JCSE Journal of Computing Science and Engineering

Regular Paper
Journal of Computing Science and Engineering, Vol. 16, No. 3, September 2022, pp. 121-128

Analysis of Theoretical Bounds in Noisy Threshold Group Testing

Jin-Taek Seong*
Department of Convergence Software, Mokpo National University, Muan, Korea
jtseong@mokpo.ac.kr

Abstract

The objective of this study was to describe a noisy threshold group testing model where positive and negative cases could occur depending on virus concentration in coronavirus disease 2019 (COVID-19) diagnosis with output results flipped due to measurement noise. We investigated lower bounds for successful reconstruction of a small set of defective samples in the noisy threshold group testing framework. To this end, using Fano's inequality, we derived the minimum number of tests required to find unknown signals with defective samples. Our results showed that the minimum number of tests on probability of error for reconstruction of unknown signals was a function of the defective rate and noise probability. We obtained lower bounds for on performance of the noisy threshold group testing framework with respect to noise intervals. In addition, the relationship between defective rate of signals and sparsity of group matrices to design optimal noisy threshold group testing systems is presented.

Category: Algorithms and Complexity

Keywords: Noisy threshold group testing; Lower bound; Defective sample; Fano's inequality

I. INTRODUCTION

Group Testing is a combinatorial problem [1]. Numerous algorithms have emerged to find the solution to the combinatorial problem. In recent years, group testing has been of interest using various approaches such as probabilistic approaches. Compressed sensing [2] has led to the rediscovery of group testing that can solve the problem of sparse signals [3]. In addition, explosive interest and research results of compressed sensing have served as a driving force for solving numerous problems in various fields over the past decade. There has been a more elaborate step to examine the performance of group testing, further revealing performance limits that can be reconstructed depending on the sparsity of signals [3]. In addition, recently, academia is using group testing as a method to find coronavirus disease 2019 (COVID-19) confirmed cases [4, 5].

The beginning of group testing can be found in a report published by Dorfman [1]. During World War II, the Department of Public Health in USA embarked on a large-scale project to find all men with syphilis. The individual syphilis testing is to take each blood sample and analyze the sample to determine the positive or negative case of syphilis. At that time, individual syphilis testing was not only expensive, but also inefficient. It took a lot of cost and time to test all soldiers one by one [3]. Assuming that there are N soldiers as follows, the individual syphilis testing is performed with N tests. If a large number of soldiers are infected with syphilis, individual syphilis testing is more effective. However, if only a very small number of soldiers are infected with syphilis, individual syphilis testing is not effective. Thus, other methods need to be considered.

Open Access http://dx.doi.org/10.5626/JCSE.2022.16.3.121 http://jcse.kiise.org

Received 04 May 2022; **Accepted** 08 Jun 2022
*Corresponding Author

pISSN: 1976-4677 eISSN: 2093-8020

In the initial group testing presented in [1], syphilis group testing was conducted as follows. Blood samples for a set of several soldiers were grouped in one pool to determine whether they could activate syphilis. If the result of the test was positive, it ensured that at least one soldier was infected with syphilis. Otherwise, if negative, it meant that all blood samples pooled for the testing were not infected with syphilis. The reason such a pooled syphilis testing is efficient and cost-effective is because most soldiers were not infected with syphilis. Only a few soldiers were infected. Therefore, the group testing problem has been mainly studied toward two different topics. One is how to construct group testing schemes. In other words, for one test pool, which samples are included in that group. The second is to find unknown defective samples with as small number of tests as possible. As the readers know, there is no benefit of group testing if the number of tests is as large as the number of individual tests.

Group testing has been extended to various models assuming how to express positive and negative results and whether noise affects results. Traditionally, test results from group testing have expressed whether the pool has one or more defects. In other words, if any of the samples in the pool is defective, the output is positive or negative. Another variant of the group testing is the additive model, also known as quantitative group testing [3]. The result of the additive model is the number of defective samples included in the pool. Another major group testing model is called threshold group testing [6]. In threshold group testing, results are expressed as either positive or negative which is the same as traditional group testing. However, unlike the traditional group testing method, it is only positive when the number of defective samples included in the pool is greater than a given threshold. Otherwise, it is negative. The reason why such threshold group testing could be proposed is that it shows a positive or negative result depending on whether the syphilis virus concentration is high or low. There is also a model that considers whether results of group testing are added with measurement noise. Measurement noise can result in false positives and false negatives.

