Shuffled Bits in the Low-Detectability Regime

Stefano Marano
DIEM
University of Salerno
Fisciano (SA), Italy
marano@unisa.it

Peter Willett
ECE Dept.
University of Connecticut
Storrs, CT, USA
peter.willett@uconn.edu

***Abstract*—We consider a decision problem in which data are unordered (unlabeled). Recent studies of this problem provide a complete asymptotic characterization of the decision performance for large data size, which is the solution of a convex optimization problem. While this is fully satisfactory from a numerical viewpoint, limited insight is offered because a closed-form explicit expression for the decision performance is, in general, not available. For binary observations and the challenging regime of low-detectability, we derive an extremely simple analytical solution, investigate its properties and discuss the obtained physical insights.**

***Index Terms*—Unlabeled detection. Unordered data. Permuted observations. Error exponent function.**

I. Introduction

A. Use Cases and Abstraction

1) Social Networks Use Case: An event happens and users of a network take consequent actions, such as visiting specific webpages, contacting friends on social networks, posting comments, and so forth. Users' profiles are known to a network analyzer, meaning that he/she knows the probability of taking each action as consequence of each possible event. The task of the network analyzer is to discover the event but users' actions are anonymized: The network analyzer has access to the actions and knows their probabilities, but cannot make an association between actions and users. Are the users' profiles still useful in this situation? And how to exploit them?

2) Sensor Networks Use Case: A sensor network is engaged in a decision task. The nodes of the network collect independent measurements about a phenomenon of interest and deliver these measurements to a common fusion center, where the decision is taken. But, either because of an external attack or because of inherent system limitations related to the nature and number of sensors, data arrive at the fusion center without identity: they are *unlabeled*. There is no way to relate data to sensors. Even though a complete statistical characterization of the sensors' observations is available at the fusion center, it is by no means obvious how to proceed. How much information is retained within the unlabeled data? What is the performance of an optimal – but label-unaware – decision maker?

3) Abstraction: The previous examples can be abstracted as follows. Suppose there are H possible states of nature, $\mathcal{H}_0, \dots, \mathcal{H}_{H-1}$. A vector $\mathbf{X} = (X_1, \dots, X_n)$ of independent observations is collected, where X_i is a random variable from a finite alphabet $\mathcal{X}$ with known distribution $\mathbb{P}_h(X_i = x)$, $x \in \mathcal{X}$, under state of nature $\mathcal{H}_h$, $h \in \{0, \dots, H-1\}$. We now introduce a change of paradigm with respect to the standard decision problems: *Vector* $\mathbf{X} = (X_1, \dots, X_n)$ is not available, and we only observe the *set* $\{X_1 \dots, X_n\}$, which is an unordered ensemble of values. Two natural questions arise: How can we efficiently process the set for decision making? And: What is the optimal decision performance, with special emphasis on how much can be lost when labels are removed?

B. Related Work

Unlabeled (or unordered) signal processing is an emerging paradigm, suitable to address challenging practical cases in which the processing of a data vector $\mathbf{x}$ takes place without knowing which entries of $\mathbf{x}$ correspond to which locations within the vector. The roots of this paradigm trace back to a standard problem in robotics, but in the signal-processing area the pioneering formulation is due to Unnikrishnan, Haghighatshoar, and Vetterli [1], who addressed the following problem. Let $\mathbf{x} = A\boldsymbol{r}$ be a linear transformation of vector $\boldsymbol{r} \in \Re^k$ by matrix $A \in \Re^{n \times k}$. The task is to recover vector $\boldsymbol{r}$ using the set of values $\{x_1, \dots, x_n\}$ in place of vector $\mathbf{x} = (x_1, \dots, x_n)^T \in \Re^n$. Using a suitable matrix A, it is shown in [1] that recovery is always possible provided that $n \geq 2k$, and is possible by using any subset of size $2k$ of the entries of $\mathbf{x}$. Prompted by these basic theoretical results (no practical algorithms are provided in [1]), a number of recent contributions have appeared, and several extensions of the problem have been investigated; useful entry points are [2]–[6].

