Quick Review of Basic Pooled Testing Methods

Georg Pichler[1], Günther Koliander[2*]

[1] Institute of Telecommunications, TU Wien, Vienna, Austria, georg.pichler@ieee.org

[2] Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria, gkoliander@kfs.oeaw.ac.at

March 2020

1 Introduction

The method of pooled PCR testing for the SARS-CoV-2 virus is at the moment discussed in the media in several countries [1, 2]. This simple method can increase the testing capabilities by a factor of over 10 as long as the prevalence rate of an infection is below 1% and still by a factor of 2 if the prevalence rate is about 10%. We want to give some information-theoretic insight into the potential of this proposed diagnostic methodology and compare very basic schemes from the literature to ultimate limits. However, we do not have a medical background and thus some of our assumptions might be impractical or even absurd to practitioners. Our belief that pooled testing is possible in the current situation is mainly based on the study conducted in [3]. There, the pooling of nasopharyngeal/throat swab specimens for influenza virus testing by PCR is considered. It is shown to be feasible without increasing the probability of false negatives. We hope that these methods are already used in laboratory practice. If not, we strongly encourage to consider their adoption as they have the potential to increase testing rates significantly. We also hope for feedback, in particular from the medical community, to adapt our assumptions, give more realistic estimates for the potential benefits of this promising approach, and answer questions that go beyond the basics we describe below. Unfortunately, the first feedback we received, indicated that even pooling only 2 specimens already results in a false negative probability of 5% if only one of the specimens under test is in fact positive [4]. However, a pooling strategy might still help with increasing testing rates once other testing methods are available that can potentially benefit from the approach. Nevertheless, information-theorists should be encouraged by this work to look into the topic and develop new testing strategies that can be easily deployed, and mitigate the increased probability of false negatives.

The basic idea of pooled testing is quite simple: Instead of testing the nasopharyngeal/throat swab specimens of every person individually, the specimens of several people are mixed and the combined sample is subjected to the test. If the result is negative, then none of the specimens contains the virus and no more testing is required. If it is positive, then at least one specimen contains the virus and further tests need to be conducted to determine which one.

For simplicity, we assume that each specimen has the same probability p of containing the virus. We further assume that there is no dependence between specimens, i.e., the probability of any specimen containing the virus does not depend on which other specimens contain the virus. Our task is to conduct tests that allow us to unequivocally identify which specimens contain the virus, and which ones do not. The most trivial, but also least efficient approach is to test each specimen individually. This results in 1 test per specimen. If the probability of infection is very low, say $p = 0.01$, then it is more efficient to pool individuals into larger groups and test, e.g., 16 specimens at a time. If the test is negative, then only one test was required to determine that 16 patients are healthy. If the test is positive, we could split the group of 16 into two groups of 8 and repeat the process. This process can be iterated and will eventually identify any infected specimens.

*The work of Günther Koliander is supported by the Vienna Science and Technology Fund (WWTF): MA16-053.

2 Information Theoretic Limits

From an information-theoretic viewpoint, the task is to use as few tests as possible (on average), but still unequivocally identify which specimens contain the virus, and which ones do not. To this end, we identify each specimen with a Bernoulli random variable, i.e., a random variable X that is 1 with probability p (i.e, the specimen contains the virus) and 0 with probability $1-p$, where p is the probability of infection. We assume that the specimens are mutually independent and, thus, we have an identically and independently distributed (i.i.d.) sequence $(X_1, X_2, \dots X_k)$ of k Bernoulli-p random variables, corresponding to a group of k specimens under test. In classical information theory, we would now be allowed to ask arbitrary yes/no questions about this sequence and try to use as few questions as possible to deduce the infected specimens.[1] If we were allowed to do so, information theoretic results [5, Ch. 5] show that it is impossible to achieve this task with less than $H(p)$ questions per specimen where

$$H(p) = -p \log_2 p - (1-p) \log_2 (1-p). \tag{1}$$

Since tests are in fact very specific type of yes/no questions, for testing k individuals, any valid testing strategy will need at least $kH(p)$ tests on average. Although our choice of yes/no questions is restricted to pooled tests, as described above, we will see in the following that this restriction does not significantly impact performance when faced with low probabilities p. In fact, we can design simple tests that perform close to the optimal threshold.