So far, there have been a lot of studies on threshold group testing problems, such as encoding and decoding methods [7], theoretical lower bound and upper bound on performance [8], and designing methods that are robust and efficient ways [9, 10]. However, there is a lack of research on how much noise affects performance in noisy threshold group testing. In this paper, we present guidelines on how to design a noisy threshold group testing framework that is robust and reliable for measurement noise. To this end, defective rate of signals, sparsity of group matrices, and measurement noise are expressed stochastically. In this paper, we explored theoretical performance bounds by considering noisy threshold group testing models.

II. RELATED WORKS

In this section, research results and significance related to group testing are reviewed. Group testing problems are then classified and each type is considered. The concept of the first group testing began in 1943 when Dorfman [1] introduced it in *the Annals of Mathematical Statistics*. Dorfman's work sought to minimize the number of tests required to find all soldiers infected with syphilis using a stochastic approach. Dorfman [1] proposed a simple method as follows. First, to test for syphilis, N soldiers were divided into several groups with a certain group size. Individual testing was performed only for groups that tested to find soldiers infected with syphilis. Dorfman [1] tabulated the optimal group size according to the total number of samples and defective rates. In [11], the author improved Dorfman's method. The Sterrett's method is the same as Dorfman's method to perform individual testing for the group that tested positive. However, when the first positive sample was found, remaining samples were grouped again and tested in a group. Sterrett's method is a little more efficient. The reason is that it supports the assumption that only about one person in the group that shows positive results would have been infected with syphilis. Later, general testing methods for group testing were introduced by Sobel and Groll [12]. They proposed five new methods, including the case where the defective rate is unknown. In [12], it was the first attempt to find a connection between information theory and group testing. It also introduced a new application of group testing and discussed the generalization of group testing.

Group testing problems are classified as adaptive and non-adaptive, or probabilistic and combinatorial. First, the probability model assumes that the defective sample follows a probability distribution. Attempts have been made to minimize the number of tests required for the detection of the defective sample. However, in a combinatorial model, minimizing the number of tests required to find all defective samples without error assuming that the probability distribution of defective samples is unknown is different from the probability type [13, 14], i.e., the problem of finding the minmax algorithm [15]. Recent studies have further improved this threshold [16]. As another class, the adaptive model relates to what information will be used when selecting samples to include in one test. In general, the sample to be included in one next testing may vary depending on the result of previous testing. In other words, the adaptive model refers to the work of using results obtained from the previous testing for the next testing. Conversely, the non-adaptive algorithm may independently select samples to be included in each testing and determine them in advance. Therefore, non-adaptive algorithms provide the advantage of performing tests simultaneously regardless of the sequence of the testing. Non- adaptive algorithms are generalized to multi-stages algorithms if the testing consists of multiple steps with each testing predetermined [1, 17]. Although adaptive models provide more design flexibility, adaptive group testing algorithms cannot significantly improve the number of tests required to find defective samples over a constant value [3, 18]. In addition, the non-adaptive group testing method is more efficient if all tests must be determined in advance.

We revisited the significance of recent studies related to noisy group testing. Theoretical performance bounds for group testing with noise presence or absence have been studied in [19]. In recent years, many studies have presented their research results. The developed algorithm can identify the proportion of positive returns in the group included for each sample. If this number exceeds the appropriately selected threshold, the sample is considered as defective. This is known to follow the optimal scaling law in $K \approx N^{\theta}$ domain where $\theta \in (0, 1)$, although not optimal [17]. In [19] with separate testing of inputs, all tests are used by considering samples individually. In other words, the state of a given sample uses binary values. In the case of the symmetric noise model, it was revealed that the number of tests was within log2 of the optimal information theory bound while approaching $\theta \rightarrow 0$. However, as θ goes away from 0, the convergence rate of the number of tests increases rapidly.