In the above references the focus is on signal reconstruction, in the sense that the goal is to recover $\boldsymbol{r}$. Our focus, instead, is on inference problems in which the unknown state of nature rules the statistical distribution of a random vector $\mathbf{X} = (X_1, \dots, X_n)$, and we want to infer such state by observing the unlabeled version of $\mathbf{X}$, which is the set $\{X_1 \dots, X_n\}$. A series of articles address this scenario [7]–[12]. In particular, assuming $H = 2$ possible states of nature, the authors of [11] investigate the fundamental theoretical limits of unlabeled decisions, and suggest practical algorithms for solving the decision problem with affordable computational complexity. In [11] the observation alphabet $\mathcal{X}$ is finite, but otherwise arbitrary. For the special case of binary alphabets, $\mathcal{X} = \{0, 1\}$, the authors of [12] show that the detection algo-

P. Willett was supported by AFOSR under contract FA9500-18-1-0463.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

rithms proposed in [11] reduce to simple forms: the popular GLRT (generalized likelihood ratio test) for some algorithms, and a simple occurrence-counting decision rule for another. The former (GLRT) is easily implementable (the combinatorial matching between observations and distributions can be circumvented), but surprisingly is shown to perform worse than coin-flipping for some problems of practical interest. The latter (counting-based) performs poorly when the long-run average $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{P}_h(X_i=1)$ is almost the same under the two hypotheses $\mathcal{H}_h$, $h=0,1$. An algorithm founded on a CLT (central limit theorem) approximation is suggested in [12] as valid alternative for a large class of practical problems. Thus, we see that [12] elaborates on the "practical" part of [11], in the special case of binary observations. Likewise, the present work elaborates on the "theoretical" part of [11]. The next section describes how we elaborate on [11] and what are the main differences with that article. An extended version of the present work can be found in [13].

C. *Contribution*

For the standard decision problem in which $\mathbf{X}$ is available, it is well-known that the two typical error measures – false alarm and miss probabilities – of an optimal decision maker converge exponentially to zero when the size n of vector $\mathbf{X}$ diverges. The convergence rates cannot be arbitrary: if the false alarm probability converges to zero at rate α, then the miss probability converges to zero at a rate no larger than $\Omega_{\text{lab}}(\alpha)$, where $\Omega_{\text{lab}}(\alpha)$ is a decreasing convex function known as the error exponent [14]. For α arbitrarily small, the error exponent reduces to the celebrated Chernoff-Stein exponent, and enforcing the equal-rate condition $\alpha=\Omega_{\text{lab}}(\alpha)$, the error exponent reduces to the celebrated Chernoff information number, see [15] for the definitions of these quantities. Any of these performance measures represents a fundamental unbeatable limit for the decision system. The main theoretical achievement of [11] is to derive the correspondent error exponent, denoted by $\Omega(\alpha)$, for the case in which only the unlabeled version $\{X_1\dots,X_n\}$ of vector $\mathbf{X}$ is available. Thus, in [11] the asymptotic performance of an optimal decision maker with unlabeled observations is provided. The error exponent function $\Omega(\alpha)$ is given in [11] by an analytical expression, but its form is by no means trivial and therefore insight is not easily found. The theme of the present paper is to explore the functional form of $\Omega(\alpha)$ in the case in which the observation alphabet is binary $\mathcal{X}=\{0,1\}$, with focus on the challenging situation in which the decision problem is "difficult". This qualification means that the data distributions under the two hypotheses are close to each other in some sensible metric or, equivalently, that error probabilities converge to zero exponentially but with small rate. For this scenario we provide an approximate expression for $\Omega(\alpha)$ that is amazingly simple and from which immediate insight can be gained. This represents our main contribution. We reiterate that we only elaborate on the theoretical limit of the decision system, while no attention is paid to practical strategies.

II. Problem Statement

The state of nature is either $\mathcal{H}_0$ or $\mathcal{H}_1$. We adhere to a Neyman-Pearson formulation, in which no a-priori probability is assigned to these states. Expectation and probability operators are denoted by $\mathbb{E}_h$ and $\mathbb{P}_h$, respectively, where the subindex $h=0,1$, specifies the underlying state of nature. We use capital letter to denote random variables and the corresponding lowercase symbol for their realizations. Let $\mathbf{X}=(X_1,\dots,X_n)$ be a vector of independent random variables over the alphabet $\mathcal{X}=\{0,1\}$, with $p_i=\mathbb{E}_1X_i$ and $q_i=\mathbb{E}_0X_i$, $i=1,\dots,n$. When the state of nature is left unspecified, we denote by r_i such expectation, so that $r_i=q_i$ or $r_i=p_i$, respectively, depending on the value of $h=0,1$ in $\mathcal{H}_h$. All throughout this article we assume $r_i\neq 0,1$, $\forall i=1,\dots,n$, which rules out trivialities. A shortcut notation for the sequence $(r_1,\dots,r_n)$ is $r_{1:n}$, and $r_{1:\infty}$ denotes its infinite-sized version. We define $\bar{r}=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}r_i$, with similar definitions for $\bar{q}$ and $\bar{p}$. An alternative notation for long-run averages is $\langle r_{1:\infty}\rangle=\bar{r}$.

