6.7800 Spring 2025: Bonus Lecture 02

Group Testing and Information-Theoretic Lower Bound

Jongha (Jon) Ryu
jongha@mit.edu

March 14, 2025

Contents

1 Introduction

There are p individuals who might have been infected with a certain disease. We wish to identify the infected individuals through blood tests. Naively, one may consider an exhaustive search, testing each individual separately; however, this is clearly very costly. If we believe that the number of infected individuals is very small, we can expect that exhaustive testing would be a waste of resources.

Instead, we can try a *group testing* approach: rather than testing individuals one by one, we can mix blood samples from multiple individuals and test the pooled sample. Each test yields a positive result if at least one individual in the group is infected (though it does not reveal which one(s) are infected), and a negative result otherwise (i.e., everyone in the group is healthy). The key question is: with this group testing approach, can we reduce the number of tests? Moreover, how can we determine whether a given testing scheme is *optimal*?

To motivate the question of a lower bound, let's consider a concrete testing scheme in the simplest possible setting: suppose that we know *a priori* that exactly one individual is infected, and that the blood test is always accurate. If we assume that the infected individual is chosen uniformly at random, then an exhaustive search would require, on average, $p/2$ tests, or at most $p - 1$ tests in the worst case. With group testing, however, we can identify the infected individual using binary search, requiring only $\log_2 p$ tests. More generally, if there are $k \geq 1$ infected individuals, we might

expect that around $k \log_2 p$ tests would suffice. (We will not explore this question further here, but it makes for an interesting puzzle to ponder!)

The focus of this session is to establish a *lower bound* on the number of tests required. How can we show that our testing scheme is efficient enough? Below, we will learn how information-theoretic measures such as entropy and mutual information, along with related inequalities from class, can help us answer this question.

2 Fano's Inequality

We start from the famous inequality of Robert Fano.[1][2] Throughout the note, we will assume log base 2.

Consider a discrete random variable X supported over a finite set $\mathcal{X}$, which we treat as a target quantity we wish to estimate. Given a noisy observation Y of X, we construct an estimator $\hat{X}$. Formally, the Markov chain $X - Y - \hat{X}$ captures the relationship between these three random variables. Let $P_e := \mathsf{P}\{\hat{X} \neq X\}$ be the probability of wrong estimation. The following inequality captures the relation between P_e and the conditional entropy $H(X|Y)$.

Theorem 1 (Fano's inequality)**.**

$$P_e \geq \frac{H(X|Y) - 1}{\log |\mathcal{X}|}.$$

In words, roughly, this relation shows that if the conditional entropy $H(X|Y)$ is large, then the detection error probability cannot be small. In an extreme case, if we wish to ensure $P_e = 0$, then it asserts that we must have $H(X|Y) \leq 1$. If X is drawn uniformly at random from $\mathcal{X}$, we can rewrite it as

$$P_e \geq 1 - \frac{I(X;Y) + 1}{\log |\mathcal{X}|}.$$

Proof. Let $H(q) := q \log \frac{1}{q} + (1-q) \log \frac{1}{1-q}$ denote the binary entropy function for $q \in [0,1]$. We claim that

$$H(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(X|\hat{X}) \geq H(X|Y). \tag{1}$$

Define an indicator random variable $E := \mathbb{1}_{\{\hat{X} \neq X\}}$. We will upper and lower bound $H(X, E|\hat{X})$ by applying the chain rules in two different ways.

$$\begin{aligned} H(X, E|\hat{X}) &\overset{(a)}{=} H(X|\hat{X}) + H(E|X, \hat{X}) \\ &\overset{(b)}{=} H(E|\hat{X}) + H(X|E, \hat{X}). \end{aligned}$$

First, from (a), since $H(E|X, \hat{X}) = 0$, we have

$$H(X, E|\hat{X}) = H(X|\hat{X}) \geq H(X|Y),$$

[1] You can also find the treatment from Practice Problem 5.3.