For noisy group testing, papers [20] and [21] have proposed realistic and usable group testing algorithms using belief propagation and linear programming instead of theoretical analysis. In addition, a sublinear algorithm that does not have an optimal performance in sample complexity was proposed to guarantee the number of samples and execution time [22]. Group testing, as mentioned earlier, can be applied to a wide area of fields, including biology, communication, information theory, and data science. Noise-free models have been extensively studied from a theoretical perspective. However, it is unrealistic to assume that all test results are noise-free and accurate in various applications. Most noisy group testing has attempted to solve the problem by considering the symmetric noise model. Nevertheless, asymmetric noise models are more common than symmetric noise models in many applications. In other words, it is more realistic to assume that the probability of the group testing results changes from positive to negative or from negative to positive is not the same.

Data forensics can compare files to see which files have been changed. In [23], the number of hashes for a collection of files is stored to confirm file changes. In addition, it is possible to know whether the hash has been changed by comparing values before and after the hash with the possibility of forgery. Ideas like this are examples of simple group testing problems. Compared to the group testing model, the changed file corresponds to the defect sample and the file collection corresponds to the group. The hash containing the changed file means that the group test result is positive. The file may be changed in a manner that does not change the hash value. In this sense, test results that must be positive can lead to different results with a specific probability. Although any long hash can be used to offset the noise effect, it is more appropriate to use a shorter hash than expected for efficient group testing.

III. NOISY THRESHOLD GROUP TESTING FRAMEWORK

A. Problem Statement

This section describes a noisy threshold group testing problem. First, an unknown signal $\mathbf{x}$ is expressed as a binary vector of size N, $\mathbf{x} = (x_1, x_2, \cdots, x_N)$, $\mathbf{x} \in \{0, 1\}^N$. For $i \in [N]$, x_i is the i-th element of $\mathbf{x}$. We present x_i in binary to indicate whether it is defective or not. That is, if the x_i sample is defective, it is expressed as $x_i = 1$, otherwise it is $x_i = 0$. In this paper, it is assumed that x_i has an independent probability distribution as follows,

$$\Pr(x_i = \alpha) = \begin{cases} 1-\delta & \text{if } \alpha = 0, \\ \delta & \text{if } \alpha = 1, \end{cases} \tag{1}$$

where δ is a defective rate and α is a dummy variable. Defective rate δ is less than 0.5, $0 < \delta < 0.5$, which is a small value for group testing problems.

The important construction of group testing is to decide which elements to join in a pool. To this end, we defined a group matrix. A group matrix has M rows and N columns, where $\mathbf{A}$ denotes a group matrix, $\mathbf{A} \in \{0, 1\}^{M \times N}$. If group testing is performed including the i-th sample x_i in the j-th testing, we can express it as $A_{ji} = 1$. Otherwise, $A_{ji} = 0$. Each element A_{ji} of group matrix indicates whether the corresponding sample is included in the testing or not as 0 or 1. Although the d-Separable matrix and the d-Disjunct matrix were used to design the group matrix, the approach of randomly selecting elements of the group matrix is known as a good design method [3]. In this paper, we assume that each element A_{ji} has the following probability with identically identical distribution,

$$\Pr(A_{ji} = \alpha) = \begin{cases} 1-\gamma & \text{if } \alpha = 0, \\ \gamma & \text{if } \alpha = 1, \end{cases} \tag{2}$$

where γ denotes the sparsity of the group matrix. High sparsity means that a large number of samples are included for group testing. On the other hand, if sparsity is low, it means that not many samples are included for group testing. From the point of view of performing group testing, high sparsity can increase the complexity of group testing. Therefore, if performance holds as the

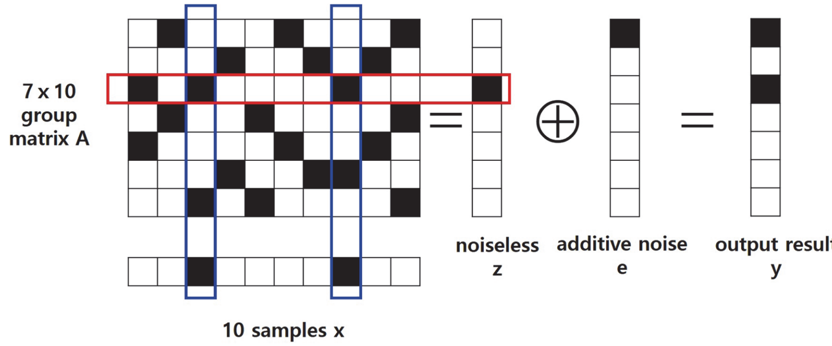

Fig. 1. One example for the noisy threshold group testing where $M = 7$, $N = 10$, $T = 2$. Black box denotes 1. White box denotes 0.

same, it is helpful for group testing design to use a group matrix with a low sparsity. In this paper, we consider how performance varies depending on the relationship between defective rate and sparsity.