The task is to decide between $\mathcal{H}_0$ and $\mathcal{H}_1$. Should the vector $\mathbf{X}$ be observed along with reliable labeling information (i.e., the standard situation), the test would be $\mathcal{H}_h:\ \mathbf{X}\sim\prod_{i=1}^{n}r_i^{x_i}(1-r_i)^{1-x_i}$, for $h=0,1$. However, we do not observe vector $\mathbf{X}=(X_1,\dots,X_n)$, but only its unlabeled entries, namely the *set* of values $\{X_1,\dots,X_n\}$. Equivalently, we observe one of the $n!$ permuted version of $\mathbf{X}$, and we do not know which. Formally, we have:

$$\begin{aligned}\mathcal{H}_1:\ \mathbf{X}&\sim\textstyle\prod_{i=1}^{n}p_i^{x_{\pi(i)}}(1-p_i)^{1-x_{\pi(i)}},\\ \mathcal{H}_0:\ \mathbf{X}&\sim\textstyle\prod_{i=1}^{n}q_i^{x_{\pi(i)}}(1-q_i)^{1-x_{\pi(i)}},\end{aligned}\qquad(1)$$

where $\pi(i)$ represents the unknown permutation. Let $k_{\mathbf{X}}\in\{0,1,\dots,n\}$ be the number of ones appearing in $\mathbf{X}$. Since identical entries of $\mathbf{X}$ are indistinguishable, upon observing the set $\{X_1,\dots,X_n\}$ only $\binom{n}{k_{\mathbf{X}}}$ out of the $n!$ permutations remains distinguishable. Still, estimating the unknown permutation according to a maximum likelihood (ML) principle seems at first glance a combinatorially complex problem, but as pointed out in [12, Prop. 1], this is not so, and in fact a GLRT approach is easily implementable. Now, in many cases of practical interest the GLRT performs very well, and the GLRT is often taken, without special concern, as being as near optimal in a composite testing situation as possible [16]. However – and remarkably – decent performance by the GLRT in the unlabeled case with binary data should not be taken for granted, at least in the asymptotic setting $n\to\infty$. This can be intuitively explained by noticing that as n grows (linearly) more data are available; but at the same time the space over which the maximum is to be found for the ML estimate grows much faster (exponentially). Actually, finding an easily-implementable asymptotically optimum decision maker remains an open problem.

Observing the set $\{X_1,\dots,X_n\}$ is tantamount to observing $k_{\mathbf{X}}/n\in[0,1]$ and, to solve test (1), the interval $[0,1]$ is partitioned in two regions $\mathcal{R}_0$ and $\mathcal{R}_1$ such that $\mathcal{R}_0\cap\mathcal{R}_1=\emptyset$, and $\mathcal{R}_0\cup\mathcal{R}_1=[0,1]$. If $k_{\mathbf{X}}/n\in\mathcal{R}_h$, then the decision is made in favor of $\mathcal{H}_h$, $h = 0, 1$. As performance figures we consider the false alarm probability $\mathbb{P}_0(\mathcal{H}_1)$ and the miss probability $\mathbb{P}_1(\mathcal{H}_0)$, where we used the simplified notation $\mathbb{P}_h(\mathcal{H}_k) = \mathbb{P}_h(k_{\mathbf{X}}/n \in \mathcal{R}_k)$. Since our focus is on the asymptotic setting $n \to \infty$, we consider the limits (if they exist):

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}_0(\mathcal{H}_1), \quad \text{false alarm rate,} \tag{2}$$

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}_1(\mathcal{H}_0), \quad \text{miss rate.} \tag{3}$$

The larger these rates, the better the decision system performs.

III. Results

For $h = 0, 1$, $\lambda \in \Re$, and $\omega \in (0, 1)$, let:

$$\psi_h(\lambda) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \log\left(r_i e^{\lambda} + 1 - r_i\right), \tag{4a}$$

$$\Psi_h(\omega) = \sup_{\lambda \in \Re} \left\{\lambda\, \omega - \psi_h(\lambda)\right\}. \tag{4b}$$

For $\alpha > 0$, let us introduce the *error exponent* function:

$$\Omega(\alpha) = \inf_{\omega\in(0,1):\, \Psi_0(\omega)<\alpha} \Psi_1(\omega). \tag{5}$$

The following result characterizes the asymptotic performance of the unlabeled decision system.

