Turkish Journal of Mathematics

http://journals.tubitak.gov.tr/math/

Research Article

Turk J Math
(2023) 47: 678 – 686
© TÜBİTAK
doi:10.55730/1300-0098.3386

Novel Fano type lower bounds on the minimum error probability of list M-ary hypothesis testing

Berkan DÜLEK*
Department of Electrical and Electronics Engineering, Hacettepe University, Beytepe Campus, Ankara, Turkey

Received: 03.01.2022 • **Accepted/Published Online:** 19.01.2023 • **Final Version:** 09.03.2023

Abstract: The problem of list M-ary hypothesis testing with fixed list size $L < M$ is considered. Based on some random observation, the test outputs a list of L candidates out of M possible hypotheses. The probability of list error is defined as the probability of the event that the list output by the test does not contain the true hypothesis that has generated the observation. An identity is derived that relates the minimum average probability of error of the optimal list hypothesis test to the minimum average probability of error of an optimal maximum a posteriori probability decision rule. The latter decides among an alternative set of hypotheses corresponding to all possible L-component mixtures of the distributions that characterize the observation under the original M candidate hypotheses. As an application, the proposed identity is employed to obtain novel Fano type lower bounds on the minimum error probability of list M-ary hypothesis testing.

Key words: List M-ary hypothesis testing, probability of error, lower bound, Fano's inequality

1. Preliminaries

Let Y be a random observation taking values in a set $\mathcal{Y}$, which can be discrete or a subset of Euclidean space. Similar to the Bayesian M-ary hypothesis testing framework, we assume that there are M competing hypotheses, denoted with $\mathcal{H}_i$, $i = 1, \ldots, M$, which represent the state of nature that generates the observation Y. Each hypothesis $\mathcal{H}_i$ is assumed to have prior probability $\pi_i > 0$ such that $\sum_{i=1}^{M} \pi_i = 1$. Under hypothesis $\mathcal{H}_i$, the random observation $Y \in \mathcal{Y}$ is drawn with density $p_i(y)$. The posterior probability of the hypothesis $\mathcal{H}_i$ conditioned on the observation $Y = y$ is expressed as

$$\pi(i|y) := \frac{\pi_i p_i(y)}{p(y)} \quad \text{for } i = 1, \ldots, M\,, \tag{1.1}$$

where $p(y) = \sum_{i=1}^{M} \pi_i p_i(y)$ denotes the marginal density of the observation. Following the notation of [6], the probability of a set $\mathcal{A} \subset \mathcal{Y}$ under hypothesis $\mathcal{H}_i$ is written as the Lebesgue-Stieltjes integral:

$$\Pr(\mathcal{A}|\mathcal{H}_i \text{ is true}) = \int_{\mathcal{A}} p_i(y) d\mu(y), \tag{1.2}$$

where μ is a finite measure on the collection of all subsets of $\mathcal{Y}$, equal to the Lebesgue measure in the continuous case ($d\mu(y) = dy$, or equivalently $\mu(\mathcal{A}) = \int_{\mathcal{A}} dy$) and the counting measure in the discrete case. The counting

*Correspondence: berkan@ee.hacettepe.edu.tr
2010 *AMS Mathematics Subject Classification:* 94A13, 62C05

This work is licensed under a Creative Commons Attribution 4.0 International License.

measure on a set of points $y_1, y_2, \ldots$ is defined as $\mu(\mathcal{A}) = \sum_k \mathbb{I}\{y_k \in \mathcal{A}\}$, where $\mathbb{I}\{\cdot\}$ denotes the indicator function.

2. List M-ary hypothesis testing

List M-ary hypothesis testing (also known as list decoding [5]) is a variation of Bayesian M-ary hypothesis testing, in which the test outputs a list of L candidates out of M possible hypotheses. There are $\binom{M}{L}$ distinct list configurations, where $\binom{M}{L} = \frac{M!}{L!(M-L)!}$. Let $\ell = \{\ell_j : j = 1, \ldots, \binom{M}{L}\}$ denote the set of all distinct list configurations, where a list $\ell_j \in \ell$ can be represented as $\ell_j = \{\mathcal{H}_{j_1}, \mathcal{H}_{j_2}, \ldots, \mathcal{H}_{j_L}\}$ for some $1 \le j_1 < j_2 < \cdots < j_L \le M$. The set of $\binom{M}{L}$-level randomized decision rules, denoted by $\mathcal{D}$, is considered, i.e. $\delta = (\delta_j)_{j=1}^{\binom{M}{L}} \in \mathcal{D}$ such that $\sum_{j=1}^{\binom{M}{L}} \delta_j(y) = 1$ and $\delta_j(y) \in [0,1]$ for every $y \in \mathcal{Y}$. More explicitly, having observed $Y = y$, the test outputs list ℓ_j with probability $\delta_j(y)$ for every $j = 1, \ldots, \binom{M}{L}$. The probability of selecting list ℓ_j under hypothesis $\mathcal{H}_i$ is then given by