Instead of the average number of tests performed per tested specimen, we will report our results in terms of an improvement factor $F = \frac{k}{T}$, where T is the expected number of performed tests, when testing k specimens. Thus, e.g., an improvement factor of 1.7 means that on average 170 people can be tested by performing 100 tests. By the preceding discussion, Shannon's source coding theorem implies that $F \leq F_{\max} = \frac{1}{H(p)}$ for any valid testing procedure. In our review of testing procedures, we focus on tests that require a minimum amount of bookkeeping and still provide good performance. Although there are more sophisticated schemes that allow for even better performance, we do not think that they are worth the effort in laboratory practice.

3 Dorfman Scheme—Two Stage Testing

This is the standard procedure of pooled testing that was already introduced in [6]. Here, a fixed number k of specimens $\{X_1, X_2, \dots, X_k\}$ are tested jointly. If the test is negative, no further testing has to be conducted. If the test is positive, all elements in the group are (re-)tested individually. The optimal group size k depends on the probability of infection p and is depicted in Fig. 1.

4 Square Matrix Scheme

An intuitive and fast method is the Square Matrix Scheme [7]. Here, all specimens in a given supergroup of size k^2 are arranged in a square matrix, i.e., the variables are indexed as $X_{i,j}$ with $i = 1, \dots, k$ and $j = 1, \dots, k$. Then the specimens belonging to the same row, i.e., $\{X_{i,1}, \dots, X_{i,k}\}$, are tested jointly for each $i = 1, \dots, k$ (i.e., we test which rows contain infected specimens). Similarly, the specimens belonging to the same column, i.e., $\{X_{1,j}, \dots, X_{k,j}\}$ are tested jointly for each $j = 1, \dots, k$ (i.e., we test which columns contain infected specimens). The elements in the submatrix of the rows and columns with positive outcome are then potentially positive and are tested one by one. The optimal value for k depends on the probability p and is depicted in Fig. 2.

The approach is illustrated in the following example. Let $k = 5$, i.e., we consider 25 specimens $X_{i,j}$. These specimens are arranged in the following matrix:
The colors indicate that the test of the groups Y_1 and Y_4 corresponding to the first and fourth column, respectively, were positively tested, as well as the groups Z_2 and Z_3 corresponding to second and third row, respectively. The specimens at all intersections, namely $X_{2,1}, X_{2,4}, X_{3,1}, X_{3,4}$ thus potentially contain the virus and are (re-)tested individually. In total, we conducted only $10 + 4 = 14$ tests for 25 specimens.

[1] Information theorists will see the similarity to variable-length source coding and we discuss this equivalence in slightly more detail in Appendix A.

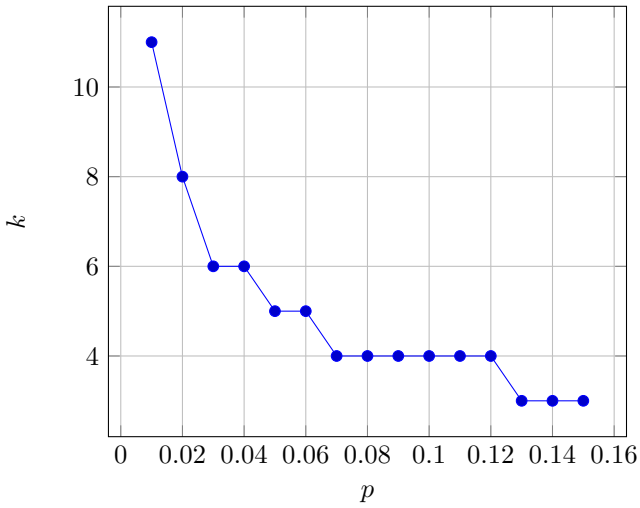


Figure 1: The optimal group size k for given probability of infection p for the Dorfman scheme.

