On single defective identification in Nested Ordered Testing

Viktor Skorniakov
Institute of Applied Mathematics
Vilnius University
Vilnius, Lithuania
viktor.skorniakov@mif.vu.lt

***Abstract*—In quality control, one oftentimes requires to identify defectives in a given set of items $I = \{I_1, \dots, I_n\}$, where each I_i can be defective with probability $p_i \in (0, 1)$. If identification is possible by testing, Group Testing (GT) stands out as a good tool offering a cost-saving identification way. We consider Nested Ordered GT procedures for identification of a single defective under non-increasing ordering of defectiveness probabilities and provide two conjectures which, if true, would allow to replace dynamic optimal testing procedure building algorithm of order $O(n \log n)$ by a constant time explicit procedure.**

***Index Terms*—Group Testing, Nested Ordered Procedure, Single Defective.**

I. Introduction

Assume that one has a set of stochastically independent items $I = \{I_1, \dots, I_n\}$. Item I_i can be defective with probability p_i and non-defective (aka pure) with probability $q_i := 1 - p_i$. Defectiveness identification is possible by testing. Each test can be conducted on any $S \subseteq I$. Test outputs 1 if S contains at least one defective. Otherwise, it outputs 0. If S tests positive, no further information is provided.

The first instance of such Group Testing (GT) procedure was described by Dorfman [2] in 1943 in the medical context. A simple example of his construction in industrial setting is as follows. To identify defective bulbs in a Christmas tree string, one can first plug in the whole string (this way making a group test): if it shines, all bulbs are non-defective. Otherwise, one needs to retest each bulb individually. Since [2], GT widespread to many fields including quality control [7]. Papers[1] [11], [1], and [9] give a good overview of the field.

II. Setting

The selection of particular GT procedure to be applied for I at hand is context dependent. However, if possible, it is desirable to choose an optimal one. For each realization of items' set $I = \{I_1, \dots, I_n\}$, the number of tests $T_X(I)$ required by testing procedure X on I is random. Therefore, optimality is understood in an average sense: given a class of GT procedures $\mathcal{X}$, procedure $X_0 \in \mathcal{C}$ is optimal if

$$\mathrm{E}\, T_{X_0}(I) = \inf_{X \in \mathcal{X}} \mathrm{E}\, T_X(I).$$

In our case, we consider Nested Ordered Procedures (ONPs) introduced in [8] and defined by the following constraints:

> **(NOP1)** One treats the set I as an n-tuple $(I_1, \dots, I_n)$ thereby introducing the order[2].
>
> **(ONP2)** If at some testing step one finds out that $D \subseteq I$ contains defectives (such D is called *defective set*) and $|D| > 1$, then, in the next testing step, one tests proper $S \subset D$. Otherwise, one tests $S \subseteq B \subseteq I$, where B is a set of items whose status (defective/pure) is unknown so far (such B is called *binomial set*).
>
> **(ONP3)** One is allowed to test only subsets consisting of the first items of the defective/binomial set.

Furthermore, we impose monotonicity constraint

$$p_1 \geq \dots \geq p_n \tag{1}$$

and focus on two specific subclasses of ONPs:

> **(C1)** The procedures for identification of a single defective (SDIPs) in the defective set.
>
> **(C2)** SDIPs in the binomial set.

III. Related work and conjectures

Let $I = (I_1, \dots, I_n)$ be the set of interest (defective or binomial) and let $p_1, \dots, p_n$ be corresponding defectiveness probabilities not necessarily satisfying (1). The general *dynamic* SDIP building algorithm is known [5], [3], [6]: in case of (C1), one has to construct Optimal Alphabetic Binary Tree (OABT) for the set of weights $w_1 = p_1$, $w_i = p_i q_1 \dots q_{i-1}$, $i = 2, \dots, n$ and convert it to the testing tree by appropriate labeling; in case of (C2), one has to do the same for $w_1 = p_1$, $w_i = p_i q_1 \dots q_{i-1}$, $i = 2, \dots, n$, and $w_{n+1} = q_1 \dots q_n$.