Proposition 1 (Adapted from [11, Th. 2]) *Consider a binary hypothesis test with unlabeled binary data as formalized in (1). For $h = 0, 1$, suppose that $\psi_h(\lambda)$ in* (4a) *is finite and twice continuously differentiable over $\lambda \in \Re$, with derivatives that can be computed term by term under the summation sign. Suppose that the decision region $\mathcal{R}_0$ is closed, and assume, for $\alpha > 0$: $\liminf_{n\to\infty} -\frac{1}{n} \log \mathbb{P}_0(\mathcal{H}_1) \geq \alpha$. Then: (i) for any decision rule*

$$\limsup_{n\to\infty} -\frac{1}{n} \log \mathbb{P}_1(\mathcal{H}_0) \leq \Omega(\alpha), \tag{6}$$

and (ii) there exists a decision rule attaining

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}_1(\mathcal{H}_0) = \Omega(\alpha). \tag{7}$$

Proof: See [11]. ●

Let us introduce a further definition. For $h = 0, 1$, let

$$\bar{\sigma}_h^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} r_i(1 - r_i), \tag{8}$$

and note that $0 < \bar{\sigma}_h^2 \leq 1/4$, where the lower bound follows by the condition $r_i \neq 0, 1$. Also, let $[x]^+ = \max(x, 0)$. The following theorem explores the challenging situation of $\bar{p}$ close to $\bar{q}$, showing that the error exponent $\Omega(\alpha)$ reduces to an elementary, closed-form expression from which insight may be gleaned.

Proposition 2 (Regime of low-detectability) *Suppose that the assumptions of Proposition 1 hold. Suppose also that $\Psi_h(\omega)$ in* (4b) *is twice continuously differentiable over $\omega \in (0, 1)$, and both $\Psi_h(\omega)$ and $\psi_h(\omega)$ have finite third derivative. Then, for sufficiently small $|\bar{p} - \bar{q}|$, the fundamental performance indices can be approximated as follows:*

$$\Omega(\alpha) \approx \Omega_{\rm LD}(\alpha) \quad = \frac{\left(\left[|\bar{p} - \bar{q}| - \sqrt{2\bar{\sigma}_0^2\, \alpha}\,\right]^+\right)^2}{2\bar{\sigma}_1^2}, \tag{9a}$$

$$\text{Chernoff-Stein } S \quad \approx \frac{(\bar{p} - \bar{q})^2}{2\bar{\sigma}_1^2}, \tag{9b}$$

$$\text{Chernoff number } C \approx \frac{(\bar{p} - \bar{q})^2}{2(\bar{\sigma}_1 + \bar{\sigma}_0)^2}. \tag{9c}$$

IV. Derivation of the Formulas in (9)

A. Convex Analysis and Implications

Expressions (9b) and (9c) follow straightforwardly from (9a): the former is $\Omega_{\rm LD}(0)$, and the latter is given by $\Omega_{\rm LD}(\alpha^*)$, where α^* verifies $\alpha^* = \Omega_{\rm LD}(\alpha^*)$. This is because Chernoff-Stein exponent corresponds to the case in which the false alarm probability goes to zero at vanishing rate (actually, is less than some arbitrarily small $\epsilon > 0$), while the Chernoff number corresponds to the case in which the false alarm and the miss probabilities converge to zero at the same exponential rate [15].

Consider hence (9a). For $h = 0, 1$, we expand $\Psi_h(\omega)$ from (4b) in a Taylor series around $\bar{r}$, as follows:

$$\begin{aligned}\Psi_h(\omega) \quad &= \quad \Psi_h(\bar{r}) + \dot{\Psi}_h(\bar{r})(\omega - \bar{r}) + \ddot{\Psi}_h(\bar{r})\frac{(\omega - \bar{r})^2}{2} \\ &\quad\quad + o((\omega - \bar{r})^2).\end{aligned} \tag{10}$$

To deal with the individual summands of (10), we need some facts about convex analysis. Note that $\psi_h(\lambda)$ in (4a) is finite and twice continuously differentiable by assumption, and is strictly convex for $\lambda \in \Re$ because infinite positively-weighted sums of strictly convex functions preserve strict convexity [17]. Recognizing $\Psi(\omega)$ as the Legendre transform of $\psi(\lambda)$, the properties of this latter imply the following [18], where dot denotes derivative:

