Universal Multiple Outlier Hypothesis Testing

Yun Li, Sirin Nitinawarat and Venugopal V. Veeravalli
Department of Electrical and Computer Engineering
and
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign, Urbana, IL 61801
Emails: {yunli2,nitinawa,vvv}@illinois.edu.

***Abstract*—The universal multiple outlier hypothesis testing problem is studied in two settings. In the first setting, each outlier can be arbitrarily distributed, and the number of outliers is fixed and known. In the second setting, the number of outliers is unknown at the outset. Nothing is known about the typical and outlier distributions other than that they are different and have full supports. For the first setting, a universally exponentially consistent test is proposed, and its achievable error exponent is characterized. The limiting error exponent achieved by such test is analyzed as the number of coordinates goes to infinity, and it is shown that the test also enjoys universally asymptotically exponential consistency. For the second setting, it is shown that with the assumption of outliers being identically distributed and the exclusion of the null hypothesis, a test based on the generalize likelihood principle is universally exponentially consistent.**

I. Introduction

In [1] and [2], we studied the inference problem of *universal outlier hypothesis testing*, which involves identifying a small subset of outlier coordinates efficiently. Universal outlier hypothesis testing finds applications in event detection and environment monitoring in sensor networks [3], understanding of visual search in humans and animals [4], fraud and anomaly detection [5], [6] in large data sets, and optimal search and target tracking [7]. It was assumed in [1] and [2] that the outlier coordinates are identically distributed according to the "outlier" distribution, which is distinct from the common "typical" distribution that governs the rest of the coordinates. The number of outliers is fixed and known at the outset. This inference problem was studied in the universal setting without any prior knowledge about the outlier and typical distributions.

The main finding in [1] and [2] is that one can construct universal tests for the outlier hypothesis testing problem that are far more efficient than those for the other inference problems previously studied in the universal setting, such as homogeneity testing or classification [8]–[12]. In particular, the test that we proposed in [1] and [2] for the outlier hypothesis testing problem is universally exponentially consistent, and it is impossible to achieve *universally exponential consistency* for homogeneity testing or classification without training data [11], [12]. In addition, we showed that our proposed test was *asymptotically efficient* in the sense that its achievable error exponent converged to the absolutely optimal error exponent when both the outlier and typical distributions were known.

It is to be noted that we made two important simplifying assumptions in [2]: first, that all the outlier coordinates were identically distributed, and second, that the number of outliers was known exactly at the outset. The purpose of this paper is to show that these assumptions can be relaxed substantially. To this end, we consider two new models, each an extension of the model considered in [2].

In the first new model, each outlier can be arbitrarily distributed as long as it is different from the typical distribution. It is interesting that the universal test proposed in our previous work [2] does *not* rely on the assumption of identically distributed outliers, and is directly applicable to this extension. We characterize the achievable error exponent for our proposed test and show that it is universally exponentially consistent. In this new model, the absolutely optimal (positive) error exponent (when the outlier and typical distributions are known) depends on the underlying distributions and the number of coordinates, and can vanish to zero as the number of coordinates goes to infinity. We show that the limit of the error exponent achievable by our universal test is always positive whenever the limit of the absolutely optimal error exponent is. We call this property *universally asymptotically exponential consistency*.

In the second new model, we relax the assumption that the number of outliers is known exactly at the outset. In particular, we show that the assumption of the outliers being identically distributed and the exclusion of the null hypothesis with no outlier present are critical for the existence of a universally exponentially consistent test. If either fails to hold, we show that there cannot exist a universally exponentially consistent test. In all other cases with uncertainty in the number of outliers, we show, through the use of a universal test that *strictly* follows the generalized likelihood principle, that it is always possible to obtain universally exponential consistency.

978-1-4673-3146-3/13/$31.00 ©2013IEEE

The outline of the paper is as follows. We start by describing a generic model in Section II with possibly distinctly distributed outliers and without the number of outliers being known exactly. Some useful decision-theoretic distance metrics between distribution pairs and technical facts are also reviewed in this section. Models pertaining to the first extension with distinctly distributed outliers but with the number of outliers being known exactly are discussed in Section III. The extension with identically distributed outliers but without the knowledge of the number of outliers is treated in Section IV.

II. Preliminaries

Throughout the paper, we denote random variables by capital letters and their realizations by the corresponding lower-case letters. All random variables are assumed to take values in finite sets, and all logarithms are the natural ones.