In general, the complexity of building OABT is $O(n \log n)$. Though it can be reduced to $O(n)$ when weights form a valley sequence [13], [4], the computational effort is still non-negligible for n large. Therefore, complexity reduction remains important problem. Considering homogeneous case $p_1 = \dots = p_n = p$, Hwang and Yao [15] demonstrated that OABT based (C1) procedure can be described explicitly for $p \in \left[1 - \frac{1}{\sqrt{2}}, \frac{3-\sqrt{5}}{2}\right]$ and there is no need to invoke dynamic construction. Building on their result, we conjecture that the

[1] though the title of [9] does not pretend to review, the paper contains a very detailed account on the subject as a whole

[2] if $I = \{I_1, I_2\}$, then $(I_1, I_2) \neq (I_2, I_1)$

same holds true provided p_i's satisfy (1) with $p_1 \leq \frac{3-\sqrt{5}}{2}$ and $p_n \geqslant a \approx 0.245$, where $1-a$ is a root of polynomial $(0,1) \ni x \mapsto x(1+x(1-x^2))-1$. Our conjecture is supported by preliminary calculations and small scale simulation[3] showing that, under imposed constraints, OABTs defining optimal (C1)–(C2) type procedures are of simple structure like in figures 1a–1b: (i) they can test only in pairs or one-by-one; (ii) testing in pairs, if any, can occur only in the initial portion of the whole testing process (described in the figures' caption).

IV. Several concluding remarks

Single defective identification procedures may be applied in the manufacturing where it is possible to estimate defectiveness probabilities from auxiliary information. They are interesting not only because sometimes it is enough to isolate first defective but also because they serve as building blocks for optimal ONPs aiming at complete classification of the whole set I [14]. As noted in [8], ONPs may provide a good substitute for globally optimal procedure unknown to date. Furthermore, if proved, conjectures stated here will make a further step towards solving important conjecture announced[4] by Malinovsky [10].

References

[1] Matthew Aldridge, Oliver Johnson, and Jonathan Scarlett. Group testing: An information theory perspective. *Foundations and Trends® in Communications and Information Theory*, 15(3-4):196–392, 2019.

[2] R. Dorfman. The detection of defective members of large populations. *The Annals of Mathematical Statistics*, 14(4):436–440, 1943.

[3] M. R. Garey and F. K. Hwang. Isolating a single defective using group testing. *Journal of the American Statistical Association*, 69(345):151–153, 1974.

[4] T. C. Hu and Man-Tak Shing. *Combinatorial algorithms: T.C. Hu and M.T. Shing*. Dover Publications, Mineola, N.Y, 2. ed edition, 2002.

[5] F. K. Hwang. A group testing problem. In *Proceedings of the Fourth Southeastern Conference on Combinatorics, Graph Theory and Computing*, pages 345–355, 1973.

[6] FK Hwang. Optimal group testing procedures in identifying a single defective from a finite population. *Bulletin of the Institute of Mathematics, Academia Sinica*, 8(2/3):129, 1980.

[7] Norman L. Johnson, Samuel Kotz, and Xizhi Wu. *Inspection Errors for Attributes in Quality Control*. Springer US, Boston, MA, 1991.

[8] D. Kurtz and M. Sidi. Multiple access algorithms via group testing for heterogeneous population of users. *IEEE Transactions on Communications*, 36(12):1316–1323, 1988.

[9] Nicholas C. Landolfi and S. Lall. Optimal Dorfman Group Testing for Symmetric Distributions. *SIAM J. Math. Data Sci.*, 6:731–760, 2023.

[10] Yaakov Malinovsky. Conjectures on optimal nested generalized group testing algorithm. *Applied Stochastic Models in Business and Industry*, 36(6):1029–1036, 2020.

[11] Yaakov Malinovsky and Paul S. Albert. Revisiting Nested Group Testing Procedures: New Results, Comparisons, and Robustness. *The American Statistician*, 73(2):117–125, April 2019.

