AN IMPROVEMENT OVER THE INFORMATION LOWER BOUND IN
BINOMIAL GROUP TESTING

BY

Y. C. YAO
COLORADO STATE UNIVERSITY
AND
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

Abstract

The group testing problem for a binomial set of items is considered. It is desired to classify all items as good or defective with a minimum expected number of group tests. Via a weak concavity property, an improvement over the information lower bound is made for the minimum expected number of group tests.

Running heading. Group Testing.

Key Words. Group testing, information lower bound, concavity.

1. Introduction.

We consider the group testing problem for a set of independent items $I = \{I_1, \dots, I_n\}$ where I_i has probability p_i of being defective and probability $q_i = 1 - p_i$ of being good. It is desired to identify the states (good or defective) of all the items with a minimum expected number of group tests. Here, a group test is a test on a subset S of I with two possible outcomes: either S is "pure" (i.e. contains no defectives) or S is "contaminated" (i.e. contains at least one defective, but with no information about which or how many). Sobel and Groll (1959) illustrated many applications of group testing. Recently Berger, Mehravari, Towsley and Wolf (1984) applied group testing procedures to conflict resolution in multiple access channel communication.

Denote by $E(p_1, \dots, p_n)$ the minimum expected number of group tests required to identify the states of the I_i , $1 \leq i \leq n$. In case that all the p_i are equal to p , we write $E(p_1, \dots, p_n)$ as $E(n * p)$. It is well known (Sobel and Groll (1959), Unger (1960)) that

$$n^{-1} E(n * p) \geq g(p), \quad 0 < p < 1, \quad (1)$$

and

$$n^{-1} E(n * p) = 1, \quad p_1 \leq p < 1, \quad (2)$$

where $p_1 = (3 - \sqrt{5})/2$ and $g(p) = -p \log_2 p - q \log_2 q$ is the information lower bound ($q = 1 - p$). Since no optimal procedure for $p < (3 - \sqrt{5})/2$ is known in general, a sharp lower bound for $n^{-1} E(n * p)$ is useful. Let

$$f(p) = \inf_{n \geq 1} n^{-1} E(n * p). \quad (3)$$

Clearly, $f(p) \geq g(p)$ for $0 < p < 1$, and $f(p) = 1$ for $p_1 \leq p < 1$. The information lower bound is sharp for small p in the sense that $f(p)/g(p) \rightarrow 1$

as $p \rightarrow 0+$ (See Sobel and Groll (1959)). But since $f(p_1) = 1 > g(p_1)$ (≈ 0.959), there is room for improvement in a neighborhood of $p = p_1$. In the next section, a weak concavity property of $f(p)$ is established. Using this property, we derive in section 3 a lower bound $h(p)$, which increases continuously from 0 to 1 as p increases from 0 to p_1 . In addition, $h(p)$ equals $g(p)$ for $p \leq p_0$ and is greater than $g(p)$ for $p_0 < p \leq p_1$ where $p_0 = 1 - q_0$ (≈ 0.363) is determined by

$$1 + \log_2 p_0 - (p_0/q_0) \log_2(p_0/q_0) = 0, \quad 0 < p_0 < p_1. \quad (4)$$

2. A weak concavity property of $f(p)$.

In this section, we prove the following theorem.

Theorem 1. For fixed $0 < u < p_1$, let $u^* = (1-2u)/(1-u) > p_1$. Then the curve $(p, f(p))$, $u < p < u^*$, is never below the line connecting $(u, f(u))$ with $(u^*, f(u^*)) = (u^*, 1)$; i.e., for $u < p < u^*$,

$$f(p) \geq f(u) + (p-u)(1-f(u))/(u^*-u).$$

To prove Theorem 1, we need two lemmas.

Lemma 1. For $0 < p < 1$,

$$f(p) = \lim_{n \rightarrow \infty} n^{-1} E(n * p).$$

Proof. By the definition of $f(p)$, it is obvious that

$$\liminf_{n \rightarrow \infty} n^{-1} E(n * p) \geq f(p).$$

Suppose that for some $0 < p < 1$,

$$\limsup_{n \rightarrow \infty} n^{-1} E(n * p) > f(p);$$

i.e. there exist $\epsilon > 0$ and a sequence $\{n_k\}$ such that $n_k \rightarrow \infty$ as $k \rightarrow \infty$ and

