Information and Inference: A Journal of the IMA (2023) **12**, 1–19
https://doi.org/10.1093/imaiai/iaad001
Advance Access publication on 20 March 2023

Minimum probability of error of list M-ary hypothesis testing

Ehsan Asadi Kangarshahi
Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK

and

Albert Guillén i Fàbregas†
Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK
Department of Information and Communications Technologies, Universitat Pompeu Fabra, C/Roc Boronat 138, 08018 Barcelona, Spain
†Corresponding author: Email: albert.guillen@eng.cam.ac.uk

[Received on 27 October 2021; revised on 1 December 2022; accepted on 24 January 2023]

We study a variation of Bayesian M-ary hypothesis testing in which the test outputs a list of L candidates out of the M possible upon processing the observation. We study the minimum error probability of list hypothesis testing, where an error is defined as the event where the true hypothesis is not in the list output by the test. We derive two exact expressions of the minimum probability or error. The first is expressed as the error probability of a certain non-Bayesian binary hypothesis test and is reminiscent of the meta-converse bound by Polyanskiy, Poor and Verdú (2010). The second, is expressed as the tail probability of the likelihood ratio between the two distributions involved in the aforementioned non-Bayesian binary hypothesis test. Hypothesis testing, error probability, information theory.

Keywords: hypothesis testing; error probability; information theory.

1. Introduction

Statistical hypothesis testing is the problem of deciding on one of M possible statistical hypotheses after processing some observation data modeled by a random variable. Hypothesis testing is one of the main problems in statistics and inference and finds applications in areas such as social, biological, medical, computer sciences, signal processing and information theory. Depending on the subject area and underlying assumptions, it can be referred to as model selection, classification, discrimination or detection. Hypothesis testing problems are typically classified as binary or non-binary, depending on the number of hypotheses, and Bayesian or non-Bayesian, depending on whether or not priors on the hypotheses are known.

The minimum average probability of error of Bayesian binary hypothesis testing is attained by the likelihood ratio test. Similarly, the minimum average error probability of Bayesian M-ary hypothesis testing is attained by the maximum a posteriori (MAP) test [8]. For non-Bayesian binary hypothesis testing, Neyman and Pearson formulated the optimal tradeoff between the pairwise error probabilities and showed that the likelihood ratio test attains the optimum tradeoff [9].

We study a variation of Bayesian M-ary hypothesis testing, where the test outputs a list with L candidate hypotheses. This setting has its roots in list decoding of error-correcting codes for reliable communication [3]. The decoding process of an error-correcting code with M codewords can be naturally cast as an M-ary hypothesis testing problem. Indeed, the intimate connections between information theory and hypothesis testing have been noted over the years. Specifically, hypothesis testing arguments have been key in the derivation of lower bounds to the error probability [2, 5, 10, 14], including bounds that connect the error probability with information measures [4, 7, 11, 12, 15], possibly with a list-decoding option [1, 5, 11, 12]. For general hypothesis testing problems, list hypothesis testing can be helpful when the number of hypotheses is very large and, for complexity reasons, one might wish to implement staggered or iterative testing. At each stage, a bank of tests of smaller dimension is run, but a candidate list is output instead of a single candidate, in order to facilitate information exchange at the next stage or iteration. List hypothesis testing is also implicitly employed in approximate recovery problems related to statistical estimation where a reduction to multiple hypothesis testing is performed (see e.g. [13, Sec. 16.2.2.]). In communications, list detection is employed in large linear multiple-input multiple-output systems iteratively exchanging information with iterative decoders of error correcting codes (see e.g. [16]).

From a theoretical perspective, it is important to understand what is the minimum error probability in order to establish a performance benchmark for practical tests. In this paper, we study the minimum probability of error of list hypothesis testing. We provide two new families of bounds to the minimum probability of error. The first one, bounds the minimum probability of error by that of a suitably optimized non-Bayesian hypothesis test and is reminiscent of the meta-converse bound in [10]. Instead, the second family bounds the minimum probability of error by the tail probability of the likelihood ratio, termed the information spectrum in the information theory literature [6]. When these bounds are optimized over an auxiliary output distribution, inspired by the work in [17], we show that the bounds are actually tight and provide two different expressions of the minimum probability of error. We show that the solution of the optimization of the second bound is unique and provide an expression for the optimal auxiliary distribution. In turn, the identities not only help in better understanding the minimum probability of error, but also help assessing the tightness of the bounds.

This paper is structured as follows. Section 2 introduces the relevant notation for binary hypothesis testing. Section 3 describes the list hypothesis testing problem and derives the minimum probability of error. Section 4 proves the first identity for the minimum probability of error and connects it with non-Bayesian binary hypothesis testing. Section 5 proves the second identity for the minimum probability of error and connects it to the information spectrum. In proving this result, it is shown that the optimal auxiliary distribution is unique. Proofs of auxiliary results can be found the Appendix.

2. Binary hypothesis testing

Let Y be a random variable taking values on set $\mathscr{Y}$. We consider two hypotheses H, 0 and 1, which correspond to Y being distributed according to two distributions, P or Q, respectively. A binary hypothesis test is a probabilistic mapping $\mathscr{Y} \to \{0, 1\}$ that upon observing a certain $y \in \mathscr{Y}$, decides which of the hypothesis represents the observation. We let $\hat{H}$ be the random variable associated with the output of test and T, the test mapping, the random variable associated with the conditional distribution $P_{\hat{H}|Y}$.

The performance of binary hypothesis testing is characterized by the type-0 and type-1 error probabilities, respectively defined as follows,

$$\epsilon_0(T, P) = \sum_y P(y)T(1|y) \tag{2.1}$$

$$\epsilon_1(T, Q) = \sum_y Q(y)T(0|y). \tag{2.2}$$

In the Bayesian setting, given a prior probability $P_H(0), P_H(1)$, the smallest average probability of error is given by

$$\bar{\epsilon} = \min_T \left\{ P_H(0) \cdot \epsilon_0(T,P) + P_H(1) \cdot \epsilon_1(T,P) \right\} \tag{2.3}$$

and the corresponding optimal T is known to be the likelihood ratio test [8]; the likelihood ratio $\frac{P(y)}{Q(y)}$ is checked against the ratio of the priors.

In the non-Bayesian setting, no knowledge about the prior probabilities $P_H(h_i)$, $i = 0, 1$ is assumed. The trade-off between the error probabilities $\epsilon_0(T,P)$ and $\epsilon_1(T,Q)$ is characterized by the function $\alpha_\beta(P,Q)$ defined as follows:

$$\alpha_\beta(P,Q) = \min_{T:\epsilon_1(T,Q)\le\beta} \epsilon_0(T,P). \tag{2.4}$$

Similarly, one can define the alternative tradeoff $\beta_\alpha(P,Q)$ as

$$\beta_\alpha(P,Q) = \min_{T:\epsilon_0(T,P)\le\alpha} \epsilon_1(T,Q). \tag{2.5}$$

It is well known that a minimizing test for (2.4) is the likelihood-ratio threshold test [9]. It is known that every optimal test is a threshold test where the likelihood ratio between the two distributions is compared with a threshold $\lambda_{\mathrm{NP}} \in \mathbb{R}$ such that optimal test can be expressed by the following equation:

$$T_{\mathrm{NP}}(1|y) = \begin{cases} 1 & \frac{P(y)}{Q(y)} > \lambda_{\mathrm{NP}} \\ 0 & \frac{P(y)}{Q(y)} < \lambda_{\mathrm{NP}} \\ \delta_y & \frac{P(y)}{Q(y)} = \lambda_{\mathrm{NP}} \end{cases}, \tag{2.6}$$

where in order to solve (2.4), δ_y and λ_{NP} are chosen such that $\epsilon_1(T,Q) = \beta$. The minimizing test is not unique in general since all values of δ_y and λ_{NP} with the property $\epsilon_1(T,Q) = \beta$ yield an optimal test.

