Statistics and Probability Letters 140 (2018) 44–47

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

On optimal policy in the group testing with incomplete identification

Yaakov Malinovsky

Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD 21250, USA

ARTICLE INFO

Article history:
Received 8 December 2017
Received in revised form 7 April 2018
Accepted 13 April 2018
Available online 24 April 2018

Keywords:
Dynamic programming
Optimal design
Partition problem
Shur-convexity

ABSTRACT

Consider a very large (infinite) population of items, where each item independent from the others is defective with probability p, or good with probability $q = 1 - p$. The goal is to identify N good items as quickly as possible. The following group testing policy (policy A) is considered: test items together in the groups, if the test outcome of group i of size n_i is negative, then accept all items in this group as good, otherwise discard the group. Then, move to the next group and continue until exact N good items are found. The goal is to find an optimal testing configuration, i.e., group sizes, under policy A, such that the expected waiting time to obtain N good items is minimal. Recently, Gusev (2012) found an optimal group testing configuration under the assumptions of constant group size and $N = \infty$. In this note, an optimal solution under policy A for finite N is provided.

1. Introduction and problem formulation

Consider a subset of x items, where each item has the probability p to be defective, and $q = 1-p$ to be good independently from the other items. Following the accepted notation in the group testing literature, we call that model a binomial model (Sobel and Groll, 1959). A group test applied to the subset x is a binary test with two possible outcomes, positive or negative. The outcome is negative if all x items are good, and the outcome is positive if at least one item among x items is defective.

In 1943, Robert Dorfman introduced the concept of group testing based on the need to administer syphilis tests to a very large number of individuals drafted into the U.S. army during World War II. The goal was *complete identification* of all drafted people. The Dorfman procedure (Dorfman, 1943) is a two-stage procedure, where the group is tested in the first stage and if the outcome is positive, then in the second stage individual testing is performed. If the group test outcome is negative in the first stage, then all items in the group are accepted as good. In this simple procedure, the saving of time may be substantial, especially for the small values of p. For example, if $p = 0.01$, when compared with individual testing, the reduction in the expected number of tests is about 80%.

Since the Dorfman work, group testing has wide-spread applications from communication networks (Wolf, 1985) to DNA and blood screening (Du and Hwang, 2006; Bar-Lev et al., 2017). Until today, an optimal group testing procedure for complete identification under binomial model is unknown for $p < (3 - 5^{1/2})/2$. For $p \geq (3 - 5^{1/2})/2$ Ungar (1960) proved that the optimal group testing procedure is an individual, one-by-one testing (at the boundary point it is an optimal). However, substantial improvements of Dorfman's procedure were obtained (Sterrett, 1957; Sobel and Groll, 1959; Hwang, 1976). For the review and comparisons among group testing procedures under binomial model see Malinovsky and Albert (2018).

To the best of our knowledge, the *incomplete identification* problem was introduced by Bar-Lev et al. (1990) and extended by Bar-Lev et al. (1995). In their model, demand D of good items should be fulfilled by purchasing two kinds of items. The first kind is 100% quality items with the purchasing cost s per unit, and the second kind is 100q% quality items with purchasing

E-mail address: yaakovm@umbc.edu.

https://doi.org/10.1016/j.spl.2018.04.015
0167-7152/

cost c per unit. In addition, there is cost K for each group-test regardless of the size of the tested group with the items of $100q\%$ quality. Under these constrains/assumptions, the authors found an optimal number of $100q\%$ quality to purchase (once) and an optimal group size chosen from the purchased group, in each stage of the testing procedure. It is related to the problem we discuss here, but with different assumptions and constrains.

Consider the binomial model with a very large (infinite) population of items. The goal is to identify N good items as quickly as possible. This is an *incomplete identification* problem. We consider the following group testing policy (policy A): Test items together in the groups, if test outcome of the group i of size n_i is negative, then accept all items in this group as good, otherwise discard the group. Then, move to the next group and continue until exact N good items will be found. The goal is to find an optimal testing configuration, i.e., group sizes, under policy A, such that the expected waiting time to obtain N good items is minimal.