For a finite set $\mathcal{Y}$, let $\mathcal{Y}^m$ denote the m Cartesian product of $\mathcal{Y}$, and $\mathcal{P}(\mathcal{Y})$ denote the set of all probability mass functions (pmfs) on $\mathcal{Y}$. The empirical distribution of a sequence $\boldsymbol{y} = y^m = (y_1, \ldots, y_m) \in \mathcal{Y}^m$, denoted by $\gamma = \gamma_{\boldsymbol{y}} \in \mathcal{P}(\mathcal{Y})$, is defined as

$$\gamma(y) \triangleq \frac{1}{m} \big| \{k = 1, \ldots, m : y_k = y\} \big|, \quad y \in \mathcal{Y}.$$

Consider n independent and identically distributed (i.i.d.) vector observations, each of which has M independent coordinates. We denote the i-th coordinate of the k-th observation by $Y_k^{(i)} \in \mathcal{Y}$. It is assumed that most coordinates are commonly distributed according to the "typical" distribution $\pi \in \mathcal{P}(\mathcal{Y})$ except for a small (possibly empty) subset $S \subset \{1, \ldots, M\}$ of "outlier" coordinates, each of which is assumed to be distributed according to an outlier distribution $\mu_i,\ i \in S$. *Nothing is known about* $\{\mu_i\}_{i=1}^M$ *and* π *except that each* $\mu_i \neq \pi,\ i = 1, \ldots, M$, *and that all* $\mu_i,\ i = 1, \ldots, M$, *and* π *have full supports.* In the following presentation, we sometimes consider the special case when all the outlier coordinates are identically distributed, i.e., $\mu_i = \mu,\ i = 1, \ldots, M$.

For a hypothesis corresponding to an outlier subset $S \subset \{1, \ldots, M\},\ |S| < \frac{M}{2}$, the joint distribution of all the observations is

$$\begin{aligned} p_S\left(y^{Mn}\right) &= p_S\left(\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(M)}\right) \\ &= \prod_{k=1}^{n} \left\{ \prod_{i \in S} \mu_i\left(y_k^{(i)}\right) \prod_{j \notin S} \pi\left(y_k^{(j)}\right) \right\}, \end{aligned}$$

where

$$\boldsymbol{y}^{(i)} = \left(y_1^{(i)}, \ldots, y_n^{(i)}\right),\ i = 1, \ldots, M.$$

We refer to the unique hypothesis corresponding to the case with *no* outlier, i.e., $S = \emptyset$, as the *null* hypothesis. In the following sections, we shall consider different settings, each being described by a suitable set $\mathcal{S}$ comprising all possible outlier subsets.

The test for the (true) outlier subset is done based on a *universal* rule $\delta : \mathcal{Y}^{Mn} \to \mathcal{S}$. In particular, the test δ is not allowed to depend on $\left(\{\mu_i\}_{i=1}^M, \pi\right)$.

For a universal test, the maximal error probability, which is a function of the test and $\left(\{\mu_i\}_{i=1}^M, \pi\right)$, is

$$e\left(\delta, \left(\{\mu_i\}_{i=1}^M, \pi\right)\right) \triangleq \max_{S \in \mathcal{S}} \sum_{y^{Mn}:\ \delta(y^{Mn}) \neq S} p_S\left(y^{Mn}\right),$$

and the corresponding error exponent is defined as

$$\alpha\left(\delta, \left(\{\mu_i\}_{i=1}^M, \pi\right)\right) \triangleq \lim_{n \to \infty} -\frac{1}{n} \log e\left(\delta, \left(\{\mu_i\}_{i=1}^M, \pi\right)\right).$$

Our results will be stated in terms of various distance metrics between a pair of distribution $p, q \in \mathcal{P}(\mathcal{Y})$. In particular, we shall consider two symmetric distance metrics: the *Bhattacharyya distance* and *Chernoff information*, denoted respectively by $B(p, q)$ and $C(p, q)$, and defined as (see, e.g., [13])

$$B(p, q) \triangleq -\log\left(\sum_{y \in \mathcal{Y}} p(y)^{\frac{1}{2}} q(y)^{\frac{1}{2}}\right) \tag{1}$$

and

$$C(p, q) \triangleq \max_{s \in [0,1]} -\log\left(\sum_{y \in \mathcal{Y}} p(y)^s q(y)^{1-s}\right), \tag{2}$$

respectively. Another distance metric, which will be key to our study, is the relative entropy, denoted by $D(p\|q)$ and defined as

$$D(p\|q) \triangleq \sum_{y \in \mathcal{Y}} p(y) \log \frac{p(y)}{q(y)}. \tag{3}$$

Unlike the Bhattacharyya distance (1) and Chernoff information (2), the relative entropy in (3) is a *non-symmetric* distance [13].