3. List hypothesis testing

Consider now a Bayesian M-ary hypothesis testing problem, with two random variables X, Y jointly distributed according to P_{XY}, such that X, Y take values on $\mathcal{X}, \mathcal{Y}$, respectively with $|\mathcal{X}| = M$. The observation alphabet $\mathcal{Y}$ is a general alphabet that encompasses the Cartesian product of n-observations and many other standard settings. Upon observing $y \in \mathcal{Y}$ we wish to decide what X was. Standard M-ary hypothesis tests output a single candidate hypothesis $\hat{X} \in \{1, \ldots, M\}$. Instead, we consider list hypothesis testing. A list hypothesis test with list size L is a possibly random mapping $P_{\hat{\boldsymbol{X}}|Y}$, where $\hat{\boldsymbol{X}} = (\hat{X}_1, \ldots, \hat{X}_L) \in \mathcal{X}^L$ denotes the random vector containing a list of candidates $\{\hat{X}_1, \ldots, \hat{X}_L\}$. For simplicity of the presentation, we assume that all candidates in the list are distinct; this does not have an effect on the structure of the test that minimizes the probability of error. We define the set of distinct unordered sequences of length L over alphabet $\mathcal{X}$ by $\mathcal{S}_L(\mathcal{X})$. We say that the true hypothesis has been successfully estimated if the true X is one of the entries of the list vector $\hat{\boldsymbol{X}} = (\hat{X}_1, \ldots, \hat{X}_L)$, i.e., if $X \in \{\hat{X}_1, \ldots, \hat{X}_L\}$. The problem is of interest when $L \ll M$.

Since the joint distribution P_{XY} defines a prior distribution P_X over the alternatives, the problem is naturally cast within the Bayesian framework. The average probability of error of a given list hypothesis test $P_{\hat{X}|Y}$, defined as $\bar{\epsilon}(P_{\hat{X}|Y})$, is written as

$$\bar{\epsilon}(P_{\hat{X}|Y}) \triangleq \mathbb{P}\big[X \notin \{\hat{X}_1, \ldots, \hat{X}_L\}\big] \tag{3.1}$$

$$= \mathbb{P}\big[\{\hat{X}_1 \neq X\} \cap \cdots \cap \{\hat{X}_L \neq X\}\big] \tag{3.2}$$

$$= 1 - \mathbb{P}\big[\{\hat{X}_1 = X\} \cup \cdots \cup \{\hat{X}_L = X\}\big] \tag{3.3}$$

$$= 1 - \mathbb{E}\big[\mathbb{1}\big\{\{\hat{X}_1 = X\} \cup \cdots \cup \{\hat{X}_L = X\}\big\}\big] \tag{3.4}$$

$$= 1 - \sum_{\substack{x\in\mathscr{X}, y\in\mathscr{Y} \\ (\hat{x}_1,\ldots,\hat{x}_L)\in\mathscr{S}_L(\mathscr{X})}} P_{XY}(x,y) P_{\hat{X}|Y}(\hat{x}_1,\ldots,\hat{x}_L|y) \mathbb{1}\big\{\{\hat{x}_1 = x\} \cup \cdots \cup \{\hat{x}_L = x\}\big\} \tag{3.5}$$

$$= 1 - \sum_{\substack{y\in\mathscr{Y} \\ (\hat{x}_1,\ldots,\hat{x}_L)\in\mathscr{S}_L(\mathscr{X})}} P_{\hat{X}|Y}(\hat{x}_1,\ldots,\hat{x}_L|y) \mathbb{P}\big[X \in \{\hat{x}_1,\ldots,\hat{x}_L\}, Y = y\big], \tag{3.6}$$

where the probabilities and expectation in (3.1)–(3.4) are computed with respect to the joint distribution between the true hypothesis X, the observation Y and the list $\hat{\boldsymbol{X}}$, and where

$$\mathbb{P}\big[X \in \{\hat{x}_1,\ldots,\hat{x}_L\}, Y = y\big] = \sum_{x\in\mathscr{X}} P_{XY}(x,y) \mathbb{1}\big\{\{\hat{x}_1 = x\} \cup \cdots \cup \{\hat{x}_L = x\}\big\} \tag{3.7}$$

$$= P_{XY}(\hat{x}_1, y) + \cdots + P_{XY}(\hat{x}_L, y). \tag{3.8}$$

Equation (3.8) holds since all elements on the list are assumed to be distinct, and thus, the events $\{\hat{X}_\ell = X\}$ for $\ell = 1,\ldots,L$, are disjoint.

Further define

$$P_{XY}(x_1,\ldots,x_L,y) \triangleq \frac{1}{\binom{M-1}{L-1}} \big(P_{XY}(x_1,y) + \cdots + P_{XY}(x_L,y)\big), \tag{3.9}$$

where $\binom{a}{b} = \frac{a!}{b!(a-b)!}$. Observe that (3.9) is, by assumption, defined only for distinct $x_1,\ldots,x_L \in \mathscr{X}$. In order to show that the above definition induces a probability distribution on $\mathscr{X}^L \times \mathscr{Y}$, we write

$$\sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X}),y} P_{XY}(x_1,\ldots,x_L,y) = \sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X}),y} \frac{1}{\binom{M-1}{L-1}} \big(P_{XY}(x_1,y) + \cdots + P_{XY}(x_L,y)\big) \tag{3.10}$$

$$= \frac{1}{\binom{M-1}{L-1}} \sum_{x,y} \binom{M-1}{L-1} P_{XY}(x,y) \tag{3.11}$$

$$= 1, \tag{3.12}$$

where (3.10) follows from the definition of $P_{XY}(x_1, \ldots, x_L, y)$ in (3.9), (3.11) follows from the fact that for any given $x \in \mathscr{X}$ in the sum (3.11), there are $\binom{M-1}{L-1}$ possible list configurations.

We now turn to the minimum probability of error over all tests, defined as

$$\bar{\epsilon} = \min_{P_{\hat{X}|Y}} \bar{\epsilon}(P_{\hat{X}|Y}). \tag{3.13}$$

The following result finds a test that achieves the minimal probability of error.

LEMMA 3.1. An optimal test achieving the minimal probability of error $\bar{\epsilon}$ chooses distinct $(\hat{x}_1, \ldots, \hat{x}_L) \in \mathscr{X}^L$ such that $P_{XY}(\hat{x}_1, \ldots, \hat{x}_L, y)$ is maximized, yielding

$$\bar{\epsilon} = 1 - \binom{M-1}{L-1} \sum_{y \in \mathscr{Y}} \max_{(\hat{x}_1,\ldots,\hat{x}_L) \in \mathscr{S}_L(\mathscr{X})} P_{XY}(\hat{x}_1, \ldots, \hat{x}_L, y). \tag{3.14}$$

Proof. Any test that maximizes $P_{XY}(\hat{x}_1, \ldots, \hat{x}_L, y)$ will maximize the probability of success and thus minimize the probability of error. Thus, we set

$$P_{\hat{X}|Y}(\hat{x}_1, \ldots, \hat{x}_L|y) = \begin{cases} \frac{1}{|\mathscr{T}(y)|} & (\hat{x}_1, \ldots, \hat{x}_L) \in \mathscr{T}(y) \\ 0 & \text{otherwise} \end{cases} \tag{3.15}$$

where

$$\mathscr{T}(y) = \Big\{(x_1, \ldots, x_L) \in \mathscr{S}_L(\mathscr{X}) \;\; P_{XY}(x_1, \ldots, x_L, y) = \max_{(\hat{x}_1,\ldots,\hat{x}_L) \in \mathscr{S}_L(\mathscr{X})} P_{XY}(\hat{x}_1, \ldots, \hat{x}_L, y)\Big\} \tag{3.16}$$

is the set of list vectors that maximize P_{XY}; there might be more than one maximizing list in which case the specific maximizer does not change the probability of error. With this particular choice, we obtain

$$\bar{\epsilon}(P_{\hat{X}|Y}) = 1 - \sum_{\substack{y \in \mathscr{Y} \\ (\hat{x}_1,\ldots,\hat{x}_L) \in \mathscr{S}_L(\mathscr{X})}} P_{\hat{X}|Y}(\hat{x}_1, \ldots, \hat{x}_L|y)\big(P_{XY}(\hat{x}_1, y) + \cdots + P_{XY}(\hat{x}_L, y)\big) \tag{3.17}$$