	Y_1	Y_2	Y_3	Y_4	Y_5
Z_1	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$
Z_2	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$
Z_3	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$
Z_4	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$
Z_5	$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$	$X_{5,5}$

5 Multi Stage Testing

An extension of the Dorfman Scheme is to repeat the clustering procedure for positively tested groups, instead of testing individuals one-by-one [8]. More specifically, we start as in the Dorfman scheme and test a fixed number k of specimens $\{X_1, X_2, \ldots, X_k\}$ jointly. If the test is negative, no further testing has to be conducted. If the test is positive, we split the specimens $\{X_1, X_2, \ldots, X_k\}$ into l equally sized groups of size $\frac{k}{l}$, i.e., $\{X_1, X_2, \ldots, X_{\frac{k}{l}}\}$, $\{X_{\frac{k}{l}+1}, X_2, \ldots, X_{\frac{2k}{l}}\}$, ... These groups are then (re-)tested. We proceed with groups that test positive and reduce the size of the groups until it reduces to one.

For $l = 2$, corresponding to a binary search, we present the optimal initial group size k for given p in Fig. 3.

6 Comparison of the Schemes and Examples

In Fig. 4, we compare the schemes presented in the sections above with optimal parameter choices for various probabilities p. For comparison, we also depict the theoretically optimal upper bound $\frac{1}{H(p)}$ and the naive testing of each individual. One can see that already the simple Dorfman scheme increases the number of specimens that can be tested significantly, in particular, for low probabilities of infection p. From the more complicated schemes mainly the multi stage scheme give a significant improvement compared to the Dorfman scheme. The square matrix scheme has the benefit of being more efficient than the Dorfman scheme while also requiring just two stages. This might be important if each test takes a long time and one cannot wait for several stages to complete before performing further tests. While we calculated the optimal k and resulting improvement factor F exactly for the binary search and the Dorfman schemes, we had to resort to a Monte-Carlo simulation to obtain data for the matrix test.

As a simple example, we take the case of a testing scenario where we expect that $p = 0.1$. Here, we see that most schemes are very close to an improvement factor of 2.More precisely, we have an optimal limit of 2.13 while the Dorfman scheme with group size of 4 already obtains an improvement factor of 1.68. The binary search has an improvement factor of 1.96 for group size 6 which might still be convenient for testing and increases the number of

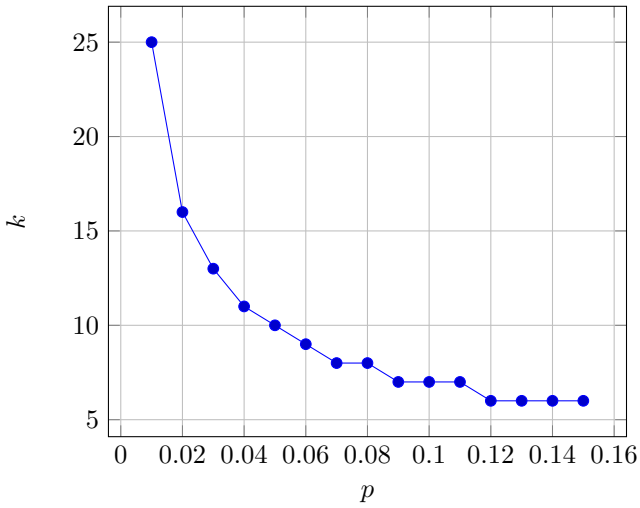


Figure 2: The optimal group size k for given probability of infection p for the Square Matrix scheme of supergroup size k^2.

persons tested on average by another 16.5% compared to the Dorfman scheme. A graph of the binary tree search in this case is depicted in Fig. 5.

