On Three Hypotheses Robust Detection Design Under Mismatch

Evgueni Haroutunian

Institute for Informatics and Automation Problems of NAS RA
Yerevan, Armenia

e-mail: eghishe@sci.am

ABSTRACT

Generalizing the result of D. Kazacos for two hypotheses, we consider the ternary detection problem of the Neyman-Pearson type under mismatch. For the case of independent identically distributed observations, the sufficiency condition of existence of test with an exponentially dicreasing probability of error is formulated in terms of the new notion of "divergence for three distributions in certain order".

Keywords

Multiple statistical hypotheses, Continious distributions, Robust tests, Exponentially decreasing error probabilities.

1. INTRODUCTION

One of the general problems in statistics is the choice between different explanations (hypotheses) for the observed data concerning the studied object.

The considerable part in the stream of publications on the problem make up works and results for the cases of binary hypotheses and (or) discrete distributions [1]-[8], [10]-[13], [19], [22].

A natural extention of the binary hypotheses testing problem is multiple hypotheses testing. We examined ternary hypotheses under the vector of independent identically distributed observations.

Our investigation deals only with the case of exponential convergence of the error probability to zero, under the presence of mismatch. The exponential rate of decrease of the error probabilities is considered as a measure of test performance.

Frequently in practical problems the definition of a specific cost structure in testing is not possible. In such cases another, Neyman-Pearson criterion is imposed. In the present paper we apply this alternative formulation.

We consider the situation in which inaccurate versions $g_1(\mathbf{x})$, $g_2(\mathbf{x})$, $g_3(\mathbf{x})$ of the true densities $f_1(\mathbf{x})$, $f_2(\mathbf{x})$, $f_3(\mathbf{x})$ are used in the decision rule.

The reason for using incorrect *probability density functions* (pdfs) in the test implementation is that a suboptimal and computationally convenient decision rule may be preferable due to its simplicity.

In recent years several studies have considered the design of robust decision procedures [14]-[18], [20], [21], [24]. A statistical operation is called robust when its action does not feel small deviations of the situation from the given model. We may know the exact pdfs, but it may be expedient to use nominal pdfs. The test must have a build in tolerance which ensures that the test is performed appropriately not only for the given model, but for the entire class of models to the vicinity of it.

In section 2 the suboptimal binary detection scheme of D. Kazacos [16] is enlarged to the ternary test. Error probabilities upper bounds are proved proposing a log-liklihood ratio test analogical to that used by the author in [12] to generalize the Neyman-Pearson fundamental lemma for more than two hypotheses.

In section 3 sufficiency conditions for the existence of a test are formulated for the case of independent identically distributed observations using elegant formula for " the divergence of three distributions in a certain order".

In section 4 a simple case of a test with exact distributions is considered.

2. SETUP OF THE ROBUST TEST

Let $\mathbf{x} = (x_1, x_2, ..., x_N) \in \mathcal{X}^N$ be the vector of observations.

$f_1(\mathbf{x})$, $f_2(\mathbf{x})$, $f_3(\mathbf{x})$ are the true pdfs under three hypotheses H_1, H_2, H_3. $g_1(\mathbf{x})$, $g_2(\mathbf{x})$, $g_3(\mathbf{x})$ are the corresponding inaccurate pdfs used in a liklihood ratio (LR) test for the corresponding hypotheses. The following assumptions on the pdfs are made for $k, l, m = 1, 2, 3$:

(I) $f_k(\mathbf{x})$ and $g_k(\mathbf{x})$ have common support.

(II) $g_1(\mathbf{x})$, $g_2(\mathbf{x})$, $g_3(\mathbf{x})$ are not equal to each other almost everywhere with respect to either of the probability measures induced by the pdfs $f_1(\mathbf{x})$, $f_2(\mathbf{x})$, $f_3(\mathbf{x})$.

(III) $\int f_k(\mathbf{x})(g_l(\mathbf{x})/g_m(\mathbf{x}))d\mathbf{x} < +\infty$.