In the recent work (Gusev, 2012) the problem of incomplete identification was considered. The policy A was applied under assumptions $N = \infty$ and a constant group size. The author found an optimal group size as a function of q. It can be explained as follows: Each time a group of size n is tested, if the test outcome of the group is negative, then accept all n items in this group as good, otherwise discard the group and take the next group of size n and so on. The waiting time (number of tests until first good group) is a geometric random variable with expectation $\frac{1}{q^n}$. Therefore, the mean waiting time per one good item is $\frac{1}{nq^n}$. We want to minimize this quantity. It is equivalent to maximizing the function $\mu(n, q) = nq^n$, which is concave as a function of continuous variable n. But, since the feasible solution is an integer, the maximizer is not necessarily unique. In the proposition below we present a slightly modified result by Gusev (2012), which found an optimal group size as the function of q. We also follow the accepted notation in the group testing literature and denote p as the probability to be defective, which is different from Gusev (2012) notation.

Proposition 1 (*Gusev, 2012*)**.** *Define $n^{**} = \frac{1}{\ln(1/q)}$. Under policy A with the constant group size and $N = \infty$, the optimal group size for the $q \geq 1/2$ is*

$$n^* = \begin{cases} n^{**} & \text{if} \quad \text{integer} \\ \lfloor n^{**} \rfloor & \text{if} \quad \mu(\lfloor n^{**} \rfloor, q) > \mu(\lceil n^{**} \rceil, q) \\ \lceil n^{**} \rceil & \text{if} \quad \mu(\lfloor n^{**} \rfloor, q) < \mu(\lceil n^{**} \rceil, q) \\ \lfloor n^{**} \rfloor \text{ or } \lceil n^{**} \rceil & \text{if} \quad \mu(\lfloor n^{**} \rfloor, q) = \mu(\lceil n^{**} \rceil, q), \end{cases} \tag{1}$$

where $\lfloor x \rfloor$ ($\lceil x \rceil$) for $x > 0$ is defined as the largest (smallest) integer, which is smaller (larger) than or equal to x. For $q < 1/2$, the optimal group size n^* equals 1.

Comment 1 (*Cut-off Point*)**.** *There is an analogy of Ungar's cut-off point for the complete identification. It seems that for $N = 2$, the policy A with the groups of size 2 is the only reasonable policy for an incomplete identification problem. For $N = 2$, policy A is better than the individual testing if the expected waiting time $1/q^2$ is less than the expected waiting time $2/q$ under individual testing, i.e., $q > 1/2$. Now, following Ungar (1960) with adoption to incomplete identification case, one can show that if $q < 1/2$, then individual testing is the optimal among all possible strategies for any N. In the boundary case $q = 1/2$, the individual testing is an optimal strategy.*

The problem formulation: Finite N

Under policy A, we are interested in finding an optimal partition $\{m_1, \ldots, m_J\}$ with $m_1 + \cdots + m_J$ for some $J \in \{1, \ldots, N\}$ such that the expected total waiting time to obtain N good items is minimal, i.e.,

$$\{m_1, \ldots, m_J\} = \arg \min_{n_1, \ldots, n_I} \left\{ \frac{1}{q^{n_1}} + \cdots + \frac{1}{q^{n_I}} \right\},$$
$$\text{subject to} \quad \sum_{i=1}^{I} n_i = N, \quad I \in \{1, \ldots, N\}. \tag{2}$$

2. Dynamic programming algorithm and alternative efficient solutions

Denote n $(n = 1, \ldots, N)$ as a number of good items remains yet unidentified and $H(n)$ an optimal total expected time to obtain n good items. Then, if we test a group of size x $(x = 1, \ldots, n)$, we have

$$H(n) = q^x H(n - x) + (1 - q^x) H(n), \quad n = 2, \ldots, N; \quad x = 1, \ldots, n, \tag{3}$$

where $H(0) = 0, \ H(1) = 1$.

Combining $H(n)$ from the left and right-hand side of (3) we obtain the dynamic programming (DP) algorithm:

$$H(0) = 0, H(1) = 1, \tag{4}$$
$$H(n) = \min_{x=1,\ldots,n} \left\{ \frac{1}{q^x} + H(n - x) \right\}, \quad n = 2, \ldots, N.$$

The complexity of calculation of $H(N)$ is $O(N^2)$.