As a second example, we take the case of a testing scenario where we expect that $p = 0.01$. Here, the differences in the schemes become more prominent. For the Dorfman scheme, we have an improvement factor of 5.11 for the optimal group size of 11 specimens. The square array test already gives an improvement factor of 7.3 for the optimal supergroup size of $k^2 = 625$ (i.e., groups of 25 specimens are tested jointly). The improvement factor also for smaller k is depicted in Fig. 6b. Finally, the binary test gives an improvement factor of 10.23 for optimal group size of 74. Note however, that already for smaller group sizes starting from $k = 40$ the improvement factor is larger than 10 and stays almost constant over a wide range of k as we can see in Fig. 6a.

If there are groups of specimens with very different assumptions about their probability of infection (e.g., routine testing of health personnel at an expected infection rate of $p = 0.01$ and testing of patients with symptoms with an expected infection rate of 0.1) then the easiest solution is to test these two groups separately using their individual optimal testing strategies.

7 Possible Extensions and Information Theoretic Insights

We note that all schemes presented above are not novel, but well established. For instance, [9] provides an overview of group testing strategies. We see that these schemes perform close to the best possible theoretic limit and more complicated schemes could only yield an additional performance gain of at most another 20% for the values of p considered here.

However, the theoretic limit was derived based on the assumption that the infections of individuals are independent. If we consider the more realistic assumption that the infections of people who, e.g., live in the same household, are tightly coupled (either all are infected or none), then the theoretical limit changes. We did not yet look into this topic in more detail and hope that other researchers will join us in the effort of designing more efficient testing strategies for the application of endemic monitoring where the infections in certain groups are highly correlated. Also a time variable can be introduced and dependencies from day to day tracked and optimal retesting scenarios, immunity, and other aspects can be modeled.

Another aspect is the error probability of these tests, i.e., a false positive test of a healthy group, or a false negative test of a contaminated group. If these probabilities are known, they can be taken into consideration and more accurate and efficient testing schemes can be designed. Only recently, the information-theoretic literature on group testing has been reviewed in [10]. Although there are many new insights and algorithms, we did not find any

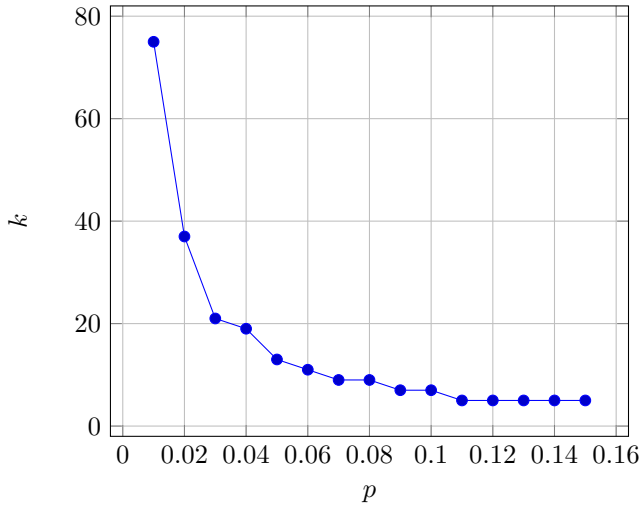


Figure 3: The optimal group size k for given probability of infection p for the Binary Searach Scheme.

practically feasible algorithms that are significantly more efficient than the ones discussed above. Also the focus in [10] is on identifying infected specimens with high probability and not with probability one. Although we agree that this is more reasonable from an information-theoretic viewpoint (and there is after all a positive probability for erroneous test results anyway), we think that in the current situation a testing methodology allowing for erroneous results will not be trusted.