We consider the following log-LR test, that is analogical to one, which was first applied in [12], for $T_1 > 0$, $T_2 > 0$:

$$\text{accept} \quad H_1 \ \text{ if } \mathbf{x} \in \mathcal{A}_1 =$$

$$\{\mathbf{x} : g_1(\mathbf{x})/g_2(\mathbf{x}) > \exp\{NT_1\}, g_1(\mathbf{x})/g_3(\mathbf{x}) > \exp\{NT_1\}\}, \tag{1}$$

accept H_2 if

$$\mathbf{x} \in \mathcal{A}_2 = \overline{\mathcal{A}_1} \cap \{\mathbf{x} : g_2(\mathbf{x})/g_3(\mathbf{x}) > \exp\{NT_2\}\}, \quad (2)$$

$$\text{accept } \ H_3 \ \text{ if } \mathbf{x} \in \mathcal{A}_3 = \overline{\mathcal{A}_1 \cup \mathcal{A}_2}. \quad (3)$$

The probablities $\alpha_k^N, k = 1, 2, 3$, of erroneous acceptance of other hypotheses, provided that H_k is true, are

$$\alpha_k^N \triangleq 1 - \int_{A_k} f_k(x)dx, k = 1, 2, 3.$$

The thresholds $\exp\{NT_1\}$ and $\exp\{NT_2\}$ are chosen so for achieving error probabilities exponential decreasing to zero with $N \to \infty$.

Theorem 1: *Assume that $f_k(\mathbf{x})$ and $g_k(\mathbf{x})$ have common support, or equivalently, are absolutely continuous measures with respect to each other, for $k = 1, 2, 3$, $N = 1, 2, ...$ For any $s > 0$, with*

$$M_N(s, F_k, G_l, G_m) = \log \int f_k(\mathbf{x})(g_l(\mathbf{x})/g_m(\mathbf{x}))^s d\mathbf{x},$$

$$k, l, m = 1, 2, 3,$$

error probabilities of test defined in (1)- (3) can be upper bounded as follows:

$$\alpha_1^N \leq \exp\{sNT_1 + M_N(s, F_1, G_2, G_1)\}+$$

$$\exp\{sNT_1 + M_N(s, F_1, G_3, G_1)\}, \quad (4)$$

$$\alpha_2^N \leq \min[\exp\{-sNT_1 + M_N(s, F_2, G_1, G_2)\},$$

$$\exp\{-sNT_1 + M_N(s, F_2, G_1, G_3)\}]+$$

$$\exp\{sNT_2 + M_N(s, F_2, G_3, G_2)\}, \quad (5)$$

$$\alpha_3^N \leq \min[\exp\{-sNT_1 + M_N(s, F_3, G_1, G_2)\},$$

$$\exp\{-sNT_1 + M_N(s, F_3, G_1, G_3)\}]+$$

$$\exp\{-sNT_2 + M_N(s, F_3, G_3, G_2)\}]. \quad (6)$$

Proof: It follows from (II) that if $g_1 = 0$ then $f_1 = 0$, thus $Pr(g_1 = 0|H_1) = 0$. Denoting by E_1 the expectation under H_1 and using the Markov inequality $Pr(z > 1) \leq Ez^s$, $s > 0$ we obtain

$$\alpha_1^N = F_1(\overline{A}_1) \leq F_1(\mathbf{x} : g_1(\mathbf{x}) < g_2(\mathbf{x}) \exp\{NT_1\})+$$

$$F_1(\mathbf{x} : g_1(\mathbf{x}) \leq g_3(\mathbf{x}) \exp\{NT_1\}) =$$

$$F_1(\exp\{NT_1\}g_2(\mathbf{x})/g_1(\mathbf{x}) > 1)+$$

$$F_1(\exp\{NT_1\}g_3(\mathbf{x})/g_1(\mathbf{x}) > 1) \leq$$

$$\exp(sNT_1 + M_N(s, F_1, G_2, G_1))+$$

$$\exp\{sNT_2 + M_N(s, F_1, G_3, G_1)\}.$$

For α_2^N and α_3^N estimates (4) and (5) are proved analogically.