$$= 1 - \sum_{y \in \mathscr{Y}} \sum_{\hat{x}_1,\ldots,\hat{x}_L \in \mathscr{T}(y)} \frac{1}{|\mathscr{T}(y)|} \max_{(\hat{x}_1,\ldots,\hat{x}_L) \in \mathscr{S}_L(\mathscr{X})} \big(P_{XY}(\hat{x}_1, y) + \cdots + P_{XY}(\hat{x}_L, y)\big) \tag{3.18}$$

$$= 1 - \sum_{y \in \mathscr{Y}} \max_{(\hat{x}_1,\ldots,\hat{x}_L) \in \mathscr{S}_L(\mathscr{X})} \big(P_{XY}(\hat{x}_1, y) + \cdots + P_{XY}(\hat{x}_L, y)\big) \sum_{(\hat{x}_1,\ldots,\hat{x}_L) \in \mathscr{T}(y)} \frac{1}{|\mathscr{T}(y)|} \tag{3.19}$$

$$= 1 - \sum_{y \in \mathscr{Y}} \max_{(\hat{x}_1,\ldots,\hat{x}_L) \in \mathscr{S}_L(\mathscr{X})} \big(P_{XY}(\hat{x}_1, y) + \cdots + P_{XY}(\hat{x}_L, y)\big), \tag{3.20}$$

where (3.17) is the same as (3.6) using (3.8), (3.18) follows from the definition of the test (3.15). The final result is obtained from definition (3.9). Finally, observe that in order for the optimal test to maximize $\mathbb{P}\big[X \in \{\hat{x}_1, \ldots, \hat{x}_L\}, Y = y\big]$ it is needed that $\hat{x}_\ell$ for $\ell = 1, \ldots, L$ are distinct, since otherwise there would be fewer than L summands in (3.8). □

4. Metaconverse

In [10] Polyanskiy, Poor and Verdú introduced a lower bound to the minimum probability of error of conventional M-ary hypothesis testing. The bound, termed metaconverse bound, is expressed as the error probability of a non-Bayesian binary hypothesis test as

$$\bar{\epsilon} \geq \alpha_{\frac{1}{M}}(P_{XY}, Q_X \times Q_Y), \tag{4.1}$$

where $Q_X(x) = \frac{1}{M}$ for every $x \in \mathscr{X}$ and Q_Y is an arbitrary auxiliary output distribution. It was shown in [17] that optimizing over Q_Y results in the bound being tight thus providing the exact minimum probability of error. In this section, we show a similar family of bounds for list hypothesis testing and provide an identity that connects the minimum error probability of M-ary list hypothesis testing and the error probability of a non-Bayesian binary hypothesis test by means of an optimization over the auxiliary distribution.

First, define an auxiliary probability distribution over the list vector

$$Q_{\boldsymbol{X}}(x_1, \ldots, x_L) \triangleq \begin{cases} \frac{1}{\binom{M}{L}} & \text{for distinct } x_1, \ldots, x_L \in \mathscr{X} \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

where $\boldsymbol{X}$ is a random vector defined on $\mathscr{X}^L$.

The following theorem states the main result of this paper for list hypothesis testing.

THEOREM 4.1. The minimum probability of error $\bar{\epsilon}$ of Bayesian M-ary list hypothesis testing with list size L can be bounded as

$$\frac{1}{\binom{M-1}{L-1}}(1 - \bar{\epsilon}) \leq 1 - \alpha_{\frac{1}{\binom{M}{L}}}(P_{\boldsymbol{X}Y}, Q_{\boldsymbol{X}} \times Q_Y), \tag{4.3}$$

where $P_{\boldsymbol{X}Y}$ and $Q_{\boldsymbol{X}}$ are defined in (3.9) and (4.2), respectively, and Q_Y is an arbitrary distribution over the observation alphabet $\mathscr{Y}$. In addition,

$$\frac{1}{\binom{M-1}{L-1}}(1 - \bar{\epsilon}) = 1 - \max_{Q_Y} \alpha_{\frac{1}{\binom{M}{L}}}(P_{\boldsymbol{X}Y}, Q_{\boldsymbol{X}} \times Q_Y), \tag{4.4}$$

where the following distribution is a maximizer for expression (4.4)

$$Q_Y^*(y) \triangleq \frac{1}{\mu} \max_{(x_1,\ldots,x_L) \in \mathscr{S}_L(\mathscr{X})} P_{\boldsymbol{X}Y}(x_1, \ldots, x_L, y) \tag{4.5}$$

with

$$\mu = \sum_{y' \in \mathscr{Y}} \max_{(x_1,\ldots,x_L) \in \mathscr{S}_L(\mathscr{X})} P_{\boldsymbol{X}Y}(x_1,\ldots,x_L,y') \tag{4.6}$$

being a normalization constant.

Proof. Since $P_{\hat{\boldsymbol{X}}|Y}$ is a probabilistic function for estimating X with a list $\hat{\boldsymbol{X}} \in \mathscr{X}^L$, we have for all $y \in \mathscr{Y}$

$$\sum_{(\hat{x}_1,\ldots,\hat{x}_L) \in \mathscr{S}_L(\mathscr{X})} P_{\hat{\boldsymbol{X}}|Y}(\hat{x}_1,\ldots,\hat{x}_L|y) = 1. \tag{4.7}$$

We proceed by defining a binary hypothesis test T between two distributions on $\mathscr{X}^L \times \mathscr{Y}$ such that

$$\text{hypothesis } 0: \quad (\boldsymbol{X}Y) \sim P_{\boldsymbol{X}Y} \tag{4.8}$$

$$\text{hypothesis } 1: \quad (\boldsymbol{X}Y) \sim Q_{\boldsymbol{X}} \times Q_Y. \tag{4.9}$$

The binary test T is chooses hypothesis 0 as

$$T(0|x_1,\ldots,x_L,y) = P_{\hat{\boldsymbol{X}}|Y}(x_1,\ldots,x_L|y) \tag{4.10}$$

and chooses hypothesis 1 in all other cases. Thus, the error probabilities are given by

$$\epsilon_0(T,P) = 1 - \mathbb{P}\big[\hat{H} = 0|H = 0\big] \tag{4.11}$$

$$= 1 - \sum_{(x_1,\ldots,x_L) \in \mathscr{S}_L(\mathscr{X}), y \in \mathscr{Y}} P_{\boldsymbol{X}Y}(x_1,\ldots,x_L,y) T(0|x_1,\ldots,x_L,y) \tag{4.12}$$

$$= 1 - \sum_{(x_1,\ldots,x_L) \in \mathscr{S}_L(\mathscr{X}), y \in \mathscr{Y}} P_{\boldsymbol{X}Y}(x_1,\ldots,x_L,y) P_{\hat{\boldsymbol{X}}|Y}(x_1,\ldots,x_L|y) \tag{4.13}$$

$$= 1 - \frac{1}{\binom{M-1}{L-1}}(1-\bar{\epsilon}), \tag{4.14}$$

where (4.14) uses (3.6) and definition (3.9). As for the other error probability, we have that

$$\epsilon_1(T,Q) = \mathbb{Q}\big[\hat{H} = 0|H = 1\big] \tag{4.15}$$

$$= \sum_{(x_1,\ldots,x_L) \in \mathscr{S}_L(\mathscr{X}), y \in \mathscr{Y}} Q_{\boldsymbol{X}}(x_1,\ldots,x_L) Q_Y(y) T(0|x_1,\ldots,x_L,y) \tag{4.16}$$

$$= \sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X}),y\in\mathscr{Y}} Q_{\boldsymbol{X}}(x_1,\ldots,x_L)Q_Y(y)P_{\hat{\boldsymbol{X}}|Y}(x_1,\ldots,x_L|y) \tag{4.17}$$

$$= \sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X}),y\in\mathscr{Y}} \frac{1}{\binom{M}{L}} Q_Y(y)P_{\hat{\boldsymbol{X}}|Y}(x_1,\ldots,x_L|y) \tag{4.18}$$

$$= \sum_{y\in\mathscr{Y}} \frac{1}{\binom{M}{L}} Q_Y(y) \sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{\hat{\boldsymbol{X}}|Y}(x_1,\ldots,x_L|y) \tag{4.19}$$

$$= \sum_{y\in\mathscr{Y}} \frac{1}{\binom{M}{L}} Q_Y(y) \tag{4.20}$$

$$= \frac{1}{\binom{M}{L}}, \tag{4.21}$$

where (4.13) and (4.17) follow from the definition of the binary test $T(0|x_1,\ldots,x_L,y)$ and (4.18) from the definition of $Q_{\boldsymbol{X}}$ in (4.2).