Finally, the following questions that may be answered by specialists in other fields (either medicine, chemistry, or epidemiology/mathematics) are of interest to design even better testing strategies:

- How many specimens can maximally be in one group without affecting the accuracy of the test ([3] states at least 10)?
- Each specimen needs to be subdivided for potentially performing multiple test. How many samples can be obtained from one specimen without affecting the accuracy of each test? (We believe, at least two, as positive tests are repeated at the moment.)
- Information on probabilities of infection in specific groups that are to be tested.
- Information on dependencies between the infection status within a population.
- Conditional probabilities given symptoms (can be used as side information in coding strategies)?
- What is the probability of a false negative?
- What is the probability of a false positive?
- Are false positives/negatives mainly due to human or mechanical errors/mistakes when taking the sample or are these errors predominantly caused by random events during testing? How likely are different outcomes when testing the same specimen twice?

A Equivalence to Source Coding

From an information-theoretic viewpoint, the task of testing a sequence of specimens is equivalent to a variable-length source coding problem. There, we want to use as few bits as possible (on average) to encode a sequence of zeros and ones. To this end, we use short sequences to encode sequences that are very likely and longer ones for the unlikely cases. The bits correspond to the tests and the sequence of zeros and ones corresponds to the Bernoulli-p

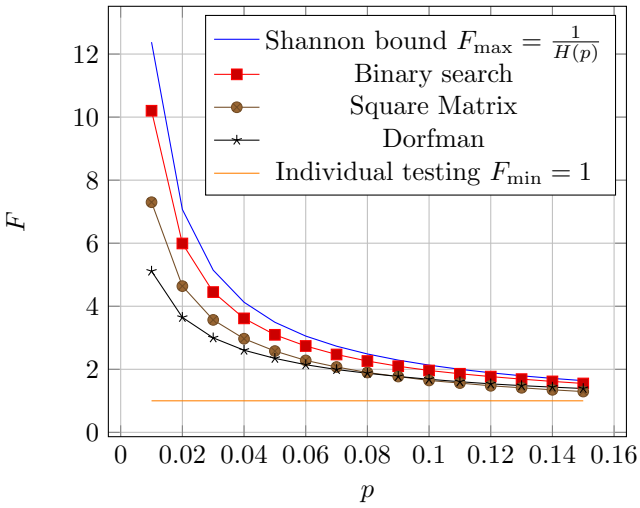


Figure 4: Comparison of the improvement factor $F = \frac{k}{T}$ for the schemes discussed, the theoretic upper bound $F_{\max}$ and the naive individual testing. In all schemes the optimal group size k is used.

sequence of specimens. Any valid testing procedure then corresponds to a particular source code, where the output of the code is given by the results of all the tests performed on a group of k specimens. As a valid testing procedure unequivocally identifies the infected specimens, the resulting code is a valid variable-length source code and its average codeword length is the average number of tests performed.

However, there is a small, but important difference between the source coding and pooled testing problems: In source coding, arbitrary "tests" with binary answers are allowed, whereas in pooled testing, we can only perform tests that lets us decide whether any of a given set of specimens contains the virus or none. Translated to our source-coding problem, this is equivalent to the max function, e.g., $\max\{X_4, X_7, X_{42}\}$ is one if and only if any one of X_4, X_7, X_{42} is one. We thus face a restricted source coding problem and cannot expect to be able to approach the optimal performance predicted by information theory.

References

[1] M. Beiglböck, P. Grohs, and W. Schachermayer, "Wie Corona Tests viel effizienter werden könnten," https://www.derstandard.at/story/2000116031860/wie-corona-tests-viel-effizienter-werden-koennten, 2020, [Online; accessed 26-March-2020].

[2] N. Jeffay, "To ease global virus test bottleneck, Israeli scientists suggest pooling samples," https://www.timesofisrael.com/to-ease-global-virus-test-bottleneck-israeli-scientists-suggest-pooling-samples/, 2020, [Online; accessed 26-March-2020].

[3] T. T. Van, J. Miller, D. M. Warshauer, E. Reisdorf, D. Jernigan, R. Humes, and P. A. Shult, "Pooling nasopharyngeal/throat swab specimens to increase testing capacity for influenza viruses by PCR," *Journal of clinical microbiology*, vol. 50, no. 3, pp. 891–896, 2012.