$$n_k^{-1} E(n_k * p) > f(p) + \epsilon, \text{ for all } k.$$

Since $f(p)$ is the infimum of $n^{-1} E(n * p)$, there exists m such that

$$m^{-1} E(m * p) < f(p) + \epsilon/2. \text{ Write}$$

$$n_k = [n_k/m]m + r_k$$

where $0 \leq r_k < m$ and $[x]$ denotes the largest integer not exceeding x . Using the inequality $E((a+b) * p) \leq E(a * p) + E(b * p)$ for integers a, b ,

$$\begin{aligned} f(p) + \epsilon &< n_k^{-1} E(n_k * p) \leq n_k^{-1} \{ [n_k/m] E(m * p) + E(r_k * p) \} \\ &\rightarrow m^{-1} E(m * p) \quad \text{as } k \rightarrow \infty, \end{aligned}$$

a contradiction. This completes the proof.

Lemma 2. For $0 < p \leq p_1$ and $(1-2p)/(1-p) \leq p' < 1$,

$$E(m * p, \ell * p') = E(m * p) + \ell.$$

Proof. This lemma states that for a set of $m+\ell$ items with two different defective probabilities p and p' , there exists an optimal procedure which tests individually those ℓ items with defective probability p' if $p' \geq (1-2p)/(1-p)$. But this is a consequence of Corollary 2 of Yao and Hwang (1988).

Proof of Theorem 1. Fix p with $u < p < u^*$. Let $\{I_1, \dots, I_n\}$ and $\{I'_1, \dots, I'_n\}$ be two independent sequences of items such that $I_1, \dots, I_n$ each independently have probability u of being defective, and $I'_1, \dots, I'_n$ each independently have probability u^* of being defective. Let $\{X_1, \dots, X_n\}$ be a sequence of Bernoulli random variables (independent of the I_i and I'_i) such that $P(X_i = 1) = s = 1 - P(X_i = 0)$ where $s = (p-u)/(u^*-u)$. Define

$$J_i = \begin{cases} I_i, & \text{if } X_i = 0 \\ I'_i, & \text{if } X_i = 1 \end{cases}$$

Clearly, $J_1, \dots, J_n$ each independently have probability p of being defective. If the X_i are not known, the minimum expected number of tests required to identify the states of the J_i equals $E(n * p)$. If the X_i are known, the minimum expected number of tests equals

$$\sum_{k=0}^n \binom{n}{k} s^k (1-s)^{n-k} E((n-k) * u, k * u^*)$$

$$= \sum_{k=0}^n \binom{n}{k} s^k (1-s)^{n-k} \{E((n-k) * u) + k\}, \text{ by Lemma 2}$$

$$= \sum_{k=0}^n \binom{n}{k} s^k (1-s)^{n-k} E((n-k) * u) + ns.$$

So,

$$n^{-1} E(n * p) \geq n^{-1} \sum_{k=0}^n \binom{n}{k} s^k (1-s)^{n-k} E((n-k) * u) + s. \quad (5)$$

Letting $n \rightarrow \infty$, the left-hand side of (5) tends to $f(p)$ by Lemma 1, while the first term on the right-hand side approaches, by Lemma 1, the Central Limit Theorem and the inequality $E((n-k) * u) \leq n-k$,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sum_{|k-ns| \leq n^{2/3}} \binom{n-k}{n} \left\{ \frac{1}{n-k} E((n-k) * u) \right\} \binom{n}{k} s^k (1-s)^{n-k} \\ & + \lim_{n \rightarrow \infty} \sum_{|k-ns| > n^{2/3}} \left\{ \frac{1}{n} E((n-k) * u) \right\} \binom{n}{k} s^k (1-s)^{n-k} \\ & = (1-s)f(u), \end{aligned}$$

completing the proof.

Remark. Motivated by Theorem 1, we conjecture that $f(p)$ is concave for $0 < p < 1$. If this conjecture is true, then it is easily seen, by using $f(p) \geq g(p)$ and $f(p_1) = 1$, that the curve $(p, f(p))$, $\tilde{p}_0 < p < p_1$ is never below the line connecting $(\tilde{p}_0, g(\tilde{p}_0))$ with $(p_1, 1)$ where $\tilde{p}_0 = 1 - \tilde{q}_0$ is determined by

$$1 + p_1 \log_2 \tilde{p}_0 + q_1 \log_2 \tilde{q}_0 = 0, \quad \tilde{p}_0 < p_1.$$

(Note that this line is tangent to $P(p, g(p))$ at $p = \tilde{p}_0$.)