1) $\Psi_h(\omega)$ is strictly convex (and essentially smooth, see [18]) on $\omega \in (0, 1)$. Recall that $\Psi_h(\omega)$ is twice continuously differentiable by assumption.
2) $\psi_h(\lambda) = \sup_{\omega\in(0,1)} \{\lambda\, \omega - \Psi_h(\omega)\}, \quad \lambda \in \Re$.
3) $\Psi_h(\omega) = \omega\, \dot{\Psi}_h(\omega) - \psi_h(\dot{\Psi}_h(\omega)), \;\; \omega \in (0, 1)$.
4) $\psi_h(\lambda) = \lambda\, \dot{\psi}_h(\lambda) - \Psi_h(\dot{\psi}_h(\lambda)), \quad \lambda \in \Re$.
5) $\dot{\psi}_h(\cdot)$ and $\dot{\Psi}_h(\cdot)$ are inverse functions of each other.
6) The variables in the two domains λ and ω are related by the one-to-one continuous mapping defined by $\omega = \dot{\psi}_h(\lambda)$, $\lambda = \dot{\Psi}_h(\omega)$.
7) When λ and ω are related as in 6), we have $\ddot{\psi}_h(\lambda)\ddot{\Psi}_h(\omega) = 1$ [19].
8) It holds: $\dddot{\Psi}_h(\omega) = -\dddot{\psi}_h(\dot{\psi}_h^{-1}(\omega))/\ddot{\psi}_h^3(\dot{\psi}_h^{-1}(\omega))$. This property can be obtained by differentiating with respect to ω the relationship in 7), and exploiting 6). Recall that the third derivatives of $\psi_h(\cdot)$ and $\Psi_h(\cdot)$ exist by assumption.

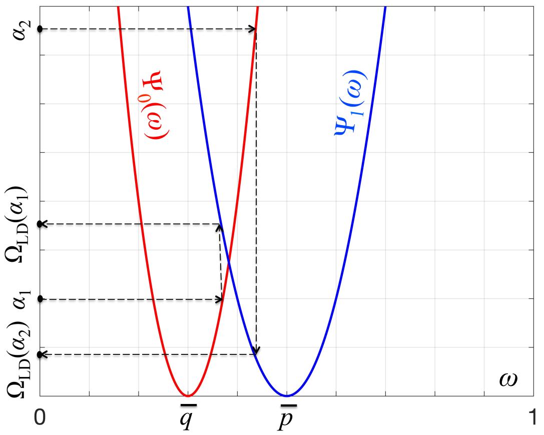

Fig. 1. Graphical construction of $\Omega_{\rm LD}(\alpha)$ from the two convex functions $\Psi_0(\omega) \approx (\omega - \bar{q})^2/(2\bar{\sigma}_0^2)$ and $\Psi_1(\omega) \approx (\omega - \bar{p})^2/(2\bar{\sigma}_1^2)$. Two values of $\alpha = \alpha_{1,2}$ and the corresponding $\Omega_{\rm LD}(\alpha_{1,2})$ are shown.

Coming back to (10), from (4a) we have $\dot{\psi}_h(0) = \bar{r}$, which, accounting for 6), gives $\dot{\Psi}_h(\bar{r}) = 0$. Using the latter into 3), yields $\Psi_h(\bar{r}) = 0$. Exploiting 7) one gets $\ddot{\Psi}_h(\bar{r})\ddot{\psi}_h(0) = 1$, and we also see by direct calculus that $\ddot{\psi}_h(0) = \bar{\sigma}_h^2$. Thus, we arrive at $\Psi_h(\omega) = \frac{(\omega-\bar{r})^2}{2\bar{\sigma}_h^2} + o((\omega - \bar{r})^2)$.

We know that $\Psi_h(\omega)$, for $h = 0, 1$, are two strictly convex twice continuously differentiable functions attaining a unique minimum at $\bar{q}$ and $\bar{p}$, respectively. As a consequence, the values of ω involved in the definition of $\Omega(\alpha)$ in (5) (when this function is not zero) are those lying in the interval with extremes $\bar{p}$ and $\bar{q}$, see Fig. 1 for a graphical representation. For $|\bar{p} - \bar{q}|$ sufficiently small, higher order terms of (10) can be neglected, and $\Psi_h(\omega)$, $h = 0, 1$, are approximated by two parabolas. Then, the approximation of $\Omega(\alpha)$ shown in (9a) follows. With reference to definition (5) and Fig. 1, suppose $\bar{q} < \bar{p}$, and pick $\alpha > 0$. The largest solution of $\alpha = (\omega - \bar{q})^2/(2\bar{\sigma}_0^2)$, is $\omega^* = \bar{q} + \sqrt{2\bar{\sigma}_0^2\alpha}$, which inserted in $(\omega - \bar{p})^2/(2\bar{\sigma}_1^2)$ gives $\left(p - q - \sqrt{2\sigma_0^2\alpha}\right)^2/(2\bar{\sigma}_1^2)$. Considering the opposite case $\bar{p} < \bar{q}$, and handing separately the values of α for which $\Omega(\alpha) = 0$, we arrive at the final approximation (9a).