III. Models with Known Number of Outliers

We start by considering the case in which the number of outliers, denoted by $T > 1$, is known at the outset, i.e., $|S| = T$, for every $S \in \mathcal{S}$. Note that unlike in [2][Section VI] wherein it was assumed that all outlier coordinates are identically distributed, in the model being currently considered in this section, the distributions of outlier coordinates $\mu_i,\ i \in S$, can be distinct.

A. Proposed Universal Test

In [2], we proposed a universal test based on the generalized likelihood principle for two setups: the setup when only π is known and the completely universal setup. We employ this same test in the current setup of this section. We now give a summary of this test; its detailed derivation can be found in [2].

The test is done based on the following statistics. For each $S \in \mathcal{S}$,

$$U_S^{\text{typ}}\big(y^{Mn}\big) \triangleq \sum_{j \notin S} D(\gamma_j \| \pi) \tag{4}$$

for the setup when only π is known, and

$$U_S^{\text{univ}}\big(y^{Mn}\big) \triangleq \sum_{j \notin S} D\left(\gamma_j \,\middle\|\, \tfrac{\sum_{k \notin S} \gamma_k}{M-T}\right) \tag{5}$$

for the completely universal setup, where γ_j denotes the empirical distribution of $\boldsymbol{y}^{(j)}$. These statistics are related to the negative of the generalized log-likelihood of y^{Mn} of the hypothesis corresponding to an outlier subset $S \subset \{1, \ldots, M\}$ for the respective setups (see [2]). The test then selects the hypothesis with the largest such generalized log-likelihood (ties are broken arbitrarily), i.e.,

$$\delta\big(y^{Mn}\big) = \underset{S \subset \{1,\ldots,M\},\ |S|=T}{\operatorname{argmin}} U_S^{\text{typ}}\big(y^{Mn}\big) \tag{6}$$

when only π is known, and

$$\delta\big(y^{Mn}\big) = \underset{S \subset \{1,\ldots,M\},\ |S|=T}{\operatorname{argmin}} U_S^{\text{univ}}\big(y^{Mn}\big), \tag{7}$$

for the completely universal setup.

B. Performance of the Proposed Test

Proposition 1. *For every fixed number of outliers $T > 1$, when all the $\mu_i,\ i = 1, \ldots, M$, and π are known, the optimal error exponent is equal to*

$$\min_{1 \le i < j \le M} C\left(\mu_i(y)\,\pi(y'), \pi(y)\,\mu_j(y')\right). \tag{8}$$

When all outlier coordinates are identically distributed, i.e., $\mu_i = \mu \neq \pi,\ i = 1, \ldots, M$, this optimal error exponent is independent of M and is equal to (cf. [2][Theorem 4])

$$2B(\mu, \pi). \tag{9}$$

Theorem 2. *For every fixed number of outliers $T > 1$, when only π is known but none of $\mu_i,\ i = 1, \ldots, M$ is known, the error exponent achievable by our test in (4), (6) is equal to*

$$\min_{1 \le i \le M} 2B(\mu_i, \pi). \tag{10}$$

When all outlier coordinates are identically distributed, i.e., $\mu_i = \mu, i = 1, \ldots, M$, this achievable error exponent is equal to, cf. [2][Theorem 4],

$$2B(\mu, \pi), \tag{11}$$

which, from Proposition 1, is the optimal error exponent when μ is also known.

