Some considerations about group testing

Julius Berner[1], Dennis Elbrächter[1], Philipp Grohs[1,2,3], and Pavol Harar[2,4]

[1]Faculty of Mathematics, University of Vienna,
[2]Research Platform Data Science, University of Vienna
[3]Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences
[4]Department of Telecommunications, Brno University of Technology

Abstract

This is a collection of visualizations illustrating the possible advantages that might be gained by different group testing methods. Despite the similar format, it is by no means a research paper but simply intended to present some information about group testing in an easily accessible way. It is also supposed to provide some theoretical background for the group-testing-simulations repository [4] used to produce the visualizations and its web application currently available at www.group-testing.com. This work was extended into a research article [9].

1 Methods

The following is intended to provide a brief overview over some basic group testing methods, which have been studied in the literature in connection with medical testing. Section 1.1 introduces the terminology we use, Section 1.2 describes the methods and provides some formulas, and Section 1.3.1 provides an example of how one could use the informative setting to gain some improvements in efficiency. Lastly, Section 1.3.2 considers a potentially new[1] non-adaptive method (i.e. all necessary tests can be performed in parallel), which is still more efficient than individual testing while introducing only a fairly minor amount of additional false positives.

The contribution of this work lies not only in providing a quick reference to anyone interested in group testing but also in providing precomputed results for the presented methods and a repository of source code [4] for parallel computation and comparative visualizations shown in the Appendices. The repository is easily runnable in the cloud using the Binder service[2]. Parts of the repository are based on the R package binGroup [2, 10]. There also exists a web interface for this package [1], which provides some of the functionalities of the package in an easily accessible manner.

1.1 Terminology

The following two key properties of a method constitute requirements that need to be achievable in order for the method to be feasible at all.

- **Divisibility**: This is the limit on how many times a given person can be tested, i.e. how many samples of each person must be available.
- **Number of Stages**: Different stages of a method must be performed sequentially as the tests performed in any stage depend on the results acquired in the previous stage. A method is called non-adaptive if it only requires a single stage, i.e. all required test can be performed in parallel. It is called adaptive if it requires more than one stage.

The following are quantities that determine how well a given method is expected to perform.

- **Infection rate** p: This is the assumed infection rate of the population which is tested. Unless more concrete information is available (see section on the informative setting), the commonly used assumption is that every person is positive with probability p, negative with probability $q = 1-p$, and that knowledge

Authors are listed in alphabetical order.
[1]In the sense that we have not managed to find in the literature.
[2]https://mybinder.org/

about the status of one person does not imply anything about the status of any other person (i.e. they are modeled as i.d.d. Bernoulli random variables).

- **Group size** k: This is the size of the groups which are used, meaning that it must be possible to combine k many samples, such that a reliable test can be performed on the pooled sample.

- **Sensitivity** S_e: Probability that the test returns a positive results, when applied to a positive sample (pool).

- **Specificity** S_p: Probability that the test returns a positive result, when applied to a negative sample (pool).

- **Positive predictive value PPV**: Probability that a person who tested positive, is actually positive. It is given by

$$PPV = \frac{pS_e}{pS_e + q(1 - S_p)}.$$

- **Expected number of tests per person** E: As measure of efficiency we consider the expected number of tests per person; for a discussion on alternate objective functions which take into account the decrease in specificity and sensitivity see [5]. The efficiency of a method will depend mostly on p and k as well as mildly on S_e and S_p.

An overview of the formulas presented in the following can be found in [7] and their derivation in [6]. The more general informative case is treated in [3, 8]. For ease of presentation we make the common assumption that sensitivity and specificity are the same for all group sizes considered. In case one has more specific information on how the sensitivity/specificity changes with the pool size k, one can simply use the same formulas but let S_e and S_p be dependent on k.
For brevity of notation we will write

$$f(k) = (1 - S_p)q^k + S_e(1 - q^k)$$

for the probability of a group of size k to test positive. In a slight abuse of notation, we will use S_e, S_p, and PPV to denote sensitivity, specificity, and positive predictive value of the test itself, while $S_e(X)$, $S_p(X)$, and $PPV(X)$ denote sensitivity, specificity, and positive predictive value of a method X.

1.2 Common methods

Note that the methods in this section correctly identify every person as either negative or positive, if the test itself is perfectly accurate. They do, however, aggregate the inaccuracies of the test itself, since every persons result will be based on multiple test. Appendix A contains plots for each of the following three methods which illustrate their efficiency for different infection rates and group sizes. Besides providing the optimal group sizes (restricted by a maximum group size of 20) for a selection of infection rates ranging from $p = 0.0037$ to $p = 0.2$, they also indicate the methods stability with respect to misestimation of p.