B. Error analysis

We now consider more in detail the error involved in approximating $\Psi_h(\omega)$ by a parabola. By Taylor's theorem [20, Th. 5.19], we have $\Psi_h(\omega) = \frac{(\omega-\bar{r})^2}{2\bar{\sigma}_h^2} + \dddot{\Psi}_h(\xi_h(\omega, \bar{r}))\frac{(\omega-\bar{r})^3}{6}$, where $\xi_h(\omega, \bar{r})$ is some point lying in the open interval joining ω and $\bar{r}$. A suitable expression for $\dddot{\Psi}_h(\omega)$ is given in property 8) of Sec. IV, and we arrive at:

$$\Psi_h(\omega) = \frac{(\omega - \bar{r})^2}{2\bar{\sigma}_h^2} - \frac{\dddot{\psi}_h(\dot{\psi}_h^{-1}(\xi_h(\omega, \bar{r})))}{\ddot{\psi}_h^3(\dot{\psi}_h^{-1}(\xi_h(\omega, \bar{r})))}\frac{(\omega - \bar{r})^3}{6}. \quad (11)$$

Recall that in the continuous mapping $\dot{\psi}_h^{-1}(\cdot)$, to $\omega = \bar{r}$ it corresponds $\lambda = 0$. Therefore, for $|\bar{p} - \bar{q}|$ sufficiently small, we can make ω close enough to $\bar{r}$, such that the corresponding variable λ is sufficiently close to zero. Then, differentiating under the summation sign the function $\psi_h(\lambda)$ in (4a) and expanding in Taylor series around $\lambda = 0$ the resulting terms, yields $\dot{\psi}_h(\lambda) \approx \bar{r} + \bar{\sigma}_h^2\lambda$. The inverse relationship is $\dot{\psi}_h^{-1}(\omega) \approx (\omega - \bar{r})/\bar{\sigma}_h^2$, which reveals that when ω belongs to the line joining $\bar{q}$ and $\bar{p}$, $\dot{\psi}_h^{-1}(\omega)$ belongs to the line joining 0 and $\pm|\bar{p} - \bar{q}|/\bar{\sigma}_h^2$. Using this fact in the second summand at the right-hand side of (11), we get: For ω lying on the line joining $\bar{q}$ and $\bar{p}$,

$$\epsilon_h = \left|\frac{\dddot{\psi}_h(\dot{\psi}_h^{-1}(\xi_h(\omega, \bar{r})))}{\ddot{\psi}_h^3(\dot{\psi}_h^{-1}(\xi_h(\omega, \bar{r})))}\frac{(\omega - \bar{r})^3}{6}\right| \le \left|\frac{\dddot{\psi}_h(\lambda)}{\ddot{\psi}_h^3(\lambda)}\right| \frac{|\bar{p} - \bar{q}|^3}{6}, \quad (12)$$

where λ lies on the line joining 0 and $\pm|\bar{p} - \bar{q}|/\bar{\sigma}_h^2$. Expression (12) can be exploited for specific observation models, as we show next by two examples.

1) $r_{1:\infty}$ modeled as uniform: Suppose that the entries of $r_{1:\infty}$ are independent realizations of a uniform random variable. Then:

$$\begin{aligned}\ddot{\psi}_h(\lambda) &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{e^\lambda r_i(1 - r_i)}{(r_i e^\lambda + 1 - r_i)^2} \\ &= \int_0^1 \frac{e^\lambda r(1 - r)}{(re^\lambda + 1 - r)^2}\,dr = \frac{e^\lambda(\lambda + 2) + e^{2\lambda}(\lambda - 2)}{(e^\lambda - 1)^3}, \quad (13)\end{aligned}$$

$$\begin{aligned}\dddot{\psi}_h(\lambda) &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{e^\lambda r_i(1 - r_i)(1 - r_i - r_i e^\lambda)}{(r_i e^\lambda + 1 - r_i)^3} \\ &= \int_0^1 \frac{e^\lambda r(1 - r)(1 - r - re^\lambda)}{(re^\lambda + 1 - r)^3}\,dr \\ &= 2e^{2\lambda}\frac{3\sinh(\lambda) - \lambda\cosh(\lambda) - 2\lambda}{(e^\lambda - 1)^4}, \quad (14)\end{aligned}$$

where the expressions with the integrals are obtained by an application of the law of large numbers. Dividing (14) by the third power of (13), expanding the result in Taylor series around $\lambda = 0$, and neglecting the terms in λ^3 gives, $-\dddot{\psi}_h(\lambda)/\ddot{\psi}_h^3(\lambda) \approx \frac{36}{5}\lambda$. Since $\bar{\sigma}_h^2 = 1/6$, this yields:

$$\epsilon_h \le \frac{6}{5}\frac{(\bar{p} - \bar{q})^4}{\bar{\sigma}_h^2} = \frac{36}{5}(\bar{p} - \bar{q})^4, \quad (15)$$

where $\bar{p} \approx \bar{q}$ and one of these values is equal to $1/2$.