$$\Pr(\text{Select } \ell_j | \mathcal{H}_i \text{ is true}) = \int_{\mathcal{Y}} \delta_j(y) p_i(y) d\mu(y) \,. \tag{2.1}$$

An error is declared if the list output by the test does not contain the true hypothesis. Under hypothesis, $\mathcal{H}_i$, the probability of error of a list hypothesis test specified by $\delta(y)$ is obtained as

$$P_{e,i}(\delta(y)) = \sum_{\substack{j=1 \\ \mathcal{H}_i \notin \ell_j}}^{\binom{M}{L}} \Pr(\text{Select } \ell_j | \mathcal{H}_i \text{ is true}) \,, \tag{2.2}$$

and the corresponding average probability of error is $P_e(\delta(y)) = \sum_{i=1}^{M} \pi_i P_{e,i}$. As intuition suggests, the optimal test that minimizes the average probability of error outputs a list according to the L largest posterior probability values in $\{\pi(i|y)\}_{i=1}^{M}$ [5, Section II.C], [3, Lemma 1]. For completeness, we present proof of this result in the following based on the notation used in this paper.

Lemma 2.1 *Given observation $y \in \mathcal{Y}$, the optimum test $\delta^*(y) \in \mathcal{D}$ that minimizes the average probability of error of the list $M-$ary hypothesis testing problem selects a list $\{\mathcal{H}_{i_1^*}, \mathcal{H}_{i_2^*}, \ldots, \mathcal{H}_{i_L^*}\}$ such that*

$$(i_1^*, i_2^*, \ldots, i_L^*) = \operatorname*{argmax}_{1 \le i_1 < i_2 < \cdots < i_L \le M} \sum_{l=1}^{L} \pi(i_l | y) \,, \tag{2.3}$$

and the minimum average probability of error is

$$P_e(\delta^*(y)) = 1 - \int_{\mathcal{Y}} \max_{1 \le i_1 < i_2 < \cdots < i_L \le M} \left\{ \sum_{l=1}^{L} \pi_{i_l} p_{i_l}(y) \right\} d\mu(y) \,. \tag{2.4}$$

Proof The average probability of error of any given test $\delta(y) \in \mathcal{D}$ can be expressed as

$$P_e(\delta(y)) = \sum_{j=1}^{\binom{M}{L}} \sum_{\substack{i=1 \\ \mathcal{H}_i \notin \ell_j}}^{M} \pi_i \, Pr(\text{Select } \ell_j | \mathcal{H}_i \text{ is true}) \tag{2.5}$$

$$= \int_{\mathcal{Y}} p(y) \sum_{j=1}^{\binom{M}{L}} \delta_j(y) \left(\sum_{\substack{i=1 \\ \mathcal{H}_i \notin \ell_j}}^{M} \pi(i|y) \right) d\mu(y) \,, \tag{2.6}$$

where the order of summation with respect to i and j is interchanged in (2.5) and (1.1) is substituted in (2.6). Defining the posterior cost of selecting list ℓ_j as

$$C(j|y) := \sum_{\substack{i=1 \\ \mathcal{H}_i \notin \ell_j}}^{M} \pi(i|y) = 1 - \sum_{\substack{i=1 \\ \mathcal{H}_i \in \ell_j}}^{M} \pi(i|y) \,, \tag{2.7}$$

(2.6) can be written as

$$P_e(\delta(y)) = \int_{\mathcal{Y}} p(y) \left(\sum_{j=1}^{\binom{M}{L}} \delta_j(y) C(j|y) \right) d\mu(y) \,. \tag{2.8}$$