1.2.1 2-stage hierarchical testing (D2)

This is the most basic 2-stage group testing method due to Dorfman. It requires a divisibility of 2 as each person is first tested as part of a group of size k in the first stage and then, given their group tests positive, tested again individually in the second stage. The probability of a pool of size k to test positive is

$$f(k) = (1 - S_p)q^k + S_e(1 - q^k).$$

The characteristics of this method are given by

$$\begin{aligned} E(\mathrm{D2}) &= \tfrac{1}{k} + f(k), \\ S_e(\mathrm{D2}) &= S_e^2, \\ S_p(\mathrm{D2}) &= 1 - (1 - S_p)f(k - 1). \end{aligned}$$

1.2.2 3-stage hierarchical testing (D3)

This is a 3-stage method requiring divisibility 3. Each person is tested as part of a group of size k in the first stage. Every group which tests positive in this stage is split into subgroups, which are tested in the second stage. Every member of a subgroup that tests positive in the second stage, is tested individually in the third stage. For simplicity we present the formulas for the case where all subgroups are of size k^* (configurations with unevenly sized subgroups tend to be suboptimal anyway).
The characteristics this method are given by

$$\begin{aligned} E(\mathrm{D3}) &= \tfrac{1}{k} + \tfrac{f(k)}{k^*} + S_e^2(1 - q^{k^*}) + (1 - S_p)f(k - k^*)q^{k^*}, \\ S_e(\mathrm{D3}) &= S_e^3, \\ S_p(\mathrm{D3}) &= 1 - (1 - S_p)^2 f(k - k^*)q^{k^*-1} \\ &\quad - (1 - S_p)S_e^2(1 - q^{k^*-1}). \end{aligned}$$

Further iteration of this process leads to hierarchical s-stage methods. In order to be an improvement over D2 and D3 the initial group sizes would need to be too large and the infection rate too low, for it

to be likely to be relevant for medical testing. This, presumably, is also the reason why it is not implemented in the binGroup package. The formulas for the general case of hierarchical s-stage methods are, however, well known and can be found, e.g., in [7].

1.2.3 (Square) Array testing (A2)

This is a 2-stage method requiring divisibility $n+1$. The core idea of this method is to test every person twice in the first stage as a part of n many different groups of size k. In the second stage a person, for which all groups they were in tested positive, is tested individually. A usual way to illustrate this for $n = 2$, is as follows.
Take a set of k^2 many persons and arrange them in an $k \times k$ array, then every row and every column is grouped together and tested (seen Appendix D for some rudimentary illustration of this). This ensures that each person is tested exactly twice as part of a group of size k and constitutes the unique intersection of the two groups it belongs to (having groups whose intersection contains multiple elements would reduce the efficiency). Note that the important part is having this unique intersection property (which requires to collect at least k^2 many persons before grouping them in some manner), whereas the concrete procedure to achieve this will depend on the laboratory setting; i.e. while 'arranging them in a array and testing rows and columns' is the usual way to visualize it, the optimal procedure in practise will most likely be different.
In the case of $S_e = S_p = 1$ it is sufficient to only retest a person if both their row and column test positive, and one gets an expected number of tests per person of

$$E(\text{A2}) = \tfrac{2}{k} + p + q(1 - q^{k-1})^2.$$

In the presence of noise, however, one needs to account for the case where one only has positive rows but no positive columns and vice versa. As this makes the formulas a bit more involved we refer to [7] for this case.

1.3 Extensions

1.3.1 Combined testing

So far everything has been in the non-informative setting, meaning that the statuses of the persons are modeled as i.i.d. Bernoulli random variables (i.e. positive with probability p, negative with probability $1-p$, and completely independent of each other). This is commonly taken to to be the most reasonable assumption in case one has nothing but the infection rate p.
In the informative setting one assumes to have a more specific idea of how the probability of infection is distributed for the people to be tested. This may, of course, change the efficiency and optimal testing configuration of the various methods[3].
While an informative setting may be used to get more 'realistic' results for the methods described above, it might also be employed to modify the testing methods in order to potentially increase efficiency. We present a simple example of this. For a detailed treatment of informative hierarchical resp. array testing we refer to [8, 3].
We consider a scenario where we want to test some people with a high infection rate p_{high} (e.g. hospital staff) as well as a larger number of people with low infection rate p_{low} (e.g. general population). It turns out pooling a small number h of high-risk persons together with a group of low-risk persons can provide an increase in efficiency. This is shown in the plots in Appendix B. For each of the methods the efficiency of this combined testing is compared to the baseline where the high-risk group is tested using the optimal configuration (of the same method[4]) for p_{high} and the low-risk group is tested using the optimal configuration (of the same method) for p_{low}. The plots show the number h of high-risk persons which should be in a group of k many (otherwise low-risk) persons. Information on 'where to put the high-risk persons' (in e.g. A2) is provided by the binGroup package, but we have not visualized this here.