Next, we describe more specifically why we consider a noisy threshold group testing model. Let us consider a model that could arise from diagnosis of COVID-19. As with COVID-19, there are cases where it is not possible to confirm whether it is positive or negative depending on the concentration of the virus. Testing for COVID-19 virus will reveal a positive result only if its concentration is above a certain level. During the initial stage of the diagnosis, the virus concentration is low. It might result in a false negative result for confirmed cases. Also, even if virus of COVID-19 is measured well, there are cases where the result is reversed due to measurement noise. We consider a noisy threshold group testing model. In other words, we consider a threshold group testing scheme because positive and negative cases can appear depending on virus concentration. We can consider a noise model because measurement noise can flip results. In a recent study [25], false positives and false negatives of COVID-19 testing results were reported to be between 0.1 and 4.5%, respectively. In this paper, we aim to analyze the theoretical performance of the noisy threshold group testing model for this mechanism.

Threshold group testing is called a different model from traditional group testing for the following reasons. In the traditional group testing, a positive result is obtained when one or more defective samples are present in the pool test. On the other hand, in threshold group testing, the result is positive when there are more than an arbitrarily determined number of defective samples where we define a threshold T. For example, a positive result only occurs if the test pool contains three defective samples in this case $T = 3$. If there is one defective sample in the pool, the result is negative. That is, to be positive, only when there are more than T defective samples in the pool, as in COVID-19 diagnosis, whether it is negative or positive depends on virus concentration.

In the traditional group testing, $T = 1$ is applied. The following equation mathematically expresses the result for the threshold group testing. Let z_j be the result of the j-th test pool without noise, where $z_j = 1$ is positive and 0 for negative result, $j \in [M]$, $\mathbf{z} = (z_1, z_2, \cdots, z_M)$.

$$z_j = \begin{cases} 0 & \text{if } \sum_{i=1}^{N} A_{ji} x_i < T, \\ 1 & \text{if } \sum_{i=1}^{N} A_{ji} x_i \geq T, \end{cases} \tag{3}$$

In this paper, we considered a noisy model in the threshold group testing problem. Noise assumes a model in which results are flipped. Suppose that noise causes z_j to convert from positive to negative model, the following noise model is defined.

$$\Pr(e_j = \alpha) = \begin{cases} 1-\eta & \text{if } \alpha = 0, \\ \eta & \text{if } \alpha = 1, \end{cases} \tag{4}$$

where η is measurement noise. We assume that e_j is independent of each other. Therefore, the j-th output y_j in the noisy threshold group testing model can be written as

$$y_j = z_j \oplus e_j \tag{5}$$

where the symbol $\oplus$ denotes XOR of logical operation. We denote $\mathbf{y} = (y_1, y_2, \ldots, y_M)$ and $\mathbf{e} = (e_1, e_2, \ldots, e_M)$.

Fig. 1 shows an example of the noisy threshold group testing. In this example, 2 out of 10 samples are defective. In noiseless, the vector $\mathbf{z}$ is $(0, 0, 1, 0, 0, 0, 0)$ for $T = 2$. Due to additive noise, the output is $\mathbf{y} = (1, 0, 1, 0, 0, 0, 0)$.

B. Bounds of Group Testing Schemes

Now let us consider theoretic performance bounds of traditional group testing schemes. The minimum number of tests required to find K defective samples among N samples using an adaptive group testing algorithm for perfect reconstruction is defined as m. We define the number of tests as $\overline{m}$ in the case that there is a non-adaptive algorithm. The number of tests required for a non-adaptive algorithm is less than the number of tests required for individual testing, $\overline{m} \le N$. Since the adaptive group testing performs the next test using the previous test result, the non-adaptive group testing always has a greater or equal number of tests than that of the adaptive group testing: $m \le \overline{m}$. If the number of defective samples is at least one, the following conditions are satisfied, $1 \le m$. As a result, the number of tests satisfies the following conditions, or vice versa. For the noise threshold group testing

