

SHARPER BOUNDS IN ADAPTIVE GROUP TESTING

Laura Riccio and Charles J. Colbourn

Abstract. Adaptive group testing in the presence of a large percentage of defectives is best done by individual testing rather than by pooling. The fraction of items which must be defective to make individual testing optimal remains unknown, and is conjectured to be $1/3$. In this paper it is shown that when the number of items is sufficiently large, and the fraction of defective items is at least $1/\log_{3/2} 3$, individual testing is optimal.

1. INTRODUCTION

Combinatorial group testing concerns itself with the identification of certain “defective” or “contaminated” samples distributed amongst a population. One of the primary goals is to determine, for a given situation, what testing algorithm minimizes the number of tests required in the worst case. An algorithm that achieves this minimum is a *minimax* algorithm. A typical scenario is as follows (see [2]). Suppose that one has n blood samples of which d are contaminated. One can test any subset (*pool*) of the n samples simultaneously, determining either that all samples in the subset are free of contamination, or that at least one of the samples is contaminated. One can await the result of the test for one pool before determining the membership of the next pool, so the testing is *adaptive*. The goal is to design a testing algorithm such that the largest possible number of tests to determine which samples are contaminated is minimized. Denote this worst case time for the best algorithm by $M(d, n)$.

The most obvious algorithm, that of individual testing, requires exactly $n - 1$ tests, as the nature of the final sample can be determined from the

0

Received October 30, 1998.

Communicated by F. K. Hwang.

2000 *Mathematics Subject Classification*: 62K99, 90B25, 62C20, 05A99.

Key words and phrases: Adaptive group testing, minimax algorithm, individual testing.

*The research was supported in part by the U.S. Army Research Office under grant DAAG55-98-1-0272.

nature of the others. A question arises: For which cases is this algorithm minimax? It has been conjectured [2] that if $n \leq 3d$, then individual testing is minimax. Du and Hwang [1] prove that when $n > 3d$, one has $M(d, n) < n - 1$. Thus, if true, the conjecture is best possible.

Hu, Hwang, and Wang [3] prove

Theorem 1.1. $M(d, n) \geq \min\{n-1, 2q + \lceil \log_2 \binom{n-q}{d-q} \rceil\}$ for $0 < q \leq d < n$.

We are interested in the ratio of n/d . Let us define $s = n/d$ and $t = q/d$. Then $0 < t \leq 1$. For a given value of s , we attempt to find $t \in (0, 1) \subset (0, 1]$ such that $2td + \lceil \log_2 \binom{sd-td}{d-td} \rceil \geq sd - 1$, at least for all sufficiently large d . This condition is equivalent to

$$\log_2 \binom{sd - td}{d - td} > sd - 2td - 2;$$

equivalently,

$$(1) \quad \binom{sd - td}{d - td} \cdot 2^{2+2td-sd} > 1.$$

Du and Hwang [1] prove that when $s = 21/8$, statement (1) holds for all d , for values of t given in their theorem. In this paper, we show that for all *sufficiently large* d , we can do better. If $n/d < \log_{3/2} 3 \approx 2.70951129$, then individual testing is minimax. We further show that this is the best possible result obtainable by Theorem 1.1.

2. MAIN RESULTS

Let us define $\Psi(s, t) = (s - t)^{s-t} 2^{2t} / (1 - t)^{1-t} (s - 1)^{s-1} 2^s$.

Theorem 2.1. $\binom{sd-td}{d-td} \cdot 2^{2+2td-sd} > 1$ for all sufficiently large d if and only if $\Psi(s, t) > 1$.

Proof. Consider the asymptotic behavior of the function

$$\begin{aligned} R(d) &= \binom{sd - td}{d - td} \cdot 2^{2+2td-sd} \\ &= \frac{4((s - t)d)!}{((1 - t)d)!((s - 1)d)!} \cdot 2^{(2t-s)d}. \end{aligned}$$