3. A lower bound for $f(p)$.

We now construct a lower bound for $f(p)$ using Theorem 1 and the fact that $f(p) \geq g(p)$ for $0 < p < 1$ and $f(p) = 1$ for $p_1 \leq p < 1$. Note that p_0 ,

defined implicitly in (4), has the property that the line connecting $(p_0, g(p_0))$ with $(p_0^*, 1)$ is tangent to the curve $(p, g(p))$ at $p = p_0$ where $p_0^* = (1-2p_0)/(1-p_0)$. Let $(p, h^{(0)}(p))$, $p_0 \leq p \leq p_1$ be the line (restricted to $p_0 \leq p \leq p_1$) connecting $(p_0, g(p_0))$ with $(p_0^*, 1)$. By Theorem 1, for $p_0 < p < p_1$,

$$g(p) < h^{(0)}(p) \leq f(p).$$

For each u with $p_0 \leq u < p_1$, define

$$h_u^{(0)}(p) = \begin{cases} h^{(0)}(p) & , p_0 \leq p \leq u, \\ h^{(0)}(u) + (p-u) \frac{1-h^{(0)}(u)}{u^*-u}, & u < p < p_1 \end{cases} \quad (6)$$

where $u^* = (1-2u)/(1-u)$. In other words, the curve $(p, h_u^{(0)}(p))$ is identical to $(p, h^{(0)}(p))$ for $p \leq u$, and identical to the line connecting $(u, h^{(0)}(u))$ with $(u^*, 1)$ for $p \geq u$. It is easily seen that each $h_u^{(0)}$ is continuous, convex, increasing and bounded from above by $f(p)$.

Define, for $p_0 \leq p < p_1$,

$$h^{(1)}(p) = \sup_{p_0 \leq u < p_1} h_u^{(0)}(p), \quad (7)$$

which is convex and increasing. Also,

$$h^{(0)}(p) \leq h^{(1)}(p) \leq f(p).$$

We define $h_u^{(1)}$ as in (6) with $h^{(0)}$ replaced by $h^{(1)}$, and define $h^{(2)}$ as in (7) with $h_u^{(0)}$ replaced by $h_u^{(1)}$. By repeating this process, we have, for $p_0 \leq p < p_1$,

$$h^{(0)}(p) \leq h^{(1)}(p) \leq h^{(2)}(p) \leq \dots \leq f(p).$$

Let

$$h(p) = \begin{cases} g(p) & , \quad 0 < p < p_0 \\ \lim_{n \rightarrow \infty} h^{(n)}(p) & , \quad p_0 \leq p < p_1 \\ 1 & , \quad p_1 \leq p < 1 \end{cases}$$

Clearly, $h(p)$ is a lower bound for $f(p)$. In what follows, we will find $h(p)$, $p_0 \leq p < p_1$, explicitly.

Lemma 3. For $n = 0, 1, 2, \dots$, and $p_0 \leq u < p_1$, $h^{(n)}(p)$ and $h_u^{(n)}(p)$ are convex and increasing in $p_0 \leq p < p_1$. Also,

$$D h^{(n)}(u+) \leq D h_u^{(n)}(u+), \quad p_0 \leq u < p_1,$$

where $D y(x+)$ denotes the right derivative of a function y at x (which always exists for convex function y .)

Proof. We proceed by induction on n . The case of $n=0$ is true obviously. Suppose the lemma holds for $n=\ell$. Since $h_u^{(\ell)}(p_0 \leq u < p_1)$ are convex and increasing by the induction hypothesis,

$$h^{(\ell+1)}(p) = \sup_u h_u^{(\ell)}(p)$$

is convex and increasing. Now it suffices to show that, for $p_0 \leq u < p_1$,

$$\begin{aligned} D h^{(\ell+1)}(u+) &\leq \{1 - h^{(\ell+1)}(u)\} / (u^* - u) \\ &= D h_u^{(\ell+1)}(u+) \end{aligned} \quad (8)$$