Remark 1. *Since the tester in Proposition 1 is more capable (with π known) than that in Theorem 2, the optimal error exponent in (8) must be no smaller than that in (10). This is verified simply by noting that for every $i, j,\ 1 \le i < j \le M$, it follows from (2) that*

$$\begin{aligned} &C\left(\mu_i(y)\,\pi(y'), \pi(y)\,\mu_j(y')\right) \\ &= \max_{s \in [0,1]} -\log\Big(\sum_{y,y'} \left(\mu_i(y)\,\pi(y')\right)^s \left(\pi(y)\,\mu_j(y')\right)^{1-s}\Big) \\ &\ge B(\mu_i, \pi) + B(\mu_j, \pi) \\ &\ge \min\left(2B(\mu_i, \pi), 2B(\mu_j, \pi)\right). \end{aligned} \tag{12}$$

Like in [2], an important consideration that we shall use to gauge the performance of a universal test is universally exponential consistency. Specifically, a universal test δ is termed *universally exponentially consistent* if for every $\mu_i,\ i = 1, \ldots, M,\ \mu_i \neq \pi$, it holds that $\alpha\left(\delta, \left(\{\mu_i\}_{i=1}^M, \pi\right)\right) > 0$. Although *universally exponential consistency* seems like a strong condition, it needs not ensure that

$$\lim_{M \to \infty} \alpha\left(\delta, \left(\{\mu_i\}_{i=1}^M, \pi\right)\right) > 0. \tag{13}$$

Of course, it follows from (8) in Proposition 1 that (13) is not possible for $\left(\{\mu_i\}_{i \ge 1}, \pi\right)$ such that

$$\lim_{M \to \infty} \min_{1 \le i < j \le M} C\left(\mu_i(y)\,\pi(y'), \pi(y)\,\mu_j(y')\right) = 0. \tag{14}$$

A test that satisfies (13) whenever (14) *does not* hold is said to enjoy *universally asymptotically exponential consistency*.

Theorem 3. *For every fixed number of outliers $T > 1$, our proposed test δ in (5), (7) is universally exponentially consistent. Furthermore, for every $\{\mu_i\}_{i=1}^M, \pi \in \mathcal{P}(\mathcal{Y}),\ \mu_i \neq \pi,\ i = 1, \ldots, M$, it holds that*

$$\begin{aligned} &\alpha\left(\delta, \left(\{\mu_i\}_{i=1}^M, \pi\right)\right) \\ &= \min_{\substack{S, S' \subset \{1,\ldots,M\} \\ |S| = |S'| = T}} \min_{q_1, \ldots, q_M} \sum_{i \in S} D(q_i \| \mu_i) + \sum_{j \notin S} D(q_j \| \pi), \end{aligned}$$

where the inner minimum above is over the set of $(q_1, \ldots, q_M)$ such that

$$\sum_{i \notin S} D\Big(q_i \,\Big\|\, \tfrac{1}{M-T} \sum_{k \notin S} q_k\Big) \ge \sum_{i \notin S'} D\Big(q_i \,\Big\|\, \tfrac{1}{M-T} \sum_{k \notin S'} q_k\Big).$$

Theorem 4. *For every fixed number of outliers $T > 1$, the error exponent achievable by our proposed test in (5), (7) is lower bounded by*

$$\min_{\substack{q \in \mathcal{P}(\mathcal{Y}) \\ D(q\|\pi) \le \frac{1}{M-T}\left(\min_{i=1,\ldots,M} 2B(\mu_i, \pi) + TC_\pi\right)}} \ \min_{i=1,\ldots,M} 2\,B(\mu_i, q),$$

where $C_\pi \triangleq -\log\left(\min_{y\in\mathcal{Y}} \pi(y)\right) < \infty$.

Furthermore, the proposed test enjoys universally asymptotically exponential consistency. In particular, as $M \to \infty$*, the error exponent achievable by our test in (5), (7) converges as*

$$\lim_{M\to\infty} \alpha\left(\delta, \left(\{\mu_i\}_{i=1}^{M}, \pi\right)\right) = \lim_{M\to\infty} \min_{i=1,\ldots,M} 2B\left(\mu_i, \pi\right),$$

which is the limit of the achievable error exponent when the typical distribution is known.

When all outlier coordinates are identically distributed, i.e., $\mu_i = \mu \neq \pi,\ i = 1, \ldots, M$*, our test is asymptotically efficient by which it means that*

$$\lim_{M\to\infty} \alpha\left(\delta, (\mu, \pi)\right) = 2B\left(\mu, \pi\right), \tag{15}$$

which from Proposition 1, is equal to the optimal error exponent when both μ *and* π *are known.*

IV. Models with Unknown Number of Outliers

In this section, we look at the case in which not all hypotheses in $\mathcal{S}$ have the same number of outliers, i.e., there is uncertainty in the number of outliers in the hypothesis testing problem.