$$1 \le m \le \overline{m} \le N. \tag{6}$$

If the group testing algorithm with an error rate of 0 by analysis using an information-theoretic bound is used, the minimum number of tests required to find K defective samples among N samples is as follows [3]:

$$M \ge \log_2 |\mathcal{S}|, \tag{7}$$

where $\mathcal{S}$ denotes a sample space. Furthermore, it is possible to obtain an information-theoretic bound even for a group testing algorithm with an error rate. It shows an upper bound on the probability of success depending on the number of tests. A group testing algorithm performed as the number of tests M satisfies the condition on the probability of success P_s to find defective samples [26]:

$$P_s \le M \log_2 \binom{N}{K}^{-1}. \tag{8}$$

Many algorithms have been developed for group testing. Among them, performance of the most efficient and practical algorithm is reviewed. Firstly, a binary splitting algorithm [3] has been used for adaptive models as a major approach for solving traditional group testing problems. The binary splitting algorithm is the simplest and most efficient. The required minimum number of tests M for the binary splitting algorithm is as follows,

$$M = \begin{cases} N & \text{if } N \le 2K-2, \\ (\log_2 \sigma + 2)K + p - 1 & \text{if } N \ge 2K-1, \end{cases} \tag{9}$$

where σ is the number of samples required for one test, p is uniquely determined nonnegative integer satisfying $p < K$. The definite defective algorithm [27] is designed to be suitable for non-adaptive models. The definite defective algorithm is an effort to eliminate false positives for non-adaptive group testing models. The definite defective algorithm provides the benefit that only false positives can occur without false negatives. Therefore, it is a useful algorithm in the area where group testing is exploited without false negatives. For a give N and K, the definite defective algorithm has the following lower bound on the performance for the number of tests M required when an error rate of E is allowed:

$$M \ge (1-\varepsilon) \log_2 \binom{N}{K}. \tag{10}$$

It can be seen that (7) and (10) converge to the same bound when the error rate goes to 0.

IV. BOUND IN NOISY THRESHOLD GROUP TESTING

A. Lower Bound using Fano's Inequality

Now let us consider the minimum number of tests required for solving an unknown signal in the noisy threshold group testing. To this end, we can use Fano's inequality theorem [28] in information theory. Fano's inequality is mainly used in channel coding theory. It describes the relationship between error rate and conditional entropy. In this paper, a lower bound on the probability of error is derived using Fano's inequality theorem

Theorem 1 (Fano's inequality [28]). *Suppose there are random variables A and B of finite size. If the decoding function Φ that finds A by measuring B is used, the following inequality holds,*

$$1 + P(\Phi(B) \ne A) \log_2 |A| \ge H(A|B), \tag{11}$$

where $P(\Phi(B) \ne A)$ is the probability of error for the decoding function Φ, the conditional entropy is defined as follows,

$$H(A|B) = \sum_{\alpha \in A} \sum_{\beta \in B} P_{AB}(\alpha, \beta) \log P_{A|B}(\alpha|\beta), \tag{12}$$

where P_{AB} and $P_{A|B}$ are joint probability and conditional probability, respectively.

In the noisy threshold group testing problem, we can obtain a lower bound on the probability of error. This lower bound shows the minimum number of tests required to reconstruct an unknown signal no matter which decoding function is used

Theorem 2. *For any decoding function, input signal defined in* (1) *and noise in* (4)*, a necessary condition for that the probability of error P_E is less such that*

$$\frac{NH(\delta) - M + MH(\eta) - 1}{N} < \rho, \tag{13}$$

where $H(\cdot)$ is the entropy function.

Proof of Theorem 2. Let $\hat{\mathbf{x}}$ be the estimated signal of $\mathbf{x}$ obtained from a decoding function. Using Markov chain, we can consider the following process, $\mathbf{x} \rightarrow (\mathbf{y}, \mathbf{A}) \rightarrow \hat{\mathbf{x}}$. For two conditional entropies, the following inequality holds,

$$H(\mathbf{x}|\mathbf{y}, \mathbf{A}) \leq H(\mathbf{x}|\hat{\mathbf{x}}). \tag{14}$$