[2] Interestingly, R. Fano derived it for his course at MIT in early 1950s, and later published in his lecture note [3].

where the inequality follows from the data processing inequality $I(X;Y) \geq I(X;\hat{X})$. Now, we upper bound using (b). Since conditioning reduces entropy $H(E|\hat{X}) \leq H(E)$. For $H(X|E,\hat{X})$, note that

$$\begin{aligned}
H(X|E,\hat{X}) &= \sum_{\hat{x}} \sum_{e\in\{0,1\}} \mathsf{P}(E=e,\hat{X}=\hat{x}) H(X|E=e,\hat{X}=\hat{x}) \\
&= \sum_{\hat{x}} \Big\{ \mathsf{P}(E=0,\hat{X}=\hat{x}) H(X|E=0,\hat{X}=\hat{x}) + \mathsf{P}(E=1,\hat{X}=\hat{x}) H(X|E=1,\hat{X}=\hat{x}) \Big\} \\
&\overset{(c)}{=} \mathsf{P}(E=1) \sum_{\hat{x}} \mathsf{P}(\hat{X}=\hat{x}|E=1) H(X|E=1,\hat{X}=\hat{x}) \\
&\overset{(d)}{\leq} P_e \log(|\mathcal{X}|-1).
\end{aligned}$$

In (c), we invoke that $H(X|E=0,\hat{X}=\hat{x}) = H(X|\hat{X}=X,\hat{X}=\hat{x}) = 0$, and (d) follows since $H(X|E=1,\hat{X}=\hat{x}) = H(X|\hat{X}\neq X,\hat{X}=\hat{x}) \leq \log(|\mathcal{X}|-1)$. Combining the inequalities proves the claim. We can get the final inequality in the statement by noting that $H(P_e) \leq 1$ and rearranging terms. □

We note that Fano's inequality is tight. To see this, suppose that there is no observation (i.e., no Y). Then, the optimal guess is $x^* = \arg\max_x p(x)$ and the probability of error is $P_e = 1 - p(x^*)$. The tight version of Fano's inequality in Eq. (1) becomes

$$H(P_e) + P_e \log(|\mathcal{X}|-1) \geq H(X).$$

It is easy to check that the equality is achieved if the pmf of X has the following form:

$$\Big(1-P_e, \frac{P_e}{|\mathcal{X}|-1}, \ldots, \frac{P_e}{|\mathcal{X}|-1}\Big).$$

3 Application: Lower Bound on Group Testing

Let S denote the set of (indicies of) infected individuals and we know that $|S| = k$ for simplicity. We assume S is uniformly at random over all possible $\binom{p}{k}$ configurations. Our group testing scheme can be described as a sequence of *test vectors* $\mathbf{X}_i \in \{0,1\}^p$, where there are n test vectors $\mathbf{X}_{1:n}$ and each records if the j-th individual was tested ($X_{ij} = 1$) or not ($X_{ij} = 0$). (We assume the *non-adaptive* setting, where we choose the design matrix a priori all the tests. That is, we assume that $\mathbf{X}_i$'s are independent and $\mathbf{X}_i$ is independent to $Y_{1:i-1}$.) If the blood test were perfect, we can write down the outcome of the i-th test as

$$U_i := \bigvee_{j\in S} X_{ij},$$

where $\bigvee$ denotes the OR operation. We consider a slightly more challenging case, where we model that a noisy testing result. That is,

$$Y_i := U_i \oplus Z_i = \bigvee_{j\in S} X_{ij} \oplus Z_i,$$

where Z_i's are independent Bernoulli random variables with probability $\varepsilon \geq 0$ and $\oplus$ denotes the XOR operation. Based on the test results $Y_{1:n} := (Y_1, \ldots, Y_n)$, we construct an estimator $\hat{S}$, and hence the (conditional) Markov chain

$$(S \to Y_{1:n} \to \hat{S}) | \mathbf{X}_{1:n}. \tag{2}$$

By applying Fano's inequality, we can show:

Theorem 2. *If $P_e \leq \delta$, then*

$$n \geq \frac{(1-\delta)\log\binom{p}{k} - 1}{1 - H(\varepsilon)}.$$

If $\delta = 0$ (exact recovery) and $\varepsilon = 0$ (perfect test), this boils down to $n \geq \log\binom{p}{k} - 1$, which matches to our prior guess.