A. Nonexistence of Universally Exponentially Consistent Tests

We start with some cases in which it is impossible to construct a universally exponentially consistent test.

Theorem 5. *Under the assumption that all the outliers are identically distributed, for every hypothesis set containing the null hypothesis, there cannot exist a universally exponentially consistent test even when the typical distribution is known.*

When the outlier coordinates can be distinctly distributed, and when the typical distribution is known, there cannot exist a universally exponentially consistent test even when the null hypothesis is excluded, i.e., there are always some outlier coordinates (regardless of the hypothesis).

B. Models with Positive Number of Identical Outliers

We now consider the case not covered in Theorem 5, i.e., when the null hypothesis is excluded and all outlier coordinates are identically distributed. In particular, we show that in this case, it becomes possible again to construct a universally exponentially consistent test.

1) Proposed Universal Test: Following the generalized likelihood principle similar to as in [2] but now *with the assumption of identical outliers being taken strictly*, the negative of the generalized log-likelihood of y^{Mn} corresponding to an outlier subset $S \in \mathcal{S}$ for the completely universal setting, denoted by $\bar{U}_S^{\text{univ}}(y^{Mn})$, can be shown to be equivalent to

$$\bar{U}_S^{\text{univ}}(y^{Mn}) \triangleq \sum_{i\in S} D\left(\gamma_i \middle\| \frac{\sum_{k\in S}\gamma_k}{T}\right) + \sum_{j\notin S} D\left(\gamma_j \middle\| \frac{\sum_{k\notin S}\gamma_k}{M-T}\right), \tag{16}$$

and our universal test can be described as

$$\delta(y^{Mn}) = \operatorname*{argmin}_{S\in\mathcal{S}} \bar{U}_S^{\text{univ}}(y^{Mn}). \tag{17}$$

2) Universally Exponential Consistency of the Proposed Test:

Theorem 6. *Under the assumption that all the outliers are identically distributed, for every hypothesis set* excluding *the null hypothesis, our proposed test in (16), (17) is universally exponentially consistent.*

Acknowledgment

This work was supported by the Air Force Office of Scientific Research (AFOSR) grant FA9550-10-1-0458, through the University of Illinois at Urbana-Champaign, by the U.S. Defense Threat Reduction Agency through subcontract 147755 at the University of Illinois from prime award HDTRA1-10-1-0086, and by the National Science Foundation grant NSF CCF 11-11342.

References

[1] Y. Li, S. Nitinawarat and V. V. Veeravalli, "Universal outlier hypothesis testing," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 7-12 2013.

[2] ——, "Universal outlier hypothesis testing," *IEEE Trans. Inf. Theory*, submitted, 2013, also available at http://arxiv.org/.

[3] J. Chamberland and V. V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Process. Mag.*, vol. 24, pp. 16–25, 2007.

[4] N. K. Vaidhiyan, S. P. Arun and R. Sundaresan, "Active sequential hypothesis testing with application to a visual search problem," in *Proc. IEEE Int. Symp. Inf. Theory*, 2012, pp. 2201–2205.

[5] R. J. Bolten and D. J. Hand, "Statistical fraud detection: A review," *Statistical Science*, vol. 17, pp. 235–249, 2002.

[6] V. Chandola, A. Banerjee and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, pp. 15.1–15.58, 2009.

[7] L. D. Stone, *Theory of Optimal Search*. Topics in Operations Research Series, INFORMS, 2004.

[8] K. Pearson, "On the probability that two independent distributions of frequency are really samples from the same population," *Biometrika*, vol. 8, pp. 250–254, 1911.

[9] O. Shiyevitz, "On Rényi measures and hypothesis testing," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 31-Aug. 5 2011, pp. 894–898.

[10] J. Unnikrishnan, "On optimal two sample homogeneity tests for finite alphabets," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 1-6 2012, pp. 2027–2031.

[11] J. Ziv, "On classification with empirically observed statistics and universal data compression," *IEEE Trans. Inf. Theory*, vol. 34, pp. 278–286, 1988.

[12] M. Gutman, "Asymptotically optimal classification for multiple tests with empirically observed statistics," *IEEE Trans. Inf. Theory*, vol. 35, pp. 401–408, 1989.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley and Sons, Inc., 1991.