right),\ l\in D_k,\ k=\overline{1,K} \quad (14)$$

Corresponding reliabilities for the second stage of test are:

$$E''_{l|s}(\varphi_2) \triangleq \limsup_{N_2\to\infty}\left\{-\frac{1}{N_2}\log\alpha''_{l|s}(\varphi_2^{N_2})\right\},\ l,s=\overline{1,S} \quad (15)$$

It follows from Eq. 14 and 15:

$$E''_{s|s}\left(\varphi_2\right) \triangleq \min_{l\ne s} E''_{l|s}\left(\varphi_2\right),\ l,s=\overline{1,S}$$

Theorem 3: Haroutunian (1990) and Hormozi Nejad and Haroutunian (2012b): If at the first stage of test the k-th family of PDs is accepted, then for given positive and finite values $E''_{s|s}$, S∈D_k, s≠C_k of the reliabilities matrix, let us investigate the regions:

$$R''_s \triangleq \left\{Q:D\left(Q\,||\,P_{s(g)}\right)\le E''_{s|s}\right\},\ s\in D_k,\ s\ne C_k$$

$$R''_{C_k} \triangleq \left\{Q:D\left(Q\,||\,P_{s(g)}\right) > E''_{s|s},\ s\in D_k,\ s\ne C_k\right\}$$

and the following values of elements of the future reliabilities matrix $E''(\varphi_2^{*})$ of the LAO test sequence:

$$E''^{*}_{s|s} \triangleq E''_{s|s},\ s\in D_k,\ s\ne C_k$$
$$E''^{*}_{l|s} \triangleq \inf_{Q\in R''_l} D\left(Q\,||\,P_{s(g)}\right),\ l,s\in D_k,\ l\ne s,$$
$$E''^{*}_{C_k|C_k} \triangleq \min_{l\ne C_k} E''^{*}_{l|C_k}$$

When the following compatibility conditions are valid:

$$E''_{C_{k-1}+1|C_{k-1}+1} < \min_{l=2,R_k} D\left(P_{(C_{k-1}+s)(g)}\,||\,P_{(C_{k-1}+1)(g)}\right),$$
$$E''_{C_{k-1}+s|C_{k-1}+s} < \min\left[\min_{l=1,s-1} E''^{*}_{C_{k-1}+l|C_{k-1}+s},\ \min_{l=s+1,R_k} D\left(P_{(C_{k-1}+l)(g)}\,||\,P_{(C_{k-1}+s)(g)}\right)\right],$$
$$s=\overline{2,R_k-1}$$

then there exists a LAO sequence of test $\varphi_2$9, elements of reliabilities matrix $E''(\varphi_2^{*})$ of which are defined above and are positive.

Even if one of the compatibility conditions is violated, then $E''(\varphi_2^{*})$ has at least one element equal to zero.

Reliabilities of the two-stage test of distributed detection: The tool of making decision according to N-sample denoted $\Phi^{*N} = (\varphi_1^{*N_1},\varphi_2^{*N_2})$ is organized by a pair of LAO tests $\varphi_1^{*N_1}$ and $\varphi_{21}^{*N_1}$. Similarly, definitions of error probabilities and reliabilities of two-stage test are as follows:

$$E''^{*}_{s|s} \triangleq E''_{s|s},\ s\in D_k,\ s\ne C_k$$
$$E''^{*}_{l|s} \triangleq \inf_{Q\in R''_l} D\left(Q\,||\,P_{s(g)}\right),\ l,s\in D_k,\ l\ne s,$$
$$E''^{*}_{C_k|C_k} \triangleq \min_{l\ne C_k} E''^{*}_{l|C_k}$$

So error probabilities can be considered as follows:

$$\alpha'''_{l|s}\left(F^{*N}\right) \triangleq P_s^N\left(U''=l\right),\ l,s=\overline{1,S},\ l\neq s,$$
$$\alpha'''_{s|s}\left(F^{*N}\right) \triangleq P_s^N\left(U''\neq s\right)=\sum_{l\neq s}\alpha'''_{l|s}\left(F^{*N}\right),\ s=\overline{1,S},$$
$$E'''_{l|s}\left(F^{*}\right) \triangleq \limsup_{N\to\infty}\left\{-\frac{1}{N}\log\alpha'''_{l|s}(F^{*N})\right\},\ l,s=\overline{1,S}$$

Using Lemma of types is defined by Cover and Thomas (1991) and Haroutunian *et al.* (2008), the following equality can be created:

$$\limsup_{N_1\to\infty}\left\{-\frac{1}{N_1}\log P_s^{N_1}\left(U'=m\right)\right\}=\inf_{Q:\ \min_{l\in D_m} D(Q\|P_{l(g)})\leq E'^{*}_{m|m}} D\left(Q\|P_{s(g)}\right)\triangleq E^{I}_{m|s},$$
$$s\notin D_m,\quad m=\overline{1,K} \tag{16}$$