Therefore, from the conditions above we can see that for any distribution Q_Y we have

$$\frac{1}{\binom{M-1}{L-1}}(1-\bar{\epsilon}) \le 1 - \alpha_{\frac{1}{\binom{M}{L}}}(P_{\boldsymbol{X}Y}, Q_{\boldsymbol{X}} \times Q_Y), \tag{4.22}$$

since the error probability $\epsilon_0(T,P)$ of the above binary test cannot be lower than the Neyman–Pearson optimal tradeoff (2.4). This proves (4.3). In addition, since Q_Y is arbitrary, this also holds for the maximizing distribution,

$$\frac{1}{\binom{M-1}{L-1}}(1-\bar{\epsilon}) \le 1 - \max_{Q_Y} \alpha_{\frac{1}{\binom{M}{L}}}(P_{\boldsymbol{X}Y}, Q_{\boldsymbol{X}} \times Q_Y). \tag{4.23}$$

In order to prove the tightness of the bound, we now need to show that

$$\frac{1}{\binom{M-1}{L-1}}(1-\bar{\epsilon}) \ge 1 - \max_{Q_Y} \alpha_{\frac{1}{\binom{M}{L}}}(P_{\boldsymbol{X}Y}, Q_{\boldsymbol{X}} \times Q_Y). \tag{4.24}$$

In order to show (4.24) we set $Q_Y = Q_Y^*$ defined in (4.5) and rewrite the α_β function as (see e.g. [18, Ch. 11])

$$\alpha_{\frac{1}{\binom{M}{L}}}(P_{\boldsymbol{X}Y}, Q_{\boldsymbol{X}} \times Q_Y^*) = \sup_{\lambda\ge 0}\left\{\mathbb{P}\left[\frac{P_{\boldsymbol{X}Y}(\boldsymbol{X},Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X})\times Q_Y^*(Y)} \le \lambda\right] + \lambda\mathbb{Q}\left[\frac{P_{\boldsymbol{X}Y}(\boldsymbol{X},Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X})\times Q_Y^*(Y)} > \lambda\right] - \frac{1}{\binom{M}{L}}\lambda\right\}, \tag{4.25}$$

where the first probability is computed with respect to P_{XY} and the second one is computed with respect to $Q_X \times Q_Y$. If we set

$$\hat{\lambda} = \binom{M}{L} \sum_{y \in \mathscr{Y}} \max_{(x_1,\dots,x_L) \in \mathscr{S}_L(\mathscr{S})} P_{XY}(x_1, \dots, x_L, y) \tag{4.26}$$

we find that

$$\mathbb{P}\left[\frac{P_{XY}(\boldsymbol{X}, Y)}{Q_X(\boldsymbol{X}) \times Q_Y^*(Y)} \le \hat{\lambda}\right] + \hat{\lambda}\mathbb{Q}\left[\frac{P_{XY}(\boldsymbol{X}, Y)}{Q_X(\boldsymbol{X}) \times Q_Y^*(Y)} > \hat{\lambda}\right] - \frac{1}{\binom{M}{L}}\hat{\lambda} = 1 - \frac{1}{\binom{M}{L}}\hat{\lambda} \tag{4.27}$$

where (4.27) follows since

$$Q_X(\boldsymbol{X}) Q_Y^*(Y)\hat{\lambda} \tag{4.28}$$

$$= \frac{1}{\binom{M}{L}} \frac{\max_{(x_1,\dots,x_L) \in \mathscr{S}_L(\mathscr{X})} P_{XY}(x_1, \dots, x_L, Y)}{\sum_{y \in \mathscr{Y}} \max_{(x_1,\dots,x_L) \in \mathscr{S}_L(\mathscr{X})} P_{XY}(x_1, \dots, x_L, y)} \binom{M}{L} \sum_{y \in \mathscr{Y}} \max_{(x_1,\dots,x_L) \in \mathscr{S}_L(\mathscr{X})} P_{XY}(x_1, \dots, x_L, y) \tag{4.29}$$

$$= \max_{(x_1,\dots,x_L) \in \mathscr{S}_L(\mathscr{X})} P_{XY}(x_1, \dots, x_L, Y) \tag{4.30}$$

implying that

$$\mathbb{P}\left[\frac{P_{XY}(\boldsymbol{X}, Y)}{Q_X(\boldsymbol{X}) \times Q_Y^*(Y)} \le \hat{\lambda}\right] = \mathbb{P}\left[P_{XY}(\boldsymbol{X}, Y) \le \max_{(x_1,\dots,x_L) \in \mathscr{S}_L(\mathscr{X})} P_{XY}(x_1, \dots, x_L, Y)\right] = 1 \tag{4.31}$$

$$\mathbb{Q}\left[\frac{P_{XY}(\boldsymbol{X}, Y)}{Q_X(\boldsymbol{X}) \times Q_Y^*(Y)} > \hat{\lambda}\right] = \mathbb{Q}\left[P_{XY}(\boldsymbol{X}, Y) > \max_{(x_1,\dots,x_L) \in \mathscr{S}_L(\mathscr{X})} P_{XY}(x_1, \dots, x_L, Y)\right] = 0. \tag{4.32}$$

From Lemma 3.1, we have that

$$\bar{\epsilon} = 1 - \binom{M-1}{L-1} \sum_{y \in \mathscr{Y}} \max_{(\hat{x}_1,\dots,\hat{x}_L) \in \mathscr{S}_L(\mathscr{X})} P_{XY}(\hat{x}_1, \dots, \hat{x}_L, y) \tag{4.33}$$

and thus, using (4.26) and (4.33)

$$1 - \frac{1}{\binom{M}{L}}\hat{\lambda} = 1 - \sum_{y \in \mathscr{Y}} \max_{(x_1,\dots,x_L) \in \mathscr{S}_L(\mathscr{X})} P_{XY}(x_1, \dots, x_L, y) \tag{4.34}$$

$$= 1 - \frac{1}{\binom{M-1}{L-1}}(1 - \bar{\epsilon}), \tag{4.35}$$

which, using (4.25), implies that

$$\alpha_{\frac{1}{\binom{M}{L}}}(P_{XY}, Q_X \times Q_Y^*) = \sup_{\lambda \geq 0} \left\{ \mathbb{P}\left[\frac{P_{XY}(X,Y)}{Q_X(X) \times Q_Y^*(Y)} \leq \lambda \right] + \lambda \mathbb{Q}\left[\frac{P_{XY}(X,Y)}{Q_X(X) \times Q_Y^*(Y)} > \lambda \right] - \frac{1}{\binom{M}{L}}\lambda \right\} \tag{4.36}$$

$$\geq \mathbb{P}\left[\frac{P_{XY}(X,Y)}{Q_X(X) \times Q_Y^*(Y)} \leq \frac{1}{\binom{M}{L}}\hat{\lambda} \right] + \hat{\lambda}\mathbb{Q}\left[\frac{P_{XY}(X,Y)}{Q_X(X) \times Q_Y^*(Y)} > \hat{\lambda} \right] - \hat{\lambda} \tag{4.37}$$

$$= 1 - \frac{1}{\binom{M-1}{L-1}}(1-\bar{\epsilon}), \tag{4.38}$$

where (4.38) uses (4.35). This proves the desired result. □

The identity established by Theorem 4.1 can be rewritten in terms of the alternative pairwise error probability tradeoff.