2) Half r and half $(1 - r)$: Suppose that half the entries of $r_{1:\infty}$ take value r and the remaining half take value $(1 - r)$, for some $r \in (0, 1)$. Omitting the details for space reasons, in this case we have

$$\epsilon_h \le \left|\frac{r(1 - r) - 1/6}{(r(1 - r))^3}\right| (\bar{p} - \bar{q})^4. \quad (16)$$

V. Discussion & Example

The error exponent function $\Omega_{\rm LD}(\alpha)$ given in (9a) has the following properties. This function depends upon the underlying distribution $r_{1:\infty}$ only through $\bar{r}$ and $\bar{\sigma}_h^2$. Namely, $p_{1:\infty}$ and $q_{1:\infty}$ play their role only through the four quantities $\bar{p}$, $\bar{\sigma}_1^2$, $\bar{q}$, $\bar{\sigma}_0^2$. $\Omega_{\rm LD}(0) = (\bar{p} - \bar{q})^2/(2\bar{\sigma}_1^2)$ (Stein-Chernoff exponent). For $\alpha \in \left(0, (\bar{p} - \bar{q})^2/(2\bar{\sigma}_0^2)\right)$, $\Omega_{\rm LD}(\alpha)$ is strictly decreasing in α, while for $\alpha \ge (\bar{p} - \bar{q})^2/(2\bar{\sigma}_0^2)$ it is identically

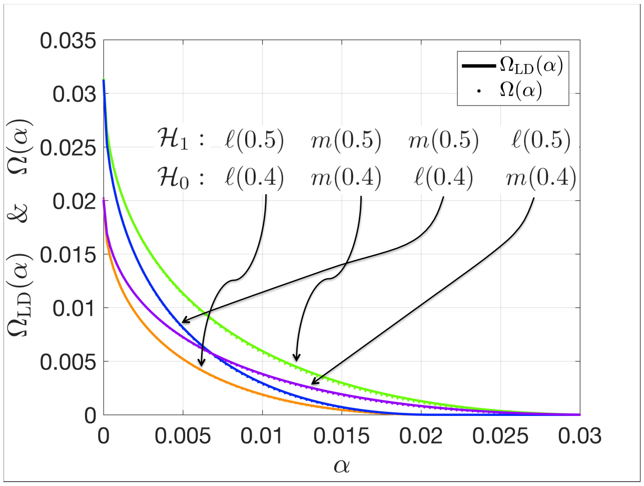

Fig. 2. The error exponent $\Omega(\alpha)$ and its approximation $\Omega_{\rm LD}(\alpha)$ in the low-detectability regime, for the four decision problems described in Sec. V.

zero. The error exponent $\Omega_{\rm LD}(\alpha)$ is infinitely differentiable and in the region where nonzero is strictly convex. Next, let us fix $\bar{r}$.

1) Most Detectable $r_{1:\infty}$*:* The most detectable sequence $r_{1:\infty}$ is the one yielding the lowest value of $\bar{\sigma}_h^2$, which is zero, a case which we have excluded. So, let us consider instead $\bar{\sigma}_h^2 = \gamma(1-\gamma) < \gamma$, for some small $\gamma > 0$. This can be achieved, for instance, by the sequence $(1-\gamma, 1-\gamma, \ldots, \gamma, \gamma, \ldots)$ containing the fraction β of values $(1-\gamma)$ and the fraction $(1-\beta)$ of values γ. For any $\gamma < \min(\bar{r}, 1-\bar{r})$, choosing $\beta = \frac{\bar{r}-\gamma}{1-2\gamma}$ gives $\langle(r_1, r_2, \ldots)\rangle = \bar{r}$.

2) Least Detectable $r_{1:\infty}$*:* To the other extreme, the least detectable sequence $r_{1:\infty}$ is that with larger $\bar{\sigma}_h^2$, compatible with the prescribed $\bar{r}$. It can be seen (not shown for space reasons) that this maximum is obtained when $r_{1:\infty} = (\bar{r}, \bar{r}, \ldots)$, which of course implies $\bar{\sigma}_h^2 = \bar{r}(1-\bar{r})$. Thus, the least detectable case is when data are iid.