[12] Yaakov Malinovsky and Viktor Skorniakov. The optimality of a nested generalized pairwise group testing procedure. *IEEE Transactions on Information Theory*, 72(2):1142–1148, 2026.

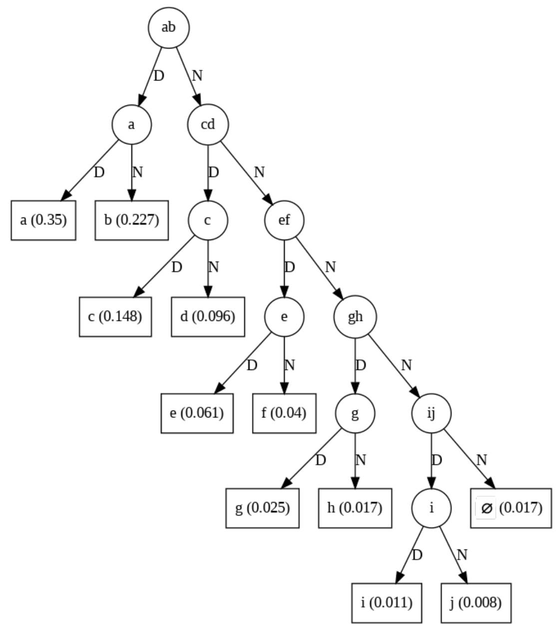
[13] S.V. Nagaraj. Optimal binary search trees. *Theoretical Computer Science*, 188(1):1–44, 1997.

[14] Viktor Skorniakov. On the Optimal Nested Ordered Algorithm. *Methodology and Computing in Applied Probability*, 27(4):83, December 2025.

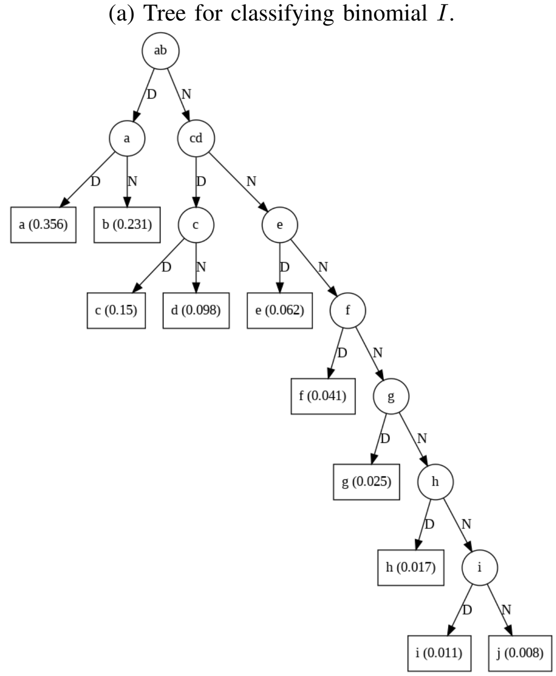
[15] Y.C. Yao and F.K. Hwang. On optimal nested group testing algorithms. *Journal of Statistical Planning and Inference*, 24(2):167–175, February 1990.

[3] conducted by randomly choosing constraints satisfying probabilities and checking that corresponding test trees have announced structure

[4] one of conjectures stated in [10] was recently solved [12]



(a) Tree for classifying binomial I.



(b) Tree for classifying defective I.

Fig. 1: Trees for finding single defective in $I' = \{a,b,c,d,e,f,g,h,i,j\}$ with $p_a = p_b = p_c = p_d = 0.35$, $p_e = p_f = 0.34$, $p_g = p_h = 0.32$, $p_i = p_j = 0.31$ when: (a) I is binomial; (b) I is defective. Letters in non-terminal (circle) nodes show the tested subsets. D (N) points to the next tested subset if node's subset tests positive (negative). Terminal nodes' labels define output—the first defective discovered by the procedure. Numbers in the braces are weights which coincide with probabilities of reaching this node and declaring that this particular item is the first defective. To test given set I, one starts in the root and moves along the tree until terminal (rectangular) node gets reached.