Using the aforementioned Fano's inequality, it is written as follows:

$$H(\mathbf{x}|\mathbf{y}, \mathbf{A}) \leq 1 + P_E \log_2(2^N - 1). \tag{15}$$

The inequality on the probability of error is rewritten as follows:

$$P_E \geq \frac{H(\mathbf{x}|\mathbf{y}, \mathbf{A}) - 1}{N}. \tag{16}$$

Now what we need to deal with is conditional entropy $H(\mathbf{x}|\mathbf{y}, \mathbf{A})$. Let us write a conditional entropy using the following definition:

$$\begin{aligned} H(\mathbf{x}|\mathbf{y}, \mathbf{A}) &= H(\mathbf{x}) - I(\mathbf{x};\mathbf{y}, \mathbf{A}) \\ &= H(\mathbf{x}) - (I(\mathbf{x};\mathbf{A}) + I(\mathbf{x};\mathbf{y}|\mathbf{A})) \\ &\overset{(a)}{=} H(\mathbf{x}) - (H(\mathbf{y}|\mathbf{A}) - H(\mathbf{y}|\mathbf{A}, \mathbf{x})), \end{aligned} \tag{17}$$

where $I(.)$ is mutual information, equality (a) comes from the fact that $\mathbf{x}$ and $\mathbf{A}$ are independent of each other. Our aim is to have the minimum conditional entropy $H(\mathbf{x}|\mathbf{y}, \mathbf{A})$ so that the right side of (16) is minimized. It means that in (17), one conditional entropy $H(\mathbf{y}|\mathbf{A})$ is maximized, while conditional entropy $H(\mathbf{y}|\mathbf{A}, \mathbf{x})$ is minimized. Let us consider that each conditional entropy is either a maximum or a minimum:

$$\begin{aligned} H(\mathbf{y}|\mathbf{A}) \leq H(\mathbf{y}) &= H(\mathbf{z} \oplus \mathbf{e}) \\ &\leq M, \end{aligned} \tag{18}$$

where the first inequality is due to the definition of conditional entropy, the last inequality is due to facts that y_j is 0 or 1, y_j are independent, and the maximum binary entropy is 1. That is, Pr $(y_j = 0)$ = Pr $(y_j = 1)$. Now consider that the other conditional entropy $H(\mathbf{y}|\mathbf{A}, \mathbf{x})$ is a minimum.

$$\begin{aligned} H(\mathbf{y}|\mathbf{A}, \mathbf{x}) &= H(\mathbf{z} \oplus \mathbf{e}|\mathbf{A}, \mathbf{x}) \\ &= H(\mathbf{e}) \\ &= MH(\eta), \end{aligned} \tag{19}$$

where the second equality comes from that that randomness of $\mathbf{z}$ vanishes if $\mathbf{x}$ and $\mathbf{A}$ are known. The last equality is due to independent events of $\mathbf{e}$. Using (18) and (19), (17) can be rewritten as follows:

$$H(\mathbf{x}|\mathbf{y}, \mathbf{A}) \leq NH(\delta) - M + MH(\eta). \tag{20}$$

Finally, if (16) is changed to satisfy the condition for $P_E < \rho$ where ρ is a small and positive value, $\rho > 0$, the following condition holds,

$$\frac{NH(\delta) - M + MH(\eta) - 1}{N} < \rho. \tag{21}$$

□

B. Effect of Noise and Discussion

We can consider the result obtained from Theorem 2. First, Theorem 2 can be expressed as the ratio of the number of tests to the total number of samples as follows:

$$\frac{M}{N} > \frac{H(\delta) - \rho}{1 - H(\eta)}. \tag{22}$$

It is advantageous to use the noisy threshold group testing framework until N and M are equal. Otherwise, when $M > N$, we see that individual testing is more effective than group testing. Therefore, the noisy threshold group testing can be theoretically used under the following noise:

$$H(\eta) < 1 + \rho - H(\delta). \tag{23}$$

If we want to design a noisy threshold group testing framework, how to construct a group matrix is important. The solution to that is shown in the proof of Theorem 2. Let us look carefully at the conditions under which the inequality of conditional entropy holds in (18). Maximum conditional entropy $H(\mathbf{y}|\mathbf{A})$ is obtained when the following