Since $\delta_j(y) \in [0,1]$ and $\sum_{j=1}^{\binom{M}{L}} \delta_j(y) = 1$, $\sum_{j=1}^{\binom{M}{L}} \delta_j(y) C(j|y) \geq \min_{1 \leq j \leq \binom{M}{L}} \{C(j|y)\}$ and we get

$$P_e(\delta(y)) \geq \int_{\mathcal{Y}} p(y) \min_{1 \leq j \leq \binom{M}{L}} \{C(j|y)\} d\mu(y) \,. \tag{2.9}$$

The lower bound in (2.9) is achieved if, for each $y \in \mathcal{Y}$, we set

$$\delta_k^*(y) = 1 \text{ for } k = \operatorname*{argmin}_{1 \leq j \leq \binom{M}{L}} C(j|y) \tag{2.10}$$

(and hence, $\delta_j^*(y) = 0$ for all $j \neq k$), i.e. the optimal test outputs the list that minimizes $C(j|y)$ over all $1 \leq j \leq \binom{M}{L}$, which can be equivalently expressed as in (2.3) yielding the average error probability given in (2.4). □

The optimum test can be easily implemented by performing the following operations sequentially:

1. compute the posterior probabilities $\pi(i|y)$ for all $1 \leq i \leq M$ based on the observation y,
2. rank them in decreasing order,
3. select the list formed by the leading L hypotheses with the highest posterior probability values.

When multiple distinct lists achieve the same maximum score in (2.3) for a given observation y, the ties can be broken by arbitrarily selecting one of them as this does not affect the average probability of error.

More insight can be obtained on the minimum average probability of erroneous list decision given in (2.4) by formulating a new Bayesian $\binom{M}{L}$-ary hypothesis testing problem in which we associate a new hypothesis $\tilde{\mathcal{H}}_j$ corresponding to each list $\ell_j = \{\mathcal{H}_{j_1}, \mathcal{H}_{j_2}, \ldots, \mathcal{H}_{j_L}\}$ for some distinct choice of $1 \leq j_1 < j_2 < \cdots < j_L \leq M$.

Proposition 2.2 *Consider the following Bayesian* $\binom{M}{L}$ *-ary hypothesis testing problem:*

$$\tilde{\mathcal{H}}_j : \; Y \sim \tilde{p}_j(y), \quad \text{for } j = 1, \dots, \binom{M}{L}, \tag{2.11}$$

where the observation $Y \in \mathcal{Y}$ *under hypothesis* $\tilde{\mathcal{H}}_j$ *is distributed with density*

$$\tilde{p}_j(y) = \frac{\sum_{l=1}^{L} \pi_{j_l} p_{j_l}(y)}{\sum_{l=1}^{L} \pi_{j_l}}, \tag{2.12}$$

and the prior probability of the hypothesis $\tilde{\mathcal{H}}_j$ *is*

$$\tilde{\pi}_j = \frac{\sum_{l=1}^{L} \pi_{j_l}}{\binom{M-1}{L-1}}, \tag{2.13}$$

corresponding to distinct choices of $1 \le j_1 < j_2 < \cdots < j_L \le M$ *under each hypothesis* $\tilde{\mathcal{H}}_j$*. The minimum average probability of error of the optimal MAP decision rule* $\tilde{\delta}_{MAP}(y)$ *for the above hypothesis testing problem is related to the minimum average probability of error of the optimal list hypothesis test* $\delta^*(y)$ *by the following identity:*

$$P_e(\delta^*) = 1 - \binom{M-1}{L-1} \times \left(1 - \tilde{P}_e(\tilde{\delta}_{MAP})\right), \tag{2.14}$$

or in terms of the average probability of correct decision:

$$P_c(\delta^*) = \binom{M-1}{L-1} \times \tilde{P}_c(\tilde{\delta}_{MAP}). \tag{2.15}$$

Proof It is straightforward to check that $\tilde{p}_j(y)$'s and $\tilde{\pi}_j$'s are valid probability densities and prior probabilities, respectively. More explicitly, $\tilde{p}_j(y) \ge 0$ for all $y \in \mathcal{Y}$, $\int_{\mathcal{Y}} \tilde{p}_j(y) d\mu(y) = 1$, $\tilde{\pi}_j > 0$, and $\sum_{j=1}^{\binom{M}{L}} \tilde{\pi}_j = 1$. It is well-known that the average probability of error for the above standard $\binom{M}{L}$-ary hypothesis testing problem is minimized by the MAP decision rule, denoted here with $\tilde{\delta}_{MAP}(y)$, which chooses hypothesis $\tilde{\mathcal{H}}_{j^*}$ if