We use Stirling's approximation, $x! \approx \sqrt{2\pi x} x^{x+1/2} e^{-x}$. Recalling that the ratio of $x!$ to its approximation approaches unity as $x \rightarrow \infty$, we may write

this in the limit as

$$\begin{aligned} & \frac{4 \cdot \sqrt{2\pi} \cdot ((s-t)d)^{(s-t)d+1/2} \cdot e^{(t-s)d}}{2\pi \cdot ((1-t)d)^{(1-t)d+1/2} \cdot ((s-1)d)^{(s-1)d+1/2} \cdot e^{(t-1)d} \cdot e^{(1-s)d}} \cdot \frac{2^{2td}}{2^{sd}} \\ &= \frac{2\sqrt{2\pi}}{\pi} \cdot \sqrt{s-t}\sqrt{1-t}\sqrt{s-1} \cdot \frac{1}{\sqrt{d}} \cdot \frac{((s-t)d)^{(s-t)d} \cdot 2^{2td}}{((1-t)d)^{(1-t)d} \cdot ((s-1)d)^{(s-1)d} \cdot 2^{sd}} \\ &= \frac{2\sqrt{2\pi}}{\pi} \cdot \frac{\sqrt{s-t}}{\sqrt{1-t}\sqrt{s-1}} \cdot \frac{1}{\sqrt{d}} \cdot \left(\frac{(s-t)^{s-t} \cdot 2^{2t}}{(1-t)^{1-t}(s-1)^{s-1}2^s} \right)^d. \end{aligned}$$

The function in parentheses is precisely $\Psi(s, t)$. The behavior of this function determines the behavior of R . Specifically,

$$\lim_{d \rightarrow \infty} R(d) = \begin{cases} 0 & \text{if } \Psi(s, t) \leq 1 \\ \infty & \text{if } \Psi(s, t) > 1 \end{cases}$$

and the result follows. ■

By the definitions of s and t , we need only consider values of $t \in (0, 1)$, and $s > 21/8 > 2$. Define $\phi = \log_{3/2} 3$ and $\xi = \log_{3/2} (3 \cdot 2^{-4/3})$.

We shall show that for all $s < \phi$, one can choose $t \in (0, 1)$ such that $\Psi(s, t) > 1$. Hence statement (1) holds for all sufficiently large d , and individual testing is minimax. Further, for all $s \geq \phi$ and for all $t \in (0, 1)$, we obtain $\Psi(s, t) \leq 1$. In this case, statement (1) holds for no sufficiently large d , and we cannot deduce that individual testing is minimax.

Lemma 2.2. $\Psi(\phi, \xi) = 1$.

Proof. Let $\tau = \log_{3/2} 2^{1/3}$. Then

$$\begin{aligned} \Psi(\phi, \xi) &= \frac{(\log_{3/2} 2^{4/3})^{(\log_{3/2} 2^{4/3})} \cdot 2^{(\log_{3/2} (9 \cdot 2^{-8/3}))}}{(\log_{3/2} 2^{1/3})^{(\log_{3/2} 2^{1/3})} (\log_{3/2} 2)^{(\log_{3/2} 2)} 2^{(\log_{3/2} 3)}} \\ &= \frac{(4\tau)^{4\tau} \cdot 2^{2-2\tau}}{\tau^\tau \cdot (3\tau)^{3\tau} \cdot 2^{1+3\tau}} = \frac{\tau^{4\tau} \cdot 4^{4\tau}}{\tau^{4\tau} \cdot 3^{3\tau}} \cdot 2^{1-5\tau} = \frac{2^{8\tau}}{3^{3\tau}} \cdot 2^{1-5\tau} = 2 \cdot (2/3)^{3\tau} \\ &= 2 \cdot (3/2)^{-\log_{3/2} 2} = 2 \cdot \frac{1}{2} = 1. \end{aligned} \quad \blacksquare$$

Lemma 2.3. $\partial\Psi/\partial s < 0$ for all $s > 2$.