The following results [16] are necessary for development of conditions of the existence of the test.

Theorem 2 ([16]): *For (III) $\int f_k(\mathbf{x})g_l(\mathbf{x})g_m^{-1}(\mathbf{x})dx < +\infty$ for $0 \leq s < 1 - e^{-1}$ the first two derivatives of M_N with respect to s exist are finite and can be evaluated by interchanging integration and differentiation.*

We have for $s = 0$, $k, l, m = 1, 2, 3$,

$$M_N'(0, F_k, G_l, G_m) = D_N(F_k||G_m) - D_N(F_k||G_l) \quad (7)$$

with

$$D_N(F_k||G_m) \triangleq \int f_k(\mathbf{x}) \log[f_k(\mathbf{x})/g_m(\mathbf{x})]d\mathbf{x} \quad (8)$$

Kullback-Leibler divergence between $f_k(\mathbf{x})$ and $g_m(\mathbf{x})$.

The second derivative of M_N with respect to s is found in [16], by direct evaluation and by interchanging of differentiations with integrations, it is nonnegative for all s and equal to zero only if $g_l(\mathbf{x}) = g_m(\mathbf{x})$, $l \neq m$, almost everywhere with respect to the measure induced by $f_k(\mathbf{x})$.

Assumption (II) has, thus, excluded the possibility of $M_N'' = 0$. Under (I), (II), and (III) the function M_N is strictly convex and has the value of 0 at $s = 0$, and its derivative at $s = 0$ is given by (7) as a difference of two informational divergence expressions.

3. TEST FOR INDEPENDENT IDENTICALLY DISTRIBUTED OBSERVATIONS

For the case of independent and identically distributed (i.i.d) observations, we have assumed that $g_k(\mathbf{x})$ are also of the product form, therefore we have

$$M_N(s, F_k, G_l, G_m) = NM_1(s, f_k, g_l, g_m) \triangleq$$

$$N \log \int f_k(x)[g_l(x)/g_m(x)]^s dx. \quad (9)$$

From (4), (5), (7) by (9) we discover analogical estimates for α_1^N, α_2^N, α_3^N with $M_N(s, F_k, G_l, G_m)$ replaced correspondingly by $NM_1(s, f_k, g_l, g_m)$.

From (7), (8), (9), we find that for $k, l, m = 1, 2, 3$

$$M_1'(0, f_k, g_l, g_m) = D_1(f_k||g_m) - D_1(f_k||g_l), \quad (10)$$

with the corresponding Kullback-Leibler divergences.

We can express the differences of divergences somewhat briefly. Let us consider and denote by $D(f_k||f_l||f_m)$

$$D_1(f_k||f_n) - D_1(f_k||f_l) = \int f_n(x) \log[f_k(x)/f_m(x)]dx -$$

$$\int f_n(x) \log[f_k(x)/f_l(x)]dx =$$

$$\int f_k(x) \log[f_l(x/f_m(x)]dx \triangleq D(f_k||f_l||f_m).$$

We propose to conventionally name this expression "divergence of three distributions f_k, f_l, f_m, in this order".