$$j^* = argmax_{1 \le j \le \binom{M}{L}} \; \tilde{\pi}(j|y), \tag{2.16}$$

where $\tilde{\pi}(j|y) = \tilde{\pi}_j \tilde{p}_j(y) / \tilde{p}(y)$ denotes the posterior probability of hypothesis $\tilde{\mathcal{H}}_j$ given the observation y and $\tilde{p}(y) = \sum_{k=1}^{\binom{M}{L}} \tilde{\pi}_k \tilde{p}_k(y) = \sum_{i=1}^{M} \pi_i p_i(y) = p(y)$. Substituting (2.12) and (2.13) into (2.16), it is seen that the posterior probability employed in (2.16) can be expressed as

$$\tilde{\pi}(j|y) = \frac{\sum_{l=1}^{L} \pi(j_l|y)}{\binom{M-1}{L-1}} \text{ for all } 1 \le j \le \binom{M}{L}. \tag{2.17}$$

The corresponding minimum average probability of error is

$$\tilde{P}_e(\tilde{\delta}_{MAP}(y)) = 1 - \int_{\mathcal{Y}} \max_{1 \le j \le \binom{M}{L}} \{\tilde{\pi}_j \tilde{p}_j(y)\} \, d\mu(y). \tag{2.18}$$

Comparing (2.18) with (2.4), the result follows. □

3. Novel lower bounds on the minimum probability of list error

The proposed lower bounds will be established by applying Fano's inequality to the hypothesis testing problem among hypotheses $\tilde{\mathcal{H}}_j$, $j = 1, \ldots, \binom{M}{L}$ described in Proposition 2.2 via (2.11), (2.12), and (2.13). To this end, we first present Fano's inequality and its application to hypothesis testing. Let $X \to Y \to \hat{X}$ denote a Markov chain. In the context of hypothesis testing, X is the unknown hypothesis from some finite set $\mathcal{X}$ with cardinality $|\mathcal{X}|$, Y is the observation generated from this hypothesis, and $\hat{X} : \mathcal{Y} \to \mathcal{X}$ is a possibly randomized hypothesis test that declares a decision for X based on the observation Y. In the following, all logarithms are with respect to base 2.

Lemma 3.1 *Fano's inequality [2, Theorem 2.10.1]: Let X take values from a finite set $\mathcal{X}$, $X \to Y \to \hat{X}$ be a Markov chain, and the probability of error be defined as $\varepsilon = Pr(\hat{X}(Y) \neq X)$. Let the conditional entropy of the underlying hypothesis X given the observation Y be denoted as:*

$$H(X|Y) = -\mathbb{E}[\log p(X|Y)] = -\int_{\mathcal{Y}} p(y) \left(\sum_{x \in \mathcal{X}} p(x|y) \log p(x|y) \right) d\mu(y)\,, \tag{3.1}$$

and the binary entropy function be denoted as $h(p) = -p\log p - (1-p)\log(1-p)$. Then,

$$H(X|Y) \leq h(\varepsilon) + \varepsilon \log(|\mathcal{X}| - 1)\,, \tag{3.2}$$

with equality if and only if

$$p(x|y) = \begin{cases} \frac{\varepsilon}{M-1}, & x \neq \hat{X}(y) \\ 1-\varepsilon, & x = \hat{X}(y)\,. \end{cases} \tag{3.3}$$

Equivalently, we have

$$\varepsilon \geq \frac{H(X|Y) - h(\varepsilon)}{\log(|\mathcal{X}| - 1)}\,. \tag{3.4}$$

Now, assume a uniform distribution on the set of hypotheses, i.e. $p(x) = 1/|\mathcal{X}|$ for all $x \in \mathcal{X}$. The conditional density of Y given $X = x$ is denoted as $p_x(y)$ and the marginal density of the observation Y is given by $p(y) = \sum_{x \in \mathcal{X}} p_x(y)/|\mathcal{X}|$. In the following, the subscript u in X_u indicates that X is uniform. Then,

$$\begin{aligned} H(X_u|Y) &= -\int_{\mathcal{Y}} \sum_{x \in \mathcal{X}} \frac{1}{|\mathcal{X}|} p_x(y) \log \left(\frac{\frac{1}{|\mathcal{X}|} p_x(y)}{p(y)} \right) d\mu(y) \\ &= \log|\mathcal{X}| - \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \int_{\mathcal{Y}} p_x(y) \log\left(\frac{p_x(y)}{p(y)} \right) d\mu(y) \\ &= \log|\mathcal{X}| - \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} D(p_x \,\|\, p)\,, \end{aligned} \tag{3.5}$$