Proof. Applying Fano's inequality to the conditional Markov chain in Eq. (2), we have

$$\delta \geq P_e \geq 1 - \frac{I(S; Y_{1:n} | \mathbf{X}_{1:n}) + 1}{\log\binom{p}{k}},$$

which leads to

$$I(S; Y_{1:n} | \mathbf{X}_{1:n}) \geq (1-\delta)\log\binom{p}{k} - 1. \tag{3}$$

Note that $\mathbf{X}_i$'s are independent, and also Y_i's are independent given $\mathbf{X}_{1:n}$. Then, we have

$$\begin{aligned}
I(S; Y_{1:n} | \mathbf{X}_{1:n}) &= \sum_{i=1}^{n} I(S; Y_i | \mathbf{X}_{1:n}, Y_{1:i-1}) \\
&= \sum_{i=1}^{n} I(S; Y_i | \mathbf{X}_{1:n}) && (\because Y_i \perp\!\!\!\perp Y_{i-1} | \mathbf{X}_{1:n}) \\
&= \sum_{i=1}^{n} I(S; Y_i | X_i). && (\because Y_i \text{ depends on } X_{1:n} \text{ only through } X_i)
\end{aligned}$$

Here,

$$\begin{aligned}
I(S; Y_i | X_i) &= H(Y_i | X_i) - H(Y_i | X_i, S) \\
&\leq H(Y_i) - H(Y_i | U_i) && (\because \text{conditioning reduces entropy \& DPI}) \\
&= I(Y_i; U_i) \\
&= I(U_i; U_i \oplus Z_i) \\
&\leq 1 - H(\varepsilon).
\end{aligned}$$

The last inequality follows from the following fact: for a Bernoulli random variable Z with probability ε, for any binary random variable U,

$$I(U; U \oplus Z) \leq 1 - H(\varepsilon).$$

Finally,

$$I(S; Y_{1:n} | \mathbf{X}_{1;n}) \leq n(1 - H(\varepsilon)),$$

and applying this inequality to Eq. (3) and dividing it by $1 - H(\varepsilon)$ proves the lower bound. □

4 Remarks and Further Readings

This puzzle was brought up by Xander Morgan (the wording is from this blog post):

> The King of a small country invites 1000 senators to his annual party. As a tradition, each senator brings the King a bottle of wine. Soon after, the Queen discovers that one of the senators is trying to assassinate the King by giving him a bottle of poisoned wine. Unfortunately, they do not know which senator, nor which bottle of wine is poisoned, and the poison is completely indiscernible. However, the King has 10 prisoners he plans to execute. He decides to use them as taste testers to determine which bottle of wine contains the poison. The poison when taken has no effect on the prisoner until exactly 24 hours later when the infected prisoner suddenly dies. The King needs to determine which bottle of wine is poisoned by tomorrow so that the festivities can continue as planned. Hence he only has time for one round of testing. How can the King administer the wine to the prisoners to ensure that 24 hours from now he is guaranteed to have found the poisoned wine bottle?

This time-constraint enforces the non-adaptive group testing as we considered, and we have the lower bound is $\log_2 1000 - 1 \approx 9$ from Fano's inequality. This can be achieved by the following binary encoding scheme: encode each jug by 10-bits of binary representation, 10 prisoners are assigned to the 10 bits, and each prisoner drinks a sip from the bottle whenever the corresponding bottle has 1 in the representation. After 24 hours, we can recover the poisonous bottle from the dead prisoners.

The idea of group testing was pioneered by R. Dorfman [2] to efficiently test for syphilis during World War II. Group testing remains relevant in the modern era, with a notable example being its application in COVID testing.

Fano's inequality plays an important role in establishing lower bounds in statistical estimation, extending beyond the discrete setup discussed in this note. We refer the interested reader to the excellent overview tutorial [4], which highlights the applicability of Fano's inequality in the literature. The group testing example is also drawn from this tutorial, which further discusses its extensions to approximate recovery and adaptive settings. A recent monograph [1] provides an information-theoretic perspective on group testing.

References

[1] Matthew Aldridge, Oliver Johnson, and Jonathan Scarlett. Group testing: an information theory perspective. *Found. Trends Commun. Inf. Theory.*, 15(3-4):196–392, 2019.

[2] Robert Dorfman. The detection of defective members of large populations. *Ann. Math. Stat.*, 14 (4):436–440, December 1943. ISSN 0003-4851,2168-8990. doi: 10.1214/aoms/1177731363.

[3] Robert Fano. *Transmission of Information: A statistical theory of communications*. MIT Press, 1966.

[4] Jonathan Scarlett and Volkan Cevher. An Introductory Guide to Fano's Inequality with Applications in Statistical Estimation. *arXiv:1901.00555*, January 2019.