Now if we assume that

$$D(f_1||g_2||g_1) + T_1 < 0,$$

then with some $s_1 > 0$ we will have

$$sT_1 + M_1(s, f_1, g_2, g_1) < 0, \quad \text{for } s \in [0, s_1].$$

Similarly, if we assume that

$$D(f_1||g_3||g_1) + T_1 < 0,$$

then with some $s_2 > 0$

$$sT_1 + M_1(s, f_1, g_3, g_1) < 0, \quad \text{for } s \in [0, s_2],$$

if

$$D(f_2||g_1||g_2) - T_1 < 0,$$

then for some $s_3 > 0$

$$-sT_1 + M_1(s, f_2, g_1, g_2) < 0, \quad \text{for } s \in [0, s_3],$$

if

$$D(f_2||g_1||g_3) - T_1 < 0,$$

then for $s_4 > 0$

$$-sT_1 + M_1(s, f_2, g_1, g_3) < 0, \quad \text{for } s \in [0, s_4],$$

if

$$D(f_2||g_3||g_2) + T_2 < 0,$$

then for $s_5 > 0$

$$sT_2 + M_1(s, f_2, g_3, g_2) < 0, \quad \text{for } s \in [0, s_5],$$

if

$$D(f_3||g_1||g_2) - T_1 < 0,$$

then for some $s_6 > 0$

$$-sT_1 + M_1(s, f_3, g_1, g_2) < 0, \quad \text{for } s \in [0, s_6],$$

if

$$D(f_3||g_1||g_3) - T_1 < 0,$$

then for $s_7 > 0$

$$-sT_1 + M_1(s, f_3, g_1, g_3) < 0, \quad \text{for } s \in [0, s_7],$$

and at last if for $s_8 > 0$

$$D(f_3||g_3||g_2) - T_2 < 0,$$

then it will be

$$-sT_2 + M_1(s, f_3, g_3, g_2) < 0, \quad \text{for } s \in [0, s_8].$$

Denote $s_0 = \min\limits_{i=\overline{1,8}} s_i$ and consider

$$T_1^+ = \min[D(f_1||g_1||g_2), D(f_1||g_1||g_3)], \tag{11}$$

$$T_1^- = \max[D(f_2||g_1||g_2), D(f_2||g_1||g_3),$$
$$D(f_3||g_1||g_3), D(f_3||g_1||g_2)], \tag{12}$$

$$T_2^+ = D(f_2||g_2||g_3), \tag{13}$$

$$T_2^- = D(f_3||g_3||g_2). \tag{14}$$

We can conclude. Conditions $s < s_0$ and $T_1^- < T_1^+$, $T_2^- < T_2^+$ guarantee exponentional decreasing to 0 of all terms in estimates (4)-(6) and then existence of the necessary test (1)-(3).

Theorem 3: *Conditions $T_1^- < T_1^+$, $T_2^- < T_2^+$ are sufficient for existence of test with exponentially decreasing error probabilities and thresholds T_1 between T_1^- and T_1^+ and T_2 between T_2^- and T_2^+.*

4. TEST WITH CORRECT DISTRIBUTIONS

Here we can return to the normal situation with exact distributions.

Let $g_k = f_k$, $k = 1, 2, 3$, instead (11)-(14) we obtain

$$\hat{T}_1^+ = \min[D(f_1||f_2), D(f_1||f_3)],$$

$$\hat{T}_1^- = \max[D(f_2||f_1||f_3), D(f_3||f_1||f_2)],$$

$$\hat{T}_2^+ = D(f_2||f_3),$$

$$\hat{T}_2^- = D(f_3||f_2).$$

As a consequence of Theorem 3 we obtain

Theorm 4: *Conditions $\hat{T}_1^- < \hat{T}_1^+$, $\hat{T}_2^- < \hat{T}_2^+$ are sufficient for existence of test with exponentially decreasing error probabilities and thresholds $\hat{T}_1$ between $\hat{T}_1^-$ and $\hat{T}_1^+$ and $\hat{T}_2$ between $\hat{T}_2^-$ and $\hat{T}_2^+$.*

This is an enlargement to the case of three hypotheses of the result of Chernoff [3] for the binary case.

5. CONCLUSION AND COMMENTS

Ternary likelihood decision rule is presented and analyszed when inaccurate version of the probability density functions is used. Upper bounds of the error probabilities are found.

For the case of independent and identically distributed observations we have provided sufficient conditions of existence of the required test.

Many problems are still open. It remains to find the necessary condition for existence of the test. Next question to study is the test for more than three hypotheses. It is desirable to know if the notion of divergance of three distributions is usefull in other situations, it is necessary to investigate its inportant properties (see [9], [23]).