where $D(p_x \,\|\, p) = \int_{\mathcal{Y}} p_x(y) \log(p_x(y)/p(y))\, d\mu(y)$ denotes the relative entropy (or Kullback-Leibler distance) between the densities $p_x(y)$ and $p(y)$. Substituting (3.5) in (3.4), the following lower bound on the average probability of error is obtained:

$$\varepsilon_u = Pr(X_u \neq \hat{X}(Y)) \geq \frac{\log|\mathcal{X}| - \frac{1}{|\mathcal{X}|}\sum_{x \in \mathcal{X}} D(p_x \,\|\, p) - h(\varepsilon_u)}{\log(|\mathcal{X}| - 1)}\,, \tag{3.6}$$

which can be relaxed to [1, Equation 1.2]:

$$\varepsilon_u \geq 1 - \frac{\frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} D(p_x \,\|\, p) + h(\varepsilon_u)}{\log(|\mathcal{X}|-1)}\,. \tag{3.7}$$

Based on (3.7), we arrive at the following bounds for list hypothesis testing.

Proposition 3.2 *The minimum average probability of error of the optimal hypothesis test that selects L candidates out of M equally likely hypotheses is lower bounded as*

$$P_e^u(\delta^*) \geq 1 - \frac{\binom{M-1}{L-1}}{\log(\binom{M}{L}-1)}\left(\frac{1}{\binom{M}{L}}\sum_{j=1}^{\binom{M}{L}} D(\tilde{p}_j \,\|\, p) + h\left(\frac{1-P_e^u(\delta^*)}{\binom{M-1}{L-1}}\right)\right), \tag{3.8}$$

which yields the following relaxed forms:

$$P_e^u(\delta^*) \geq 1 - \frac{\binom{M-1}{L-1}}{\log(\binom{M}{L}-1)}\left(\frac{1}{\binom{M}{L}M}\sum_{j=1}^{\binom{M}{L}}\sum_{i=1}^{M} D(\tilde{p}_j \,\|\, p_i) + h\left(\frac{1-P_e^u(\delta^*)}{\binom{M-1}{L-1}}\right)\right), \tag{3.9}$$

$$P_e^u(\delta^*) \geq 1 - \frac{\binom{M-1}{L-1}}{\log(\binom{M}{L}-1)}\left(\frac{1}{M}\sum_{i=1}^{M} D(p_i \,\|\, p) + h\left(\frac{1-P_e^u(\delta^*)}{\binom{M-1}{L-1}}\right)\right), \tag{3.10}$$

$$P_e^u(\delta^*) \geq 1 - \frac{\binom{M-1}{L-1}}{\log(\binom{M}{L}-1)}\left(\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} D(p_i \,\|\, p_j) + h\left(\frac{1-P_e^u(\delta^*)}{\binom{M-1}{L-1}}\right)\right), \tag{3.11}$$

and lastly

$$P_e^u(\delta^*) \geq 1 - \frac{\binom{M-1}{L-1}}{\log(\binom{M}{L}-1)}\left(\max_{\substack{1\leq i\leq M\\ 1\leq j\leq M\\ i\neq j}} D(p_i \,\|\, p_j) + h\left(\frac{1-P_e^u(\delta^*)}{\binom{M-1}{L-1}}\right)\right). \tag{3.12}$$

Proof Applying the bound in (3.7) to the hypothesis testing problem among hypotheses $\tilde{\mathcal{H}}_j$, $j = 1, \ldots, \binom{M}{L}$ described in Proposition 2.2 via (2.11) and (2.12) with equally likely priors (i.e. $\tilde{\pi}_j = 1/\binom{M}{L}$ corresponding to $\pi_i = 1/M$), we get

$$\tilde{P}_e^u(\tilde{\delta}_{MAP}) \geq 1 - \frac{\frac{1}{\binom{M}{L}}\sum_{j=1}^{\binom{M}{L}} D(\tilde{p}_j \,\|\, p) + h(\tilde{P}_e^u(\tilde{\delta}_{MAP}))}{\log(\binom{M}{L}-1)}\,, \tag{3.13}$$

where superscript u is used to indicate that the error probability is computed under uniform priors. Equivalently using $\tilde{P}_e^u(\tilde{\delta}_{MAP}) = 1 - \tilde{P}_c^u(\tilde{\delta}_{MAP})$ and $h(p) = h(1-p)$:

$$\tilde{P}_c^u(\tilde{\delta}_{MAP}) \leq \frac{\frac{1}{\binom{M}{L}}\sum_{j=1}^{\binom{M}{L}} D(\tilde{p}_j \,\|\, p) + h(\tilde{P}_c^u(\tilde{\delta}_{MAP}))}{\log(\binom{M}{L}-1)}\,. \tag{3.14}$$