((8) implies that $h_u^{(\ell+1)}(p)$ is convex and increasing since $h^{(\ell+1)}(p)$ is convex and increasing.) To prove (8), we first consider the case of u satisfying $h^{(\ell)}(u) < h^{(\ell+1)}(u)$. From the continuity of $h^{(\ell)}(p)$ and $h^{(\ell+1)}(p)$ in p and the continuity of $h_v^{(\ell)}(p)$ in both v and p , it follows that there exists $0 < \delta < p_1 - u$ such that for $u \leq x < u + \delta < p_1$, $h^{(\ell)}(x) < h^{(\ell+1)}(x)$ and

$$\sup_{u \leq v \leq x} h_v^{(\ell)}(x) < h^{(\ell+1)}(x). \quad (9)$$

For $u < x < u + \delta$, from (9), the continuity of $h_v^{(\ell)}(x)$ in v and the fact that

$$h^{(\ell+1)}(x) = \sup_{p_0 \leq v \leq x} h_v^{(\ell)}(x),$$

it follows that there exists $p_0 \leq v(x) < u$ such that

$$h^{(\ell+1)}(x) = h_{v(x)}^{(\ell)}(x). \quad (10)$$

So, for $u < x < u+\delta$,

$$\begin{aligned} D h^{(\ell+1)}(u+) &\leq \{h^{(\ell+1)}(x) - h^{(\ell+1)}(u)\}/(x-u), \text{ by convexity of } h^{(\ell+1)} \\ &= \{h_{v(x)}^{(\ell)}(x) - h^{(\ell+1)}(u)\}/(x-u) \\ &\leq \{h_{v(x)}^{(\ell)}(x) - h_{v(x)}^{(\ell)}(u)\}/(x-u). \end{aligned} \quad (11)$$

By (10) and the fact that $v(x) < u < x$, the right-hand side of (11) equals the slope of the line connecting $(v^*(x), 1)$ with $(x, h_{v(x)}^{(\ell)}(x)) = (x, h^{(\ell+1)}(x))$ where $v^*(x) = (1-2v(x))/(1-v(x))$. But since $v(x) < x$, this slope is less than that of the line connecting $(x, h^{(\ell+1)}(x))$ with $(x^*, 1)$ where $x^* = (1-2x)/(1-x) < v^*(x)$. Hence,

$$D h^{(\ell+1)}(u+) \leq \{1 - h^{(\ell+1)}(x)\}/(x^* - x). \quad (12)$$

Letting $x \downarrow u$, the right-hand side of (12) tends to

$$\{1 - h^{(\ell+1)}(u)\}/(u^* - u) = D h_u^{(\ell+1)}(u+),$$

which proves (8) for the case where $h^{(\ell)}(u) < h^{(\ell+1)}(u)$. Next, consider the case of u satisfying $h^{(\ell)}(u) = h^{(\ell+1)}(u)$. If there exists $\epsilon > 0$ such that $h^{(\ell)}(x) = h^{(\ell+1)}(x)$ for $u \leq x \leq u+\epsilon$, then (8) holds by the induction hypothesis. If there exists a sequence $\{x_k\}$ such that $x_k \downarrow u$ and $h^{(\ell)}(x_k) < h^{(\ell+1)}(x_k)$, then

$$\begin{aligned}
D h^{(\ell+1)}(u+) &\leq \lim_{k \rightarrow \infty} D h^{(\ell+1)}(x_k+), \text{ by convexity of } h^{(\ell+1)}, \\
&\leq \lim_{k \rightarrow \infty} \frac{1-h^{(\ell+1)}(x_k)}{x_k^*-x_k}, \text{ since } h^{(\ell)}(x_k) < h^{(\ell+1)}(x_k), \\
&= \{1-h^{(\ell+1)}(u)\}/(u^*-u)
\end{aligned}$$

where $x_k^* = (1-2x_k)/(1-x_k)$. So, (8) holds for all $p_0 \leq u < p_1$, completing the proof.

Lemma 4.

- (i) For $p_0 \leq p < p_1$, $h(p)$ is convex, increasing and continuous.
- (ii) For $p_0 \leq p < p_1$, $D h(p+) = \{1-h(p)\}/(p^*-p)$ where $p^* = (1-2p)/(1-p)$.
- (iii) For $p_0 < p < p_1$, $D h(p-) = D h(p+)$.

Proof. Part (i) is trivial. To prove (ii), note that, since $h^{(n)}$ converges to h , $h_u^{(n)}$ also converges to h_u so that

$$h(p) = \sup_{p_0 \leq u < p_1} h_u(p)$$

where h_u is defined as in (6) with $h^{(0)}$ replaced by h .