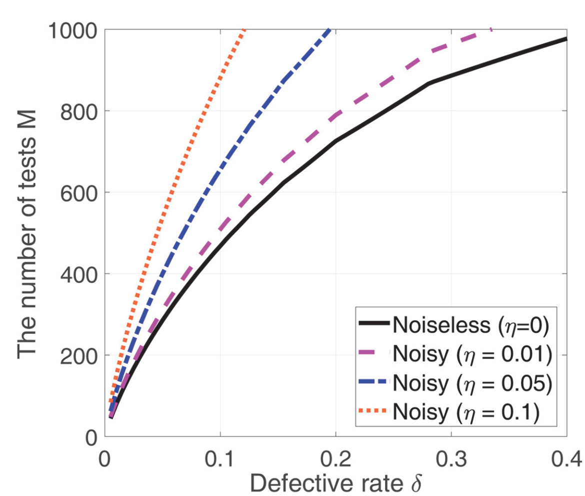

Fig. 2. The minimum number of tests M required to reconstruct unknown signals for the noisy threshold group testing with respect to different probabilities of noise, where $N = 1000$.

conditions are satisfied: Pr (y_j = 0) = Pr (y_j = 1). This means that the noisy threshold group testing system should be designed so that the output has an equal probability of 0 and 1. Since **x** and **A** are independent of each other, the probability of having an output of 0 is as follows:

$$\Pr(y_j = 0) = \sum_{t=0}^{T-1} \binom{N}{t} (\delta\gamma)^t (1-\delta\gamma)^{N-t} = \frac{1}{2}. \qquad (24)$$

As shown in (24), it can be seen that δ and γ have a trade-off with each other. In other words, when we want to reconstruct a sparse signal, we need to generate and use a high-density group matrix. Conversely, if the signal is not sparse, the group matrix should be designed with a low density.

V. CONCLUSION

In this paper, we considered a noisy threshold group testing model that was recently rediscovered in academic areas. The group testing was developed in World War II. Recently, there is a movement to exploit the theoretical idea of the group testing as a re-evaluation of related research. We considered the threshold group testing problem because positive and negative cases could occur depending on the virus concentration. We considered the noise model due to flipping results. As a result, using Fano's inequality, we obtained the minimum number of tests required for unknown signals in the noisy threshold group testing. We showed that the minimum number of tests was a function of defective rate and noise probability. We derived the lower bound for performance of the noisy threshold group testing framework with respect to noise intervals. In addition, the relationship between the defective rate of the input signal and the sparsity of the group matrix to design an optimal noisy threshold group testing system was presented.

ACKNOWLEDGEMENTS

This research was supported by Research Funds of Mokpo National University in 2020.