COROLLARY 4.1. Identity (4.4) can be rewritten as

$$\frac{1}{\binom{M}{L}} = \max_{Q_Y} \beta_{1-\frac{1}{\binom{M-1}{L-1}}(1-\bar{\epsilon})}(P_{XY}, Q_X \times Q_Y). \tag{4.39}$$

The proof of Theorem 4.1 suggests a broad family of lower bounds to the probability of error parametrized by the auxiliary distribution Q_Y. In particular, for a fixed auxiliary distribution Q_Y, we have that

$$\frac{1}{\binom{M-1}{L-1}}(1-\bar{\epsilon}) \leq 1 - \alpha_{\frac{1}{\binom{M}{L}}}(P_{XY}, Q_X \times Q_Y), \tag{4.40}$$

or equivalently,

$$\frac{1}{\binom{M}{L}} \geq \beta_{1-\frac{1}{\binom{M-1}{L-1}}(1-\bar{\epsilon})}(P_{XY}, Q_X \times Q_Y). \tag{4.41}$$

In order to efficiently compute these bounds, one must choose a convenient Q_Y. The specific choice will, naturally, depend on the specifics of the problem at hand. In the case of list decoding of error-correcting codes, these can be useful to derive converse bounds. Consider the transmission of one of M equiprobable messages over a channel described by random transformation $P_{Y|X}$. The encoder maps the message $v \in \{1, \ldots, M\}$ to a codeword $x(v)$ of codebook $\mathscr{C}$. Since there is a codeword for each message, equiprobable messages induce the following channel input distribution

$$P_X^{\mathscr{C}}(x) = \begin{cases} \frac{1}{M} & x \in \mathscr{C} \\ 0 & \text{otherwise.} \end{cases} \tag{4.42}$$

This corresponds to a mass point with probability $\frac{1}{M}$ where codewords are placed and zero otherwise. The decoder runs a list decoder with list size L. The error probability for a fixed codebook $\mathscr{C}$ is denoted by $\bar{\epsilon}(\mathscr{C})$. Therefore, for a fixed code $\mathscr{C}$ and a fixed auxiliary distribution Q_Y we rewrite (4.40) as

$$\frac{1}{\binom{M-1}{L-1}}\big(1-\bar{\epsilon}(\mathscr{C})\big) \le 1-\alpha_{\frac{1}{\binom{M}{L}}}(P^{\mathscr{C}}_{\boldsymbol{X}Y}, Q^{\mathscr{C}}_{\boldsymbol{X}} \times Q_Y), \tag{4.43}$$

where

$$P^{\mathscr{C}}_{\boldsymbol{X}Y}(x_1,\dots,x_L,y) \triangleq \frac{1}{\binom{M-1}{L-1}}\big(P^{\mathscr{C}}_X(x_1)P_{Y|X}(y|x_1)+\cdots+P^{\mathscr{C}}_X(x_L)P_{Y|X}(y|x_L)\big) \tag{4.44}$$

and

$$Q^{\mathscr{C}}_{\boldsymbol{X}}(x_1,\dots,x_L) \triangleq \begin{cases} \frac{1}{\binom{M}{L}} & \text{for distinct} x_1,\dots,x_L \in \mathscr{C} \\ 0 & \text{otherwise.} \end{cases} \tag{4.45}$$

The best bound is found by optimizing (4.43) over the distribution of the code, i.e.,

$$\frac{1}{\binom{M-1}{L-1}}(1-\bar{\epsilon}) = \max_{P^{\mathscr{C}}_X} \frac{1}{\binom{M-1}{L-1}}\big(1-\bar{\epsilon}(\mathscr{C})\big) \tag{4.46}$$

$$\le 1-\min_{P^{\mathscr{C}}_X}\alpha_{\frac{1}{\binom{M}{L}}}(P^{\mathscr{C}}_{\boldsymbol{X}Y}, Q^{\mathscr{C}}_{\boldsymbol{X}} \times Q_Y). \tag{4.47}$$

By optimizing over arbitrary distributions P_X, not necessarily those of the form (4.42), we obtain the metaconverse for list decoding

$$\frac{1}{\binom{M-1}{L-1}}(1-\bar{\epsilon}) \le 1-\min_{P_X}\alpha_{\frac{1}{\binom{M}{L}}}(P_{\boldsymbol{X}Y}, Q_{\boldsymbol{X}} \times Q_Y). \tag{4.48}$$

where now,

$$P_{\boldsymbol{X}Y}(x_1,\dots,x_L,y) \triangleq \frac{1}{\binom{M-1}{L-1}}\big(P_X(x_1)P_{Y|X}(y|x_1)+\cdots+P_X(x_L)P_{Y|X}(y|x_L)\big) \tag{4.49}$$

and $Q_{\boldsymbol{X}}$ is defined in (4.2). Equation (4.48) is the metaconverse bound for list decoding. Observe that for $L = 1$, the above bound recovers the original metaconverse bound for channel coding [10] and can thus be used to prove the converse statement of the channel coding theorem.

5. Information spectrum

In this section, we show an alternative identity for the probability of error of list hypothesis testing. Specifically, this identity is expressed as a function of the tail probability that the likelihood ratio exceeds a certain threshold, sometimes termed information spectrum [6].

THEOREM 5.1. For a fixed auxiliary distribution Q_Y and constant $\lambda \geq 0$, we have that

$$\frac{1}{\binom{M-1}{L-1}}(1-\bar{\epsilon}) \leq 1 - \left\{ \mathbb{P}\left[\frac{P_{XY}(\boldsymbol{X},Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y(Y)} \leq \lambda \right] - \frac{1}{\binom{M}{L}}\lambda \right\}. \tag{5.1}$$

In addition,

$$\frac{1}{\binom{M-1}{L-1}}(1-\bar{\epsilon}) = 1 - \max_{Q_Y} \sup_{\lambda \geq 0} \left\{ \mathbb{P}\left[\frac{P_{XY}(\boldsymbol{X},Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y(Y)} \leq \lambda \right] - \frac{1}{\binom{M}{L}}\lambda \right\} \tag{5.2}$$

$$= 1 - \sup_{\lambda \geq 0} \left\{ \mathbb{P}\left[\frac{P_{XY}(\boldsymbol{X},Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y^*(Y)} \leq \lambda \right] - \frac{1}{\binom{M}{L}}\lambda \right\} \tag{5.3}$$

where Q_Y^* defined in (4.5) is the unique maximizer of (5.2).

Proof. As shown in the proof of Theorem 4.1, we have that,

$$\alpha_{\frac{1}{\binom{M}{L}}}(P_{XY}, Q_{\boldsymbol{X}} \times Q_Y) = \sup_{\lambda \geq 0} \left\{ \mathbb{P}\left[\frac{P_{XY}(\boldsymbol{X},Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y(Y)} \leq \lambda \right] + \lambda \mathbb{Q}\left[\frac{P_{XY}(\boldsymbol{X},Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y(Y)} > \lambda \right] - \frac{1}{\binom{M}{L}}\lambda \right\} \tag{5.4}$$

$$\geq \mathbb{P}\left[\frac{P_{XY}(\boldsymbol{X},Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y(Y)} \leq \lambda \right] + \lambda \mathbb{Q}\left[\frac{P_{XY}(\boldsymbol{X},Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y(Y)} > \lambda \right] - \frac{1}{\binom{M}{L}}\lambda \tag{5.5}$$

$$\geq \mathbb{P}\left[\frac{P_{XY}(\boldsymbol{X},Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y(Y)} \leq \lambda \right] - \frac{1}{\binom{M}{L}}\lambda \tag{5.6}$$

where (5.5) holds for any fixed $\lambda \geq 0$ and (5.6) follows since the second term is always non-negative. Applying this to (4.3), the bound (5.1) follows.

For the particular choice $\hat{\lambda}$ in (4.26) we have,

$$\frac{1}{\binom{M-1}{L-1}}(1-\bar{\epsilon}) = 1 - \sup_{\lambda \geq 0} \left\{ \mathbb{P}\left[\frac{P_{XY}(\boldsymbol{X},Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y^*(Y)} \leq \lambda \right] + \lambda \mathbb{Q}\left[\frac{P_{XY}(\boldsymbol{X},Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y^*(Y)} > \lambda \right] - \frac{1}{\binom{M}{L}}\lambda \right\} \tag{5.7}$$

$$= 1 - \mathbb{P}\left[\frac{P_{XY}(\boldsymbol{X},Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y^*(Y)} \leq \hat{\lambda} \right] - \frac{1}{\binom{M}{L}}\hat{\lambda} \tag{5.8}$$

$$= 1 - 1 + \frac{1}{\binom{M-1}{L-1}}(1 - \bar{\epsilon}) \tag{5.9}$$

$$= \frac{1}{\binom{M-1}{L-1}}(1 - \bar{\epsilon}) \tag{5.10}$$

where (5.8) and (5.9) follow from (4.32) and (4.31), respectively. Equations 5.75.10 imply that $\hat{\lambda}$ in (4.26) is a maximizer of (5.7), and thus,

$$\frac{1}{\binom{M-1}{L-1}}(1 - \bar{\epsilon}) = 1 - \sup_{\lambda \geq 0} \left\{ \mathbb{P}\left[\frac{P_{XY}(\boldsymbol{X}, Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y^*(Y)} \leq \lambda \right] - \frac{1}{\binom{M}{L}}\lambda \right\}. \tag{5.11}$$

We now proceed with the proof that Q_Y^* as defined in (4.5) is the unique maximizer of (5.2). We divide the proof in two parts, depending on whether or not $Q_{\boldsymbol{X}} \times Q_Y^*$ is absolutely continuous with respect to $P_{\boldsymbol{X}Y}$.