From the identity that relates the average probability of correct decision of the optimal list hypothesis test to that of the optimal MAP decision rule given in (2.15), it follows that

$$P_c^u(\delta^*) \leq \frac{\binom{M-1}{L-1}}{\log(\binom{M}{L}-1)} \left(\frac{1}{\binom{M}{L}} \sum_{j=1}^{\binom{M}{L}} D(\tilde{p}_j \,\|\, p) + h\left(\frac{P_c^u(\delta^*)}{\binom{M-1}{L-1}} \right) \right), \tag{3.15}$$

and substituting $P_c^u(\delta^*) = 1 - P_e^u(\delta^*)$, we get the inequality in (3.8). The relaxed versions are obtained by applying Jensen's inequality using the fact that the relative entropy is convex in its arguments, and the last inequality follows by upper bounding the mean pairwise relative entropy with the maximum pairwise relative entropy. □

It is also noted that the average relative entropy term in (3.8) can be expressed as

$$\frac{1}{\binom{M}{L}} \sum_{j=1}^{\binom{M}{L}} D(\tilde{p}_j \,\|\, p) = H(p) - \frac{1}{\binom{M}{L}} \sum_{j=1}^{\binom{M}{L}} H(\tilde{p}_j), \tag{3.16}$$

where $H(p) = -\int_{\mathcal{Y}} p(y) \log p(y) d\mu(y)$ denotes the entropy corresponding to the density $p(y)$ and $H(\tilde{p}_j)$ is defined similarly. Further relaxation of the inequalities given in (3.8)-(3.12) are possible by applying some well-known upper bounds on the binary entropy function $h(\cdot)$ such as $h(p) \leq 2\sqrt{p(1-p)}$ or $h(p) \leq (4p(1-p))^{1/\ln 4}$ [12].

Corollary 3.3 *The lower bounds given in Proposition 3.2 hold for the minimax probability of list error:*

$$P^*_{max} = \inf_{\delta \in \mathcal{D}} \sup_{1 \leq i \leq M} P_{e,i}(\delta(y)), \tag{3.17}$$

where $P_{e,i}(\delta(y))$ *is as defined in* (2.2).

Proof The result follows by observing that

$$P^*_{max} \geq P_e^u(\delta^*) = \inf_{\delta \in \mathcal{D}} \sum_{i=1}^{M} \frac{1}{M} P_{e,i}(\delta(y)). \tag{3.18}$$

□

4. Concluding remarks

Several bounds exist in the literature for the error probability of list decoding [5, 9, 10]. The proposed bounds fall within the class of Fano type bounds among which the most well-known is Fano's inequality for list decoding [4, Lemma 1], [8, Chapter 3.E], [11, Section IV.C]. Let X take values from a finite set $\mathcal{X}$ with $|\mathcal{X}| = M$ and $X \to Y \to \mathcal{L}(Y)$ be a Markov chain. Consider an arbitrary mapping $\mathcal{L}(y) : \mathcal{Y} \to 2^{\mathcal{X}}$ that maps every observation $y \in \mathcal{Y}$ to a list with fixed size $|\mathcal{L}(Y)| = L < M$. Let the probability of list error, i.e. the probability of the event that the list output by the test does not contain the true hypothesis, be defined as $P_e = Pr(X \notin \mathcal{L}(Y))$. Then, Fano's inequality for list decoding states that

$$H(X|Y) \leq h(P_e) + P_e \log\left(\frac{M}{L} - 1\right) + \log L \tag{4.1}$$

with equality if and only if

$$\pi(i|y) = \begin{cases} \frac{P_e}{M-L}, & i \notin \mathcal{L}(y) \\ \frac{1-P_e}{L}, & i \in \mathcal{L}(y)\,. \end{cases} \tag{4.2}$$