REFERENCES

1. R. Dorfman, "The detection of defective members of large populations," *The Annals of Mathematical Statistics*, vol. 14, no. 4, pp. 436-440, 1943.
2. D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289-1306, 2006.
3. D. Z. Du, and F. K. M. Hwang, *Pooling Designs and Nonadaptive Group Testing: Important Tools for DNA Sequencing*. Singapore: World Scientific Publishing, 2006.
4. C. M. Verdun, T. Fuchs, P. Harar, D. Elbrachter, D. S. Fischer, J. Berner, P. Grohs, F. J. Theis, and F. Krahmer, "Group testing for SARS-CoV-2 allows for up to 10-fold efficiency increase across realistic scenarios and testing strategies," *Frontiers in Public Health*, vol. 2021, article no. 1205, 2021. https://doi.org/10.3389/fpubh.2021.583377
5. L. Mutesa, P. Ndishimye, Y. Butera, J. Souopgui, A. Uwineza, R. Rutayisire, et al., "A pooled testing strategy for identifying SARS-CoV-2 at low prevalence," *Nature*, vol. 589, no. 7841, pp. 276-280, 2021.
6. P. Damaschke, "Threshold group testing," in *General Theory of Information Transfer and Combinatorics*. Heidelberg, Germany: Springer, 2006. pp. 707-718.
7. T. V. Bui, M. Kuribayashi, M. Cheraghchi, and I. Echizen, "Efficiently decodable non-adaptive threshold group testing," *IEEE Transactions on Information Theory*, vol. 65, no. 9, pp. 5519-5528, 2019.
8. J. T. Seong, "Theoretical bounds on performance in threshold group testing schemes," *Mathematics*, vol. 8, no. 4, article no. 637, 2020. https://doi.org/10.3390/math8040637
9. H. B. Chen and A. De Bonis, "An almost optimal algorithm for generalized threshold group testing with inhibitors," *Journal of Computational Biology*, vol. 18, no. 6, pp. 851-864, 2011.
10. G. De Marco, T. Jurdzinski, D. R. Kowalski, M. Rozanski, and G. Stachowiak, "Subquadratic non-adaptive threshold group testing," in *Fundamentals of Computation Theory*. Heidelberg, Germany: Springer, 2017, pp. 117-189.
11. A. Sterrett, "On the detection of defective members of large populations," *The Annals of Mathematical Statistics*, vol. 28, no. 4, pp. 1033-1036, 1957.
12. M. Sobel and P. A. Groll, "Group testing to eliminate efficiently all defectives in a binomial sample," *Bell System Technical Journal*, vol. 38, no. 5, pp. 1179-1252, 1959.
13. A. Allemann, "An efficient algorithm for combinatorial group testing," in *Information Theory, Combinatorics, and Search Theory*. Heidelberg, Germany: Springer, 2013, pp. 569-596.
14. J. N. Srivastava, *A Survey of Combinatorial Theory*. Amsterdam, Netherlands: North-Holland Publishing, 1973.
15. L. Riccio and C. J. Colbourn, "Sharper bounds in adaptive group testing," *Taiwanese Journal of Mathematics*, vol. 4, no. 4, pp. 669-673, 2000.
16. M. G. Leu, "A note on the Hu–Hwang–Wang conjecture for group testing," *The ANZIAM Journal*, vol. 49, no. 4, pp. 561-571, 2008.
17. C. L. Chan, P. H. Che, S. Jaggi, and V. Saligrama, "Non-adaptive probabilistic group testing with noisy measurements: near-optimal bounds with efficient algorithms," in *Proceedings of 2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, 2011, pp. 1832-1839.
18. G. K. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1880-1901, 2012.
19. M. B. Malyutov, "The separating property of random matrices," *Matematicheskie Zametki* (*Mathematical Notes)*, vol. 23, no. 1, pp. 84-91, 1978.
20. D. Sejdinovic and O. Johnson, "Note on noisy group testing:

asymptotic bounds and belief propagation reconstruction," in *Proceedings of 2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, 2010, pp. 998-1003.
21. D. Malioutov and M. Malyutov, "Boolean compressed sensing: LP relaxation for group testing," in *Proceedings of 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, 2012, pp. 3305-3308.
22. S. Bondorf, B. Chen, J. Scarlett, H. Yu, and Y. Zhao, "Sublinear-time non-adaptive group testing with O(k log n) tests via bit-mixing coding," *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1559-1570, 2021.
23. M. T. Goodrich, M. J. Atallah, and R. Tamassia, "Indexing information for data forensics," in *Applied Cryptography and Network Security*. Heidelberg, Germany: Springer, 2005, pp. 206-221
24. W. Kautz and R. Singleton, "Nonrandom binary superimposed codes," *IEEE Transactions on Information Theory*, vol. 10, no. 4, pp. 363-377, 1964.
25. D. A. Mistry, J. Y. Wang, M. E. Moeser, T. Starkey, and L. Y. Lee, "A systematic review of the sensitivity and specificity of lateral flow devices in the detection of SARS-CoV-2," *BMC Infectious Diseases*, vol. 21, article no. 828, 2021. https://doi.org/10.1186/s12879-021-06528-3
26. L. Baldassini, O. Johnson, and M. Aldridge, "The capacity of adaptive group testing," in *Proceedings of 2013 IEEE International Symposium on Information Theory*, Istanbul, Turkey, 2013, pp. 2676-2680.
27. M. Aldridge, L. Baldassini, and O. Johnson, "Group testing algorithms: bounds and simulations," *IEEE Transactions on Information Theory*, vol. 60, no. 6, pp. 3671-3687, 2014.
28. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2006.

Jin-Taek Seong

Jin-Taek Seong received the B.S. degree from the University of Seoul, Seoul, Korea, in 2006, the M.S. and the Ph.D. degree from the Gwangju Institute of Science and Technology, Gwangju, Korea, in 2008 and 2014, respectively. Since 2018, he is currently working as Associate Professor at the Department of Convergence Software, Mokpo National University in Korea. His research interests include information theory, machine learning, and coding theory.