According to Eq. 16 and definition of reliabilities are obtained:

$$E'''_{l|s}\left(\Phi^*\right)=\left(1-\psi\right)E''^{*}_{l|s},\ l,s\in D_k,\ k=\overline{1,K} \tag{17}$$

$$E'''_{l|s}\left(\Phi^*\right)=\psi E^{I}_{m|s}+\left(1-\psi\right)E''^{*}_{l|s},\ l\in D_m,\ s\in D_k,\ m,k=\overline{1,K} \tag{18}$$

$$E'''_{s|s}\left(\Phi^*\right)=\min_{l\neq s}E'''_{l|s}\left(\Phi^*\right),\ s\in D_k,\ k=\overline{1,K} \tag{19}$$

Theorem 4: If for different distributions $P_s,\ s=\overline{1,S}$ compatibility conditions of Theorems 2 and 3 are satisfied, then elements of reliabilities matrix $E'''(\phi^*)$ of the two-stage test are defined in Eq. 17-19.

When one of the compatibility conditions is violated, then at least one element of $E'''(\phi^*)$ is equal to zero.

CONCLUSION

The reliabilities of a distributed detection system investigated as the number of sensors tend to infinity. It is assumed that the sensor data are quantized into M-ary messages and transmitted to the fusion center for multiple hypotheses testing concerning many families of PDs. The optimal reliabilities in a pair of stages characterized and the compatibility conditions provided for this to happen and description of characteristics of LAO hypotheses testing of distributed detection investigated. The goal was to make a decision on many possible hypotheses, based on the messages received at the fusion centers.

NOTATIONS

N	=	Numbers of sensors
M	=	Numbers of messages
U	=	Message
U', U'', U_0	=	The fusion centre
D	=	Global decision rule
H	=	Hypothesis
g	=	Quantizer
X	=	The space of source data
U	=	The space of messages
S	=	Number of hypotheses
K	=	Number of families
R_k	=	Number of PDs in k-th family
P_k	=	k-th probability distribution(PD)
$x = (x_1, x_1,..., x_N)$	=	Vector of observations
$u = (u_1, u_1,..., u_N)$	=	Vector of transmitted messages
ϕ, φ_1, φ_2, Φ	=	Tests
α	=	Error probability
E	=	Reliability
R, D	=	Sets

REFERENCES

Ahlswede, R. and E.A. Haroutunian, 2006. On Statistical Hypotheses Optimal Testing and Identification. In: General Theory of Information Transfer and Combinatorics, Ahlswede, R., L. Baumer, N. Cai, H. Aydinian, V. Blinovsky, C. Deppe and H. Mashurian (Eds.). LNCS, Vol. 4123, Springer, Berlin, pp: 462-478.

Cover, T.M. and J.A. Thomas, 1991. Elements of Information Theory. John Wiley and Sons, New Jersey.

Haroutunian, E.A., 1990. Logarithmically asymptotically optimal testing of multiple statistical hypotheses. Prob. Control Inform. Theory, 19: 413-421.

Haroutunian, E.A., M.E. Haroutunian and A.N. Haroutunyan, 2008. Reliability criteria in information theory and in statistical hypothesis testing. Foundations Trends Commun. Inform. Theory, 4: 97-263.

Hoeffding, W., 1965. Asymptotically optimal tests for multinomial distributions. Annal. Math. Stat., 36: 369-401.

Hormozi Nejad, F., E.A. Haroutunian and P.M. Hakobyan, 2011. On LAO testing of multiple hypotheses for the pair of families of distributions. Proceedings of the Conference on Computer Science and Information Technologies, September 26-30, 2011, Yerevan, Armenia, pp: 135-138.

Hormozi Nejad, F. and E.A. Haroutunian, 2012a. Many hypotheses parallel distributed detection of the Pair of families of probability distributions. Trans. IIAP NAS RA Math. Problems Comput. Sci., 38: 61-65.

Hormozi Nejad, F. and E.A. Haroutunian, 2012b. On two-stage LAO multihypotheses testing for many distinct families of probability distributions. Proceedings of the 12th Islamic Countries Conference on Statistical Sciences Conference, Volume 23, December 19-22, 2012, Qatar University, Doha, Qatar, pp: 1-13.

Kreidl, P.O., J.N. Tsitsiklis and S.I. Zoumpoulis, 2011. On decentralized detection with partial information sharing among sensors. IEEE Trans. Signal Process., 59: 1759-1765.

Tenney, R.R. and N.R. Sandell, 1981. Detection with distributed sensors. IEEE Trans. Aerosp. Electron. Syst., 17: 501-510.

Tsitsiklis, J.N. and M. Athans, 1985. On the complexity of decentralized decision making and detection problems. IEEE Trans. Automat. Contr., 30: 440-446.

Tsitsiklis, J.N., 1988. Decentralized detection by a large number of sensors. Math. Contr. Signals Syst., 1: 167-182.

Tusnady, G., 1977. On asymptotically optimal tests. Annal. Stat., 5: 385-393.