REFERENCES

[1] E. Biglieri and L. Györty, "Some remarks on robust binary hypothesis testing", *Prociding of IEEE International Symposium on Information Theory*, pp. 566-570, 2014.

[2] R. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, Reading Publishing Company, 1987.

[3] A. Chernoff, "A measure of assymtotic efficiency for teest of a hypothesis based on the sum of observations", *Ann. Math. Statist.*, vol. 23, pp. 493-507, 1952.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Second Edition, Wiley, New York, 2006.

[5] I. Csiszár, "Information-type measures of divergence of probability distributions and inderict observations", *Studia Scientiarum Mthematica*, Hungaria, no. 2, pp. 299–318, 1967.

[6] I. Csiszár, "I-divergence geometry of probability distributions and minimization problems", *The Annals of Probability*, vol. 3, no. 1, pp. 146–158, 1975.

[7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981, (Russian translation, Mir, Moscow, 1985).

[8] I. Csiszár, and P. Shields, *Information Theory and Statistics*, Tutorial, Foundations and Trends in Communications and Infromation Theory, vol. 1, no. 4, 2004.

[9] A. Ghosh and A. Basu, "A new family of divergences originating from model adequacy tests and application to robust statistical inference", *IEEE Trans. on Information Theory*, vol. 64, no. 6, pp. 5581-5590, 2018.

[10] E. Haroutunian, "Logarithmically asymptotically optimal testing of multiple statistical hypotheses", *Problems of Controls and Information Theory*, vol. 19, no. 5, pp. 413-421, 1990.

[11] E. Haroutunian, M. Haroutunian and A. Harutyunyan, Reliability Criteria in Information Theory and in Statistical Hypotheses Testing, Foundations and Trends in Communications and Infromation Theory, vol. 4, no. 2-3, 171 pages, 2008.

[12] E. Haroutunian, "Neyman-Pearson principle for more than two hypotheses", *Abstracts of Armenian Mathematical union annual session*, pp. 49-50, 2013.

[13] E. Haroutunian and P. Hakobyan, "Most powerful test for multiple hypotheses", *Proccedings of Conference Computer Science and Information Technology*, Yerevan, Armenia, pp. 110-112, 2015.

[14] P. J. Huber, *Robust Statistics*, Wiley and Sons, N. Y. 1981.

[15] P. J. Huber, "A robust version of the probability ratio test", *Annals Mathematical Statistics*, vol. 36, pp. 1753-1758, 1965.

[16] D. Kazacos, "Signal detection under mismatch", *IEEE Transactions on Information Theory*, vol. 26, no. 4, pp. 681-684, 1982.

[17] D. Kazacos and P. Papantoni-Kazakos, *Detection and Estimation*, Computer Science Press, 1990.

[18] N. M. Khalfina, L. A. Khalfin, "On a robust version of the probability ratio test", *Teor. Veroyatnost. i priminen.*, vol. 20, no. 1, pp. 203-206, 1975.

[19] E. Lehmann and J. Romano, *Testing Statistical Hypotheses*, New York, Springer, third ed., 2005.

[20] B. C. Levy, *Principles of Singnal Detection and Parameter Estimation*, Springer, 2008.

[21] B. C. Levy, "Robust hypothesis testing with a relative entropy tolerance", *IEEE Trans. on Information Theory*, vol. 55, no. 1, pp. 413-421, 2009.

[22] V. Poor, *An Introduction to Signal Detection and Estimation*, Second Edition, Springer, 1994.

[23] T. van Even and P. Harremoës, "Renyi divergence and Kullback-Leibler divergence", *IEEE Trans. on Information Theory*, vol. 60, no. 3, pp. 3797-3820, 2014.

[24] P. Yang and B. Chen, "Robust Kullback-Leibler divergence and universal hypothesis testing for continous distributions", *IEEE Trans. on Information Theory*, vol. 65, no. 4, pp. 2360-2373, 2019.