Proof. Rewrite $\Psi(s, t)$ as

$$\exp\{(s-t) \log(s-t) + 2t \log 2 + (t-1) \log(1-t) + (1-s) \log(s-1) - s \log 2\}.$$

Then

$$\begin{aligned}\frac{\partial \Psi}{\partial s} &= \Psi(s, t) \cdot [\log(s - t) + 1 - \log(s - 1) - 1 - \log 2] \\ &= \Psi(s, t) \cdot \log\left(\frac{s - t}{2s - 2}\right).\end{aligned}$$

Since $s > 2$ and $0 < t < 1$, we have $0 < s - t < 2s - 2$, and hence $(s - t)/(2s - 2)$ is less than one. Thus the logarithm is negative, while Ψ is always positive, and hence $\partial \Psi / \partial s < 0$. ■

Lemma 2.4. *Assume $s = \phi$. Then $\partial \Psi / \partial t > 0$ for $t < \xi$, $\partial \Psi / \partial t < 0$ for $t > \xi$, and $\partial \Psi / \partial t = 0$ at $t = \xi$.*

Proof. As in Lemma 2.3,

$$\begin{aligned}\frac{\partial \Psi}{\partial t} &= \Psi(s, t) \cdot [-\log(s - t) - 1 + 2 \log 2 + \log(1 - t) + 1] \\ &= \Psi(s, t) \cdot \left[2 \log 2 + \log\left(\frac{1 - t}{s - t}\right)\right].\end{aligned}$$

The term in brackets is strictly decreasing with t , while Ψ is always positive. Further, when we evaluate this term at $(s, t) = (\phi, \xi)$, we obtain $[\log 4 + \log(\tau/4\tau)] = [\log 4 + \log(1/4)] = 0$. Together with the monotonicity of the bracketed term, this gives us the desired result. ■

Corollary 2.5. *Taking $t = \xi$ maximizes $\Psi(\phi, t)$, and this maximum is 1.*

We now establish a final corollary prior to deducing our main result.

Corollary 2.6. (a) *For $s \geq \phi$, $\Psi(s, t) \leq 1$ for all $t \in (0, 1)$.*
 (b) *For $s < \phi$, there is a neighborhood (ξ', ξ'') of ξ such that $\Psi(s, t) > 1$ for all $t \in (\xi', \xi'')$.*

Proof. Statement (a) simply recapitulates the monotonicity of Ψ with s . Statement (b) follows directly from the concavity of Ψ in the neighborhood of ξ . ■

For statement (1) to be applicable, we need $t = q/d$ to be a rational number with d as its denominator. However, for sufficiently large d , we may choose q to be an integer sufficiently close to $d \cdot \xi$ (that is, $q/d \in (\xi', \xi'')$) so that $\Psi(s, t) > 1$. This completes the proof of the main theorem:

Theorem 2.7. *If $c < \phi \approx 2.70951129$, then for sufficiently large d , the (cd, d) problem has individual testing as its minimax algorithm. By the monotonicity of $M(d, n)$, if $n \leq cd$, then the (n, d) problem has individual testing as its minimax algorithm.*

3. CONCLUSION

The value of $\phi = \log_{3/2} 3$ that plays a central role here was discovered empirically in our research. Examining the methods of Du and Hwang [1], we determined that essentially the same proof guarantees that statement (1) holds for $s = 58/22$, $140/53$, and $251/95$, among other values, as well as the value of $21/8$. For these values of s , in fact, the statement holds not just for all sufficiently large d , but for all d . Considering cases when d is sufficiently large, we found that statement (1) holds for $s = 59/22$, $62/23$, $89/33$, $116/43$, $197/73$, $311/115$, and $1971/728$. Indeed, the value of $\phi = \log_{3/2} 3$ was initially ‘guessed’ as the apparent limit of this sequence, and then verified. The value ϕ achieved here appears to be the best result possible via this line of attack. However, we believe that this is more a limitation of Theorem 1.1 than a reason to doubt the conjecture that individual testing is minimax for $n/d \leq 3$.

REFERENCES

1. D. Z. Du and F. K. Hwang, Minimizing a combinatorial function, *SIAM J. Algebraic Discrete Methods* **3** (1982), 523-528.
2. D. Z. Du and F. K. Hwang, *Combinatorial Group Testing*, World Scientific, Singapore, 1993.
3. M. C. Hu, F. K. Hwang and J. K. Wang, A boundary problem for group testing, *SIAM J. Algebraic Discrete Methods* **2** (1981), 81-87.

Laura Riccio
Mathematics and Statistics, University of Vermont,
Burlington, VT 05405, U.S.A.
E-mail: lriccio@emba.uvm.edu

Charles J. Colbourn
Computer Science, University of Vermont,
Burlington, VT 05405, U.S.A.
E-mail: charles.colbourn@uvm.edu