$Q_{\boldsymbol{X}} \times Q_Y^*$ is absolutely continuous with respect to $P_{\boldsymbol{X}Y}$

Using (5.4), we rewrite (4.4) as

$$\frac{1}{\binom{M-1}{L-1}}(1 - \bar{\epsilon}) = 1 - \max_{Q_Y} \sup_{\lambda \geq 0} \left\{ \mathbb{P}\left[\frac{P_{XY}(\boldsymbol{X}, Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y(Y)} \leq \lambda \right] + \lambda \mathbb{Q}\left[\frac{P_{XY}(\boldsymbol{X}, Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y(Y)} > \lambda \right] - \frac{1}{\binom{M}{L}}\lambda \right\}. \tag{5.12}$$

The above expression and (5.2) are both exact characterizations of the error probability. However, (5.12) has an additional non-negative term compared to (5.2). Thus, any maximizing distribution and constant Q_Y^* and λ^* of (5.2) are also maximizers of (5.12). As a result, by comparing both equations, we have that

$$\mathbb{Q}\left[\frac{P_{XY}(\boldsymbol{X}, Y)}{Q_{\boldsymbol{X}}(\boldsymbol{X}) \times Q_Y^*(Y)} > \lambda^* \right] = 0. \tag{5.13}$$

Using the definition of $Q_{\boldsymbol{X}}$ in (4.2), and the absolute continuity of $Q_{\boldsymbol{X}} \times Q_Y^*$ with respect to $P_{\boldsymbol{X}Y}$ this implies that

$$P_{XY}(x_1, \ldots, x_L, y) \leq \frac{\lambda^*}{\binom{M}{L}} Q_Y^*(y) \tag{5.14}$$

for all $(x_1, \ldots, x_L) \in \mathscr{S}_L(\mathscr{X})$ and $y \in \mathscr{Y}$. Since this expression holds for arbitrary $(x_1, \ldots, x_L) \in \mathscr{S}_L(\mathscr{X})$, in particular it holds for the maximizing $(x_1, \ldots, x_L) \in \mathscr{S}_L(\mathscr{X})$, yielding

$$\max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{XY}(x_1, \ldots, x_L, y) \leq \frac{\lambda^*}{\binom{M}{L}} Q_Y^*(y). \tag{5.15}$$

Summing over $y \in \mathscr{Y}$ yields

$$\sum_{y\in\mathscr{Y}} \max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{XY}(x_1,\ldots,x_L,y) \le \frac{\lambda^*}{\binom{M}{L}} \sum_{y\in\mathscr{Y}} Q_Y^*(y) \tag{5.16}$$

$$= \frac{\lambda^*}{\binom{M}{L}}, \tag{5.17}$$

where (5.17) follows from the fact that Q_Y^* is a probability distribution.

We have shown that for the maximizing Q_Y^*, λ^* must satisfy (5.17). Therefore, since the first term of (5.2) is increasing with λ, for any λ satisfying (5.17), the smallest λ satisfying (5.17) is the maximizer of (5.2), and thus, equality in (5.17) must hold, i.e.,

$$\sum_{y\in\mathscr{Y}} \max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{XY}(x_1,\ldots,x_L,y) = \frac{\lambda^*}{\binom{M}{L}} \tag{5.18}$$

Substituting λ^* in (5.18) into (5.15) yields

$$\frac{\max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{XY}(x_1,\ldots,x_L,y)}{\sum_{y\in\mathscr{Y}} \max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{XY}(x_1,\ldots,x_L,y)} \le Q_Y^*(y). \tag{5.19}$$

Observe that the left hand side of (5.19) is itself a probability distribution on $\mathscr{Y}$ and thus, (5.19) holds with equality for all $y \in \mathscr{Y}$, recovering (4.5).

<u>$Q_X \times Q_Y^*$ is not absolutely continuous with respect to P_{XY}</u>

Consider a distribution V_Y on $\mathscr{Y}$ and a non-Bayesian binary hypothesis test between P_{XY} and $Q_X \times V_Y$. Then, if there exists some $\hat{y} \in \mathscr{Y}$ such that $V_Y(\hat{y}) = 0$, any optimal test in the Neyman–Pearson setting T is such that

$$T(1|x_1,\ldots,x_L,\hat{y}) \cdot P_{XY}(x_1,\ldots,x_L,\hat{y}) = 0 \tag{5.20}$$

for every $(x_1,\ldots,x_L) \in \mathscr{S}_L(\mathscr{X})$. The interpretation of this statement is that whenever $V_Y(\hat{y}) = 0$, any optimal test would not choose hypothesis 1, unless $P_{XY}(x_1,\ldots,x_L,\hat{y}) = 0$ for all $(x_1,\ldots,x_L) \in \mathscr{S}_L(\mathscr{X})$.

We have the following result, whose proof can be found in Appendix A.

Lemma 5.1 Let Q_Y be a distribution on $\mathscr{Y}$. If there exists a $\bar{y}$ such that

1. $Q_Y(\bar{y}) = 0$
2. $\exists \boldsymbol{x}^1, \boldsymbol{x}^2 \in \mathscr{X}^L$, with $\boldsymbol{x}^i = (x_1^i,\ldots,x_L^i)$ such that $P_{XY}(\boldsymbol{x}^1,\bar{y})P_{XY}(\boldsymbol{x}^2,\bar{y}) > 0$,

then, there exists a distribution $\hat{Q}_Y$ on $\mathscr{Y}$ such that

$$\alpha_{\frac{1}{\binom{M}{L}}}(P_{XY}, Q_X \times Q_Y) < \alpha_{\frac{1}{\binom{M}{L}}}(P_{XY}, Q_X \times \hat{Q}_Y). \tag{5.21}$$

The above Lemma shows that, if there are two (or more) hypotheses for which $P_{XY}(\boldsymbol{x}^1, \bar{y})P_{XY}(\boldsymbol{x}^2, \bar{y}) > 0$, an auxiliary distribution Q_Y that associates zero mass to observation $\bar{y}$ cannot be optimal. In particular, the lemma shows the existence of a distribution that places non-zero mass to all $y \in \mathscr{Y}$ that is better than one that places zero mass at $\bar{y}$, thus bringing us back to the case where $Q_X \times Q_Y^*$ is absolutely continuous with respect to P_{XY}.

There is a remaining trivial case, where there are observations $y \in \mathscr{Y}$ that can only be obtained from only one individual hypothesis. In this case, there is no ambiguity as to what hypothesis caused the observation. Thus, then the problem reduces to removing those observations, i.e., the optimal distribution places zero mass on those and non-zero on the others. □

Data availability statement

No new data were generated or analyzed in support of this research.

Funding

European Research Council (Grant 725411); Spanish Ministry of Economy and Competitiveness (Grant PID2020-116683GB-C22).