That is, $h(x) \geq h_p(x)$ for all x and for all $p_0 \leq p < p_1$. Since $h(p) = h_p(p)$,

$$D h(p+) \geq D h_p(p+) = \{1-h(p)\}/(p^*-p).$$

On the other hand,

$$\begin{aligned}
D h(p+) &= \lim_{\epsilon \rightarrow 0+} \{h(p+\epsilon) - h(p)\} \epsilon^{-1} \\
&= \lim_{\epsilon \rightarrow 0+} \lim_{n \rightarrow \infty} \{h^{(n)}(p+\epsilon) - h^{(n)}(p)\} \epsilon^{-1} \\
&\leq \lim_{\epsilon \rightarrow 0+} \lim_{n \rightarrow \infty} D h^{(n)}((p+\epsilon)+), \text{ by convexity of } h^{(n)},
\end{aligned}$$

$$\begin{aligned}
&\leq \lim_{\epsilon \rightarrow 0+} \lim_{n \rightarrow \infty} \frac{1-h^{(n)}(p+\epsilon)}{(p+\epsilon)^* - (p+\epsilon)}, \text{ by Lemma 3} \\
&= \lim_{\epsilon \rightarrow 0+} \frac{1-h(p+\epsilon)}{(p+\epsilon)^* - (p+\epsilon)} = \{1-h(p)\}/(p^*-p),
\end{aligned}$$

where $(p+\epsilon)^* = (1-2(p+\epsilon))/(1-(p+\epsilon))$, completing the proof of (ii).

To prove (iii), note that, for $\epsilon > 0$ with $p_0 < p-\epsilon$,

$$\begin{aligned}
\frac{1-h(p-\epsilon)}{(p-\epsilon)^* - (p-\epsilon)} &= D h((p-\epsilon)+), \text{ by (ii)} \\
&\leq D h(p-) \quad , \text{ by convexity of } h \\
&\leq D h(p+) = \{1-h(p)\}/(p^*-p).
\end{aligned}$$

The proof is complete by letting $\epsilon \downarrow 0$.

Theorem 2. For $p_0 \leq p < p_1$,

$$h(p) = 1 - C \left[\frac{3-\sqrt{5}}{2} - p \right]^{(5-\sqrt{5})/10} \left[\frac{3+\sqrt{5}}{2} - p \right]^{(5+\sqrt{5})/10}$$

where

$$C = \{1-g(p_0)\} / \left\{ \left[\frac{3-\sqrt{5}}{2} - p_0 \right]^{(5-\sqrt{5})/10} \left[\frac{3+\sqrt{5}}{2} - p_0 \right]^{(5+\sqrt{5})/10} \right\}.$$

Proof. By Lemma 4, $h(p)$ satisfies the first order differential equation

$$D h(p) = (1-h(p))/(p^*-p)$$

with the initial condition $h(p_0) = g(p_0)$. The theorem follows by solving the differential equation.

It is easy to prove directly that h is convex and increasing, and $D h(p_0) = D g(p_0)$. So, $h(p) > g(p)$ for $p_0 < p < p_1$ since g is concave.

4. Conclusion.

We established a weak concavity property of $f(p)$ and derived a lower bound $h(p)$ which is an improvement over the information lower bound for p close to $(3-\sqrt{5})/2$. While the improvement is not very significant, it appears to be only the second result in the literature on improving the information bound, the first one being $f(p)=1$ for $(3-\sqrt{5})/2 \leq p < 1$ due to Unger (1960). Moreover, the new bound increases continuously from 0 to 1 as p increases from 0 to $(3-\sqrt{5})/2$. It would be of great interest to evaluate numerically or analytically

$$\inf_{0 < p < 1} \{h(p)/f(p)\}.$$

This quantity is a good measure of the sharpness of h over the entire range of p .

References

- Berger, T., Mehravari, N., Towsley, D. and Wolf, J. (1984) "Random multiple-access communication and group testing," *IEEE Transactions on Communication* C-34, 769-779.
- Sobel, M. and Groll, P.A. (1959) "Group testing to eliminate efficiently all defectives in a binomial sample," *Bell System Technical Journal*, 38, 1179-1252.
- Unger, P. (1960) "The cut-off point in group testing," *Communication in Pure and Applied Mathematics*, 13, 49-54.
- Yao, Y.C. and Hwang, F.K. (1988) "Individual testing of independent items in optimal group testing," *Probability in Engineering and Informational Sciences*, 2, to appear.