A. *Example*

We conclude with an example. Let $m(\bar{p})$ and $m(\bar{q})$ denote, for short, the most detectable sequences for prescribed values of $\bar{p}$ and $\bar{q}$, respectively. Likewise, let $\ell(\bar{p})$ and $\ell(\bar{q})$ denote the least detectable sequences, with γ also assigned. In Fig. 2 we set $\bar{p} = 0.5$, $\bar{q} = 0.4$, $\gamma = 0.2$, and consider four decision problems, as follows:

case no.	1	2	3	4
$\mathcal{H}_1:$	$\ell(0.5)$	$m(0.5)$	$m(0.5)$	$\ell(0.5)$
$\mathcal{H}_0:$	$\ell(0.4)$	$m(0.4)$	$\ell(0.4)$	$m(0.4)$

The figure shows $\Omega_{\rm LD}(\alpha)$ (solid lines) and $\Omega(\alpha)$ (dots). Note the accuracy of the approximation $\Omega(\alpha) \approx \Omega_{\rm LD}(\alpha)$.

VI. Conclusions

The error exponent for unlabeled detection $\Omega(\alpha)$ provides a complete characterization of the asymptotic decision performance in the setting in which the data vector is observed after an unknown shuffling of its entries. Unfortunately, the analytical form of $\Omega(\alpha)$ is involved and provides limited insight. In the challenging situation of low detectability, namely, when $|\bar{p}-\bar{q}|$ is small, we show, by standard series expansions, that the error exponent can be approximated by $\Omega_{\rm LD}(\alpha)$, which is given in elementary form, and from which it is easy to understand how the system parameters (underlying data distributions) determine the decision performance. Our characterization of the error exponent provides easy-to-read answers to the questions posed in the introduction by quantifying the information retained within the unlabeled data.

References

[1] J. Unnikrishnan, S. Haghighatshoar, and M. Vetterli, “Unlabeled sensing with random linear measurements,” *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3237–3253, May 2018.

[2] A. Pananjady, M. J. Wainwright, and T. A. Courtade, “Linear regression with shuffled data: Statistical and computational limits of permutation recovery,” *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3286–3300, May 2018.

[3] S. Haghighatshoar and G. Caire, “Signal recovery from unlabeled samples,” *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1242–1257, March 2018.

[4] A. Abid, A. Poon, and J. Zou. (2017, May 4) Linear regression with shuffled labels. [Online]. Available: http://arxiv.org/abs/1705.01342

[5] G. Elhami, A. Scholefield, B. B. Haro, and M. Vetterli, “Unlabeled sensing: Reconstruction algorithm and theoretical guarantees,” in *Proc. of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017)*, New Orleans, USA, March 5-9, 2017.

[6] A. Pananjady, M. J. Wainwright, and T. A. Courtade. (2017, April 24) Denoising linear models with permutated data. [Online]. Available: http://arxiv.org/abs/1704.07461

[7] Z. Liu and J. Zhu, “Signal detection from unlabeled ordered samples,” *IEEE Communications Letters*, vol. 22, no. 12, pp. 2431–2434, Dec. 2018.

[8] G. Wang, J. Zhu, R. S. Blum, P. Willett, S. Marano, V. Matta, and P. Braca, “Signal amplitude estimation and detection from unlabeled binary quantized samples,” *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4291–4303, Aug. 2018.

[9] J. Zhu, H. Cao, C. Song, and Z. Xu, “Parameter estimation via unlabeled sensing using distributed sensors,” *IEEE Communications Letters*, vol. 21, no. 10, pp. 2130–2133, Oct 2017.

[10] S. Marano, V. Matta, P. Willett, P. Braca, and R. Blum, “Hypothesis testing in the presence of Maxwell’s daemon: Signal detection by unlabeled observations,” in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017)*, New Orleans, LA, USA, 5-9 Mar. 2017.

[11] S. Marano and P. Willett, “Algorithms and fundamental limits for unlabeled detection using types,” *IEEE Transactions on Signal Processing*, vol. 67, no. 8, pp. 2022–2035, Apr. 2019.

[12] ——, “Making decisions with shuffled bits,” in *Proc. of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019)*, Brighton, UK, 12–17 May, 2019.

[13] S. Marano and P. Willett, “Making Decisions by Unlabeled Bits,” submitted to *IEEE Transactions on Signal Processing*, May 2019.

[14] R. Blahut, *Principles and Practice of Information Theory*. Addison-Wesley, 1987.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New Jersey, USA: Wiley-Interscience, 2006.

[16] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. Englewood Cliffs, New Jersey: Prentice Hall, 1998.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[18] R. T. Rockafellar, *Convex analysis*. Princeton, NJ: Princeton University Press, 1970.

[19] R. K. P. Zia, E. F. Redish, and S. R. McKay, “Making sense of the Legendre transform,” *American Journal of Physics*, vol. 77, no. 7, pp. 614–622, Jul. 2009.

[20] T. M. Apostol, *Mathematical Analysis*, 2nd ed. Reading, Massachusetts: Addison-Wesley Publishing Company, 1974.