References

1. Ahlswede, R., Gács, P. & Körner, J. (1976) Bounds on conditional probabilities with applications in multi-user communication. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **34**, 157–177.
2. Blahut, R. E. (1974) Hypothesis testing and information theory. *IEEE Trans. Inf. Theory*, **20**, 405–417.
3. Elias, P. (1957) List decoding for noisy channels. *Wescon Convention Record, Part 2*. IRE, pp. 95–104.
4. Fano, R. M. (1952) *Class Notes for Course 6.574: Transmission of Information*. Lecture Notes, MIT.
5. Forney, D. G. (1968) Exponential error bounds for erasure, list, and decision feedback schemes. *IEEE Trans. Inf. Theory*, **14**, 206–220.
6. Han, T. S. (2003) *Information-Spectrum Methods in Information Theory*. Springer.
7. Hellman, M. & Raviv, J. (1970) Probability of error, equivocation, and the Chernoff bound. *IEEE Trans. Inf. Theory*, **16**, 368–372.
8. Lehmann, E. L. & Romano, J. P. (2005) *Testing Statistical Hypotheses*. Springer.
9. Neyman, J. & Pearson, E. S. (1933) On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. R. Soc. Lond. Ser A*, **231**, 289–337.
10. Polyanskiy, Y., Poor, H. V. & Verdú, S. (2010) Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory*, **56**, 2307–2359.
11. Sason, I. (2019) On data-processing and majorization inequalities for f-divergences with applications. *Entropy*, **21**, 1022.
12. Sason, I. & Verdú, S. (2017) Arimoto–Rényi conditional entropy and Bayesian *M*-ary hypothesis testing. *IEEE Trans. Inf. Theory*, **64**, 4–25.
13. Scarlett, J. & Cevher, V. (2021) An introductory guide to Fano's inequality with applications in statistical estimation. *Information-Theoretic Methods in Data Science*. (M. R. D. Rodrigues & Y. C. Eldar eds). Cambridge University Press.
14. Shannon, C. E., Gallager, R. G. & Berlekamp, E. R. (1967) Lower bounds to error probability for coding in discrete memoryless channels I-II. *Inform. Control*, **10**, 522–552.
15. Shayevitz, O. (2011) On Rényi measures and hypothesis testing. *IEEE Int. Symp. Inf. Theory*, 894–898.
16. Studer, C. & Bölcskei, H. (2010) Soft–input soft–output single tree-search sphere decoding. *IEEE Trans. Inf. Theory*, **56**, 4827–4842.

17. Vazquez-Vilar, G., Tauste Campo, A., Guillén i Fàbregas, A. & Martinez, A. (2016) Bayesian M-ary hypothesis testing: the meta-converse and Verdú-Han bounds are tight. *IEEE Trans. Inf. Theory*, **62**, 2324–2333.
18. Verdú, S. (draft)(2021) *Information Theory*.

A. Proof of Lemma 5.1

Let (T, λ^*) be an optimal non-Bayesian likelihood-ratio test and the corresponding threshold for testing between $P_{\boldsymbol{X}Y}$ and $Q_{\boldsymbol{X}} \times Q_Y$ with fixed type-1 error probability $\epsilon_1(T, Q_{\boldsymbol{X}} \times Q_Y) = \frac{1}{\binom{M}{L}}$.

Consider the distribution $\hat{Q}_Y$ defined as

$$\hat{Q}_Y(y) = \begin{cases} \frac{\binom{M}{L}}{\mu} \max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{\boldsymbol{X}Y}(x_1,\ldots,x_L,y) & y = \bar{y} \\ \frac{\lambda^*}{\mu} Q_Y(y) & y \neq \bar{y} \end{cases} \tag{A1}$$

where $\mu = \binom{M}{L} \max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{\boldsymbol{X}Y}(x_1,\ldots,x_L,y) + \lambda^*$.

We first show that $\hat{Q}_Y$ is a probability distribution on $\mathscr{Y}$, i.e., that

$$\sum_y \hat{Q}_Y(y) = 1. \tag{A2}$$

We write

$$\sum_y \hat{Q}_Y(y) = \sum_{y\neq\bar{y}} \hat{Q}_Y(y) + \hat{Q}_Y(\bar{y}) \tag{A3}$$

$$= \frac{\lambda^*}{\mu} \sum_{y\neq\bar{y}} Q_Y(y) + \frac{\binom{M}{L}}{\mu} \max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{\boldsymbol{X}Y}(x_1,\ldots,x_L,y) \tag{A4}$$

$$= \frac{1}{\mu}\left(\binom{M}{L} \cdot \max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{\boldsymbol{X}Y}(x_1,\ldots,x_L,y) + \lambda^*\right) \tag{A5}$$

$$= 1, \tag{A6}$$

where (A4) follows from the definition of $\hat{Q}_Y$, (A5) follows from the fact that Q_Y is a probability distribution with $Q_Y(\bar{y}) = 0$, and (A6) follows from the definition of μ.

We now proceed with the proof that for the distribution $\hat{Q}_Y$ in (A1) we have that

$$\alpha_{\frac{1}{\binom{M}{L}}}(P_{\boldsymbol{X}Y}, Q_{\boldsymbol{X}} \times Q_Y) < \alpha_{\frac{1}{\binom{M}{L}}}(P_{\boldsymbol{X}Y}, Q_{\boldsymbol{X}} \times \hat{Q}_Y). \tag{A7}$$

In particular, we construct a binary test $\hat{T}$ to test between $P_{\boldsymbol{X}Y}$ and $Q_{\boldsymbol{X}} \times \hat{Q}_Y$ and show that

$$\epsilon_1(\hat{T}, Q_{\boldsymbol{X}} \times \hat{Q}_Y) = \frac{1}{\binom{M}{L}} \tag{A8}$$

$$\epsilon_0(\hat{T}, P_{\boldsymbol{X}Y}) > \epsilon_0(T, P_{\boldsymbol{X}Y}). \tag{A9}$$

In addition, if we show that the test $\hat{T}$ is an optimal test in the Neyman–Pearson sense, the proof will be complete.

Consider the set defined in (3.16)

$$\mathscr{T}(y) = \left\{ (x_1, \ldots, x_L) \in \mathscr{S}_L(\mathscr{X}) \;\; P_{\boldsymbol{XY}}(x_1, \ldots, x_L, y) = \max_{(\hat{x}_1,\ldots,\hat{x}_L)\in\mathscr{S}_L(\mathscr{X})} P_{\boldsymbol{XY}}(\hat{x}_1, \ldots, \hat{x}_L, y) \right\} \tag{A10}$$

and let

$$\bar{\boldsymbol{x}} = (\bar{x}_1, \ldots, \bar{x}_L) \in \mathscr{T}(y). \tag{A11}$$

We construct the test $\hat{T}$ as follows:

$$\hat{T}(1|x_1, \ldots, x_L, y) = \begin{cases} T(1|x_1, \ldots, x_L, y) & y \neq \bar{y} \\ 1 & y = \bar{y}, (x_1, \ldots, x_L) \neq (\bar{x}_1, \ldots, \bar{x}_L) \\ 0 & y = \bar{y}, (x_1, \ldots, x_L) = (\bar{x}_1, \ldots, \bar{x}_L). \end{cases} \tag{A12}$$

We next validate equality (A8),

$$1 - \epsilon_1(\hat{T}, Q_{\boldsymbol{X}} \times \hat{Q}_Y) = \sum_{y\in\mathscr{Y}} \sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} Q_{\boldsymbol{X}}(x_1, \ldots, x_L)\hat{Q}_Y(y)\hat{T}(1|x_1, \ldots, x_L, y) \tag{A13}$$

$$= \sum_{y\neq\bar{y}} \sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} Q_{\boldsymbol{X}}(x_1, \ldots, x_L)\hat{Q}_Y(y)\hat{T}(1|x_1, \ldots, x_L, y) + \sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} Q_{\boldsymbol{X}}(x_1, \ldots, x_L)\hat{Q}_Y(\bar{y})\hat{T}(1|x_1, \ldots, x_L, \bar{y}) \tag{A14}$$

$$= \sum_{y\neq\bar{y}} \sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} Q_{\boldsymbol{X}}(x_1, \ldots, x_L)\frac{\lambda^*}{\mu}Q_Y(y)\hat{T}(1|x_1, \ldots, x_L, y) + \sum_{\substack{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})\\(x_1,\ldots,x_L)\neq(\bar{x}_1,\ldots,\bar{x}_L)}} Q_{\boldsymbol{X}}(x_1, \ldots, x_L)\frac{\binom{M}{L}}{\mu} \max_{(x'_1,\ldots,x'_L)\in\mathscr{X}^L} P_{\boldsymbol{XY}}(x'_1, \ldots, x'_L, \bar{y})\hat{T}(1|x_1, \ldots, x_L, \bar{y}) \tag{A15}$$