[4] personal communication with AGES (Österreichische Agentur für Ernährungssicherheit) experts, [27-March-2020].

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.

[6] R. Dorfman, "The detection of defective members of large populations," *The Annals of Mathematical Statistics*, vol. 14, no. 4, pp. 436–440, 1943.

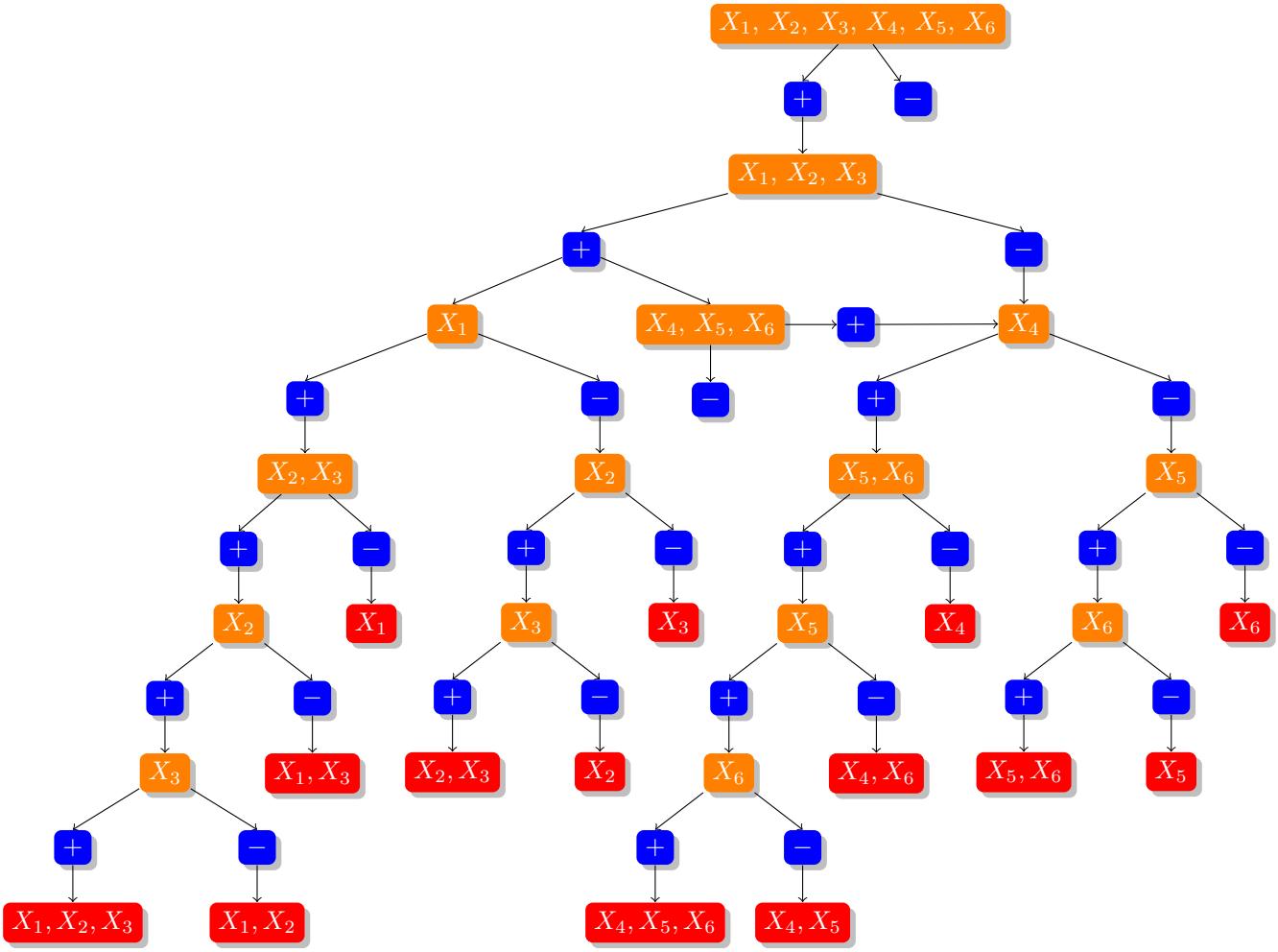
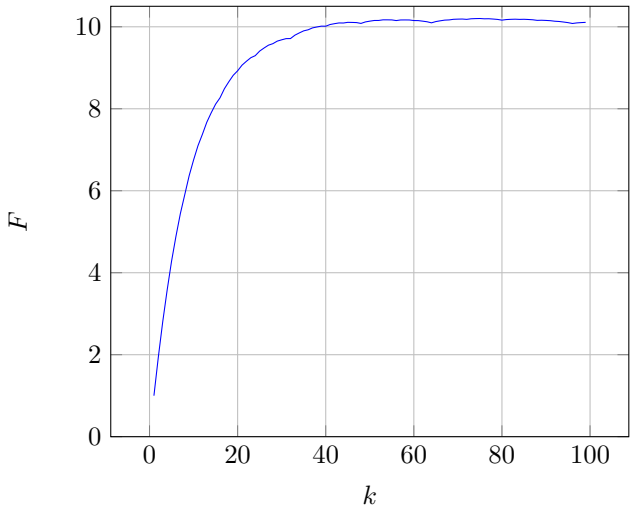
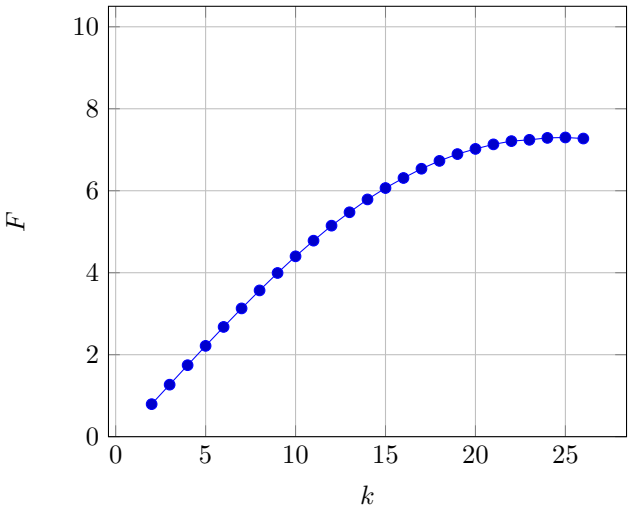