Equivalently, we have

$$P_e \geq \frac{H(X|Y) - \log L - h(P_e)}{\log\left(\frac{M}{L} - 1\right)}\,. \tag{4.3}$$

In a similar manner to the analysis above, for equally likely hypotheses, (4.3) can be expressed as

$$P_e^u \geq \frac{\log\left(\frac{M}{L}\right) - \frac{1}{M}\sum_{i=1}^{M} D(p_i \,\|\, p) - h(P_e^u)}{\log(\frac{M}{L} - 1)}\,, \tag{4.4}$$

which can be related to:

$$P_e^u \geq 1 - \frac{\frac{1}{M}\sum_{i=1}^{M} D(p_i \,\|\, p) + h(P_e^u)}{\log(\frac{M}{L} - 1)}\,. \tag{4.5}$$

It is also noted that generalizations of (4.1) and (4.3) to the Arimoto-Rényi conditional entropy are presented in [11, Theorems 8, 9, 10].

Comparing the proposed bound in (3.8) with Fano's inequality in (4.5), it is noted that $\sum_{j=1}^{\binom{M}{L}} D(\tilde{p}_j \,\|\, p)/\binom{M}{L} \leq \sum_{i=1}^{M} D(p_i \,\|\, p)/M$ by Jensen's inequality and this improves the proposed bound compared to Fano's inequality. On the other hand, if the coefficient $\binom{M-1}{L-1}/\log(\binom{M}{L} - 1)$ in (3.8) is much larger than $1/\log(\frac{M}{L} - 1)$, the improvement may fade away, indicating a trade-off between probability mixing and scaling. Nevertheless, it is well known that Fano type bounds are tight (i.e. lower bound is achieved for certain choices of the densities) [6, Theorem 7.5].

As a final remark, we note that by the identity given in (2.14), known bounds on the minimum probability of error of conventional M-ary hypothesis testing can be directly applied to derive similar bounds on the minimum probability of error of list M-ary hypothesis testing as performed in this paper. Another example of this is the bound given in [3, Theorem 1] which readily follows by applying the identity in (2.14) to the meta-converse bound introduced by Polyanskiy, Poor, and Verdu in [7].

References

[1] Birge L. A new lower bound for multiple hypothesis testing. IEEE Transactions on Information Theory 2005; 51 (4): 1611-1615. https://doi.org/10.1109/TIT.2005.844101

[2] Cover TM, Thomas JA. Elements of Information Theory. New York, USA: Wiley, 2006.

[3] Kangarshahi EA, Fabregas i AG. Minimum probability of error of list M-ary hypothesis testing. arXiv 2021.

[4] Kim Y-H, Sutivong A, Cover TM. State amplification. IEEE Transactions on Information Theory 2008; 54 (5): 1850-1859. https://doi.org/10.1109/TIT.2008.920242

[5] Merhav N. List decoding-Random coding exponents and expurgated exponents. IEEE Transactions on Information Theory 2014; 60 (11): 6749-6759. https://doi.org/10.1109/TIT.2014.2351393

[6] Moulin P, Veeravalli VV. Statistical Inference for Engineers and Data Scientists. Cambridge, UK: Cambridge University Press, 2019.

[7] Polyanskiy Y, Poor HV, Verdu S. Channel coding rate in the finite blocklength regime. IEEE Transactions on Information Theory 2010; 56 (5): 2307-2359. https://doi.org/10.1109/TIT.2010.2043769

[8] Raginsky M, Sason I. Concentration of Measure Inequalities in Information Theory, Communications, and Coding. In: Now Foundations and Trends; Second Edition; 2014.

[9] Sakai Y. Generalized Fano-type inequality for countably infinite systems with list-decoding. In: International Symposium on Information Theory and its Applications (ISITA) Proceedings; Singapore; 2018. pp. 727-731.

[10] Sakai Y. Generalizations of Fano's inequality for conditional information measures via majorization theory. Entropy 2020; 22 (3): 288. https://doi.org/10.3390/e22030288

[11] Sason I, Verdu S. Arimoto-Rényi conditional entropy and Bayesian M-ary hypothesis testing. IEEE Transactions on Information Theory 2018; 64 (1): 4-25. https://doi.org/10.1109/TIT.2017.2757496

[12] Topsøe F. Bounds for entropy and divergence for distributions over a two-element set. Journal of Inequalities in Pure & Applied Mathematics 2001; 2 (2): Art. 25.