$$= \frac{\lambda^*}{\mu} \sum_{y\neq\bar{y}} \sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} Q_{\boldsymbol{X}}(x_1, \ldots, x_L)Q_Y(y)T(1|x_1, \ldots, x_L, y) + \frac{\binom{M}{L}}{\mu} \max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{\boldsymbol{XY}}(x_1, \ldots, x_L, \bar{y}) \sum_{\substack{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})\\(x_1,\ldots,x_L)\neq(\bar{x}_1,\ldots,\bar{x}_L)}} Q_{\boldsymbol{X}}(x_1, \ldots, x_L) \tag{A16}$$

$$= \frac{\lambda^*}{\mu}\big(1 - \epsilon_1(T, Q_{\boldsymbol{X}} \times \hat{Q}_Y)\big) + \frac{\binom{M}{L}}{\mu} \max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{\boldsymbol{XY}}(x_1, \ldots, x_L, y)\left(1 - \frac{1}{\binom{M}{L}}\right) \tag{A17}$$

$$= \frac{\lambda^*}{\mu}\left(1 - \frac{1}{\binom{M}{L}}\right) + \frac{\binom{M}{L}}{\mu} \max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{XY}(x_1,\ldots,x_L,y)\left(1 - \frac{1}{\binom{M}{L}}\right) \tag{A18}$$

$$= 1 - \frac{1}{\binom{M}{L}}, \tag{A19}$$

where (A15) follows from the definition of the distribution $\hat{Q}_Y$ in (A1), (A16) from the definition of the test $\hat{T}$ in (A12), (A17) follows from the definition of the type-1 probability of error for test T and from the definition of the distribution Q_X in (4.2), (A18) follows from the fact that by the definition of test T, its type-1 error probability is $\frac{1}{\binom{M}{L}}$, and (A19) follows from the definition of the normalization constant μ.

Now, we turn to inequality (A9). We have that

$$\epsilon_0(T, P_{XY}) = \sum_{y\in\mathscr{Y}} \sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{XY}(x_1,\ldots,x_L,y)T(1|x_1,\ldots,x_L,y) \tag{A20}$$

$$= \sum_{y\neq\bar{y}} \sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{XY}(x_1,\ldots,x_L,y)T(1|x_1,\ldots,x_L,y) + \sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{XY}(x_1,\ldots,x_L,\bar{y})T(1|x_1,\ldots,x_L,\bar{y}) \tag{A21}$$

$$< \sum_{y\neq\bar{y}} \sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{XY}(x_1,\ldots,x_L,y)T(1|x_1,\ldots,x_L,y) + \sum_{\substack{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})\\ (x_1,\ldots,x_L)\neq(\bar{x}_1,\ldots,\bar{x}_L)}} P_{XY}(x_1,\ldots,x_L,\bar{y}) \tag{A22}$$

$$= \sum_{y\neq\bar{y}} \sum_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{XY}(x_1,\ldots,x_L,y)\hat{T}(1|x_1,\ldots,x_L,y) + \sum_{\substack{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})\\ (x_1,\ldots,x_L)\neq(\bar{x}_1,\ldots,\bar{x}_L)}} P_{XY}(x_1,\ldots,x_L,\bar{y})\hat{T}(1|x_1,\ldots,x_L,\bar{y}) \tag{A23}$$

$$= \epsilon_0(\hat{T}, P_{XY}) \tag{A24}$$

where (A21) follows from splitting the sum over y in two, (A22) follows from the fact that for $\bar{y}$ the test T is such that $T(1|x_1,\ldots,x_L,\bar{y})P_{XY}(x_1,\ldots,x_L,\bar{y}) = 0$, and from the fact that there are at least two hypotheses for which $P_{XY}(x_1,\ldots,x_L,\bar{y}) \neq 0$ from the statement of the lemma; we thus upper bound it by a non-zero term. Finally, (A22) follows from the definition of $\hat{T}$ in (A12) and (A24) follows from the definition of type-0 probability of error.

We now will be done if we show that the test $\hat{T}$ is an optimal test in the Neyman–Pearson sense. Since we have shown that $\epsilon_1(\hat{T}, Q_X \times \hat{Q}_Y) = \frac{1}{\binom{M}{L}}$, all we need to show is that there exists a threshold for the likelihood ratio λ_{NP} such that (2.6) holds for $\hat{T}$. We will now show that μ is indeed this threshold for test $\hat{T}$. We divide the proof in several cases:

- When $y \neq \bar{y}$, we have that

$$\frac{P_{XY}(x_1,\ldots,x_L,y)}{\mu Q_X(x_1,\ldots,x_L)\hat{Q}_Y(y)} = \frac{P_{XY}(x_1,\ldots,x_L,y)}{\mu Q_X(x_1,\ldots,x_L)\frac{\lambda^*}{\mu}Q_Y(y)} \tag{A25}$$

$$= \frac{P_{XY}(x_1,\ldots,x_L,y)}{\lambda^* Q_X(x_1,\ldots,x_L)Q_Y(y)}, \tag{A26}$$

where (A25) follows from the definition of the distribution $\hat{Q}_Y$ in (A1). Thus, since λ^* was an optimal threshold for test T, μ is an optimal threshold for test $\hat{T}$ in this case.

- When $y = \bar{y}$ and $(x_1,\ldots,x_L) \neq (\bar{x}_1,\ldots,\bar{x}_L)$, according to the definition of $\hat{T}$ in (A12), we have that $\hat{T}(1|x_1,\ldots,x_L,\bar{y}) = 1$. In addition,

$$\mu Q_X(x_1,\ldots,x_L)\hat{Q}(\bar{y}) = \mu Q_X(x_1,\ldots,x_L)\frac{\binom{M}{L}}{\mu}\max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{XY}(x_1,\ldots,x_L,\bar{y}) \tag{A27}$$

$$= \max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{XY}(x_1,\ldots,x_L,\bar{y}) \tag{A28}$$

$$\geq P_{XY}(x_1,\ldots,x_L,\bar{y}), \tag{A29}$$

where (A27) follows from the definition of $\hat{Q}_Y$ in (A1) and (A28) from the definitions of Q_X in (4.2) and of $(\bar{x}_1,\ldots,\bar{x}_L)$ in (A11). This implies that when $y = \bar{y}$ and $(x_1,\ldots,x_L) \neq (\bar{x}_1,\ldots,\bar{x}_L)$,

$$\frac{P_{XY}(x_1,\ldots,x_L,\bar{y})}{Q_X(x_1,\ldots,x_L)\hat{Q}(\bar{y})} \leq \mu. \tag{A30}$$

- When $y = \bar{y}$ and $(x_1,\ldots,x_L) = (\bar{x}_1,\ldots,\bar{x}_L)$, according to the definition of $\hat{T}$ in (A12), we have that $\hat{T}(1|\bar{x}_1,\ldots,\bar{x}_L,\bar{y}) = 0$. In addition,

$$\mu Q_X(\bar{x}_1,\ldots,\bar{x}_L)\hat{Q}(\bar{y}) = \mu Q_X(\bar{x}_1,\ldots,\bar{x}_L)\frac{\binom{M}{L}}{\mu}\max_{(x_1,\ldots,x_L)\in\mathscr{S}_L(\mathscr{X})} P_{XY}(x_1,\ldots,x_L,\bar{y}) \tag{A31}$$

$$= P_{XY}(\bar{x}_1,\ldots,\bar{x}_L,\bar{y}), \tag{A32}$$

where (A31) follows from the definition of $\hat{Q}_Y$ in (A1) and (A32) from the definitions of Q_X in (4.2) and of $(\bar{x}_1,\ldots,\bar{x}_L)$ in (A11). This implies that for $y = \bar{y}$, $(x_1,\ldots,x_L) = (\bar{x}_1,\ldots,\bar{x}_L)$, we have that

$$\frac{P_{XY}(\bar{x}_1,\ldots,\bar{x}_L,\bar{y})}{Q_X(\bar{x}_1,\ldots,\bar{x}_L)\hat{Q}(\bar{y})} = \mu. \tag{A33}$$

As a result, the test $\hat{T}$ is an optimal Neyman–Pearson test according to (2.6) and satisfies

$$\alpha_{\frac{1}{\binom{M}{L}}}(P_{XY}, Q_X \times Q_Y) < \alpha_{\frac{1}{\binom{M}{L}}}(P_{XY}, Q_X \times \hat{Q}_Y). \tag{A34}$$