Figure 5: A graphical representation of the binary search for a group of 6 specimens. We start at the top: An orange box corresponds to a test of the indicated specimens. Depending on the outcome of the test (positive (+) or negative(−), we follow the arrows through the respective blue box to the next test(s). At one point a positive result yields two new tests. Here, we have to resolve both sides independently. If a test ends in a red box, the specimens indicated in the red box are infected. Some paths terminate in a blue (−) box and no further testing is needed along this path (most prominently, if the first test is negative no specimen contains the virus).



(a) Binary Search Scheme



(b) Square Matrix Scheme

Figure 6: Improvement factor F for fixed probability of infection $p = 0.01$ and various group sizes k.

[7] R. Phatarfod and A. Sudbury, "The use of a square array scheme in blood testing," *Statistics in Medicine*, vol. 13, no. 22, pp. 2337–2343, 1994.

[8] H. M. Finucan, "The blood testing problem," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 13, no. 1, pp. 43–50, 1964.

[9] D. Du and F. K. Hwang, *Combinatorial Group Testing and Its Applications*. World Scientific, 2000, vol. 12.

[10] M. Aldridge, O. Johnson, and J. Scarlett, "Group testing: an information theory perspective," *Foundations and Trends in Communications and Information Theory*, vol. 15, no. 3-4, pp. 196–392, 2019.