1.3.2 Non-adaptive array testing (A1)

All methods so far were adaptive, meaning they require at least 2 stages to be performed sequentially, which might not always be feasible. We now consider a non-adaptive method (in the non-informative setting) which, up to false negatives caused by the test itself, is guaranteed to find every positive person, but may create additional false positive. It turns out that it is capable of using significantly fewer tests than individual testing, while only creating a surprisingly small number of false positives. These non-adaptive methods could possibly also be used to save even more tests than the adaptive ones, but (as far as we can tell) only by

[3] As we have not investigated this sufficiently, we cannot comment of how much and in what manner it changes things. Not to mention, that even though its is clear that the infection probabilities are not distributed homogeneously among the population and there certainly are dependencies, deciding which model reflects reality is a very complicated and only mildly mathematical question

[4] Individual testing is chosen if it is more efficient than the method under consideration, which may be the case for sufficiently large p_{high}.

creating a significant amount of additional false positives.
The method is given by performing the first stage of A2, i.e. taking k^2 many persons arranged in a $k \times k$ array and testing every person as a part of n many suitably chosen groups. Then, a person is simply declared positive, if every group they were in tested positive. For $n = 2$ the trade-off between efficiency and additional false positives tends not to be that good. However, this improves significantly for bigger n. This requires choosing additional ways to partition the $k \times k$ persons into k many groups of size k, optimally in such a way that every person is the unique intersection of all n groups it belongs to. For a concrete way of how to do this for $n = 3$ and (if k is an odd number) for $n = 4$ see Appendix D. We get the following formulas

$$\begin{aligned} E(\text{A1}) &= \tfrac{n}{k} \\ S_e(\text{A1}) &= S_e^n, \\ S_p(\text{A1}) &= 1 - f(k-1)^n. \end{aligned}$$

Plots for this method can be found in Appendix C. Note that this methods requires divisibility n.

References

[1] Pooled testing. `https://bilder.shinyapps.io/PooledTesting`.

[2] C. R. Bilder, B. Zhang, F. Schaarschmidt, and J. M. Tebbs. bingroup: a package for group testing. *The R journal*, 2(2):56, 2010.

[3] M. S. Black, C. R. Bilder, and J. M. Tebbs. Optimal retesting configurations for hierarchical group testing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(4):693–710, 2015. `doi:10.1111/rssc.12097`.

[4] P. Harar, J. Berner, D. Elbrächter, P. Grohs, D. Fischer, T. Fuchs, F. Krahmer, F. J. Theis, and C. Verdun. Group testing simulations. `https://gitlab.com/hararticles/group-testing-simulations`, 2020.

[5] B. D. Hitt, C. R. Bilder, J. M. Tebbs, and C. S. McMahan. The objective function controversy for group testing: Much ado about nothing? *Statistics in Medicine*, 38(24):4912–4923, 2019. `doi:10.1002/sim.8341`.

[6] N. Johnson, S. Kotz, and X. Wu. *Inspection Errors for Attributes in Quality Control*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1991. URL: `https://books.google.de/books?id=gcskYPVNjCgC`.

[7] H.-Y. Kim, M. G. Hudgens, J. M. Dreyfuss, D. J. Westreich, and C. D. Pilcher. Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics*, 63(4):1152–1163, 2007. `doi:10.1111/j.1541-0420.2007.00817.x`.

[8] C. Mcmahan, J. Tebbs, and C. Bilder. Two-dimensional informative array testing. *Biometrics*, 68:793–804, 12 2011. `doi:10.1111/j.1541-0420.2011.01726.x`.

[9] M. Verdun, Claudio, T. Fuchs, P. Harar, D. Elbrächter, S. Fischer, David, J. Berner, P. Grohs, J. Theis, Fabian, and F. Krahmer. Group testing for sars-cov-2 allows for up to 10-fold efficiency increase across realistic scenarios and testing strategies. 2020, submitted. `doi:10.1101/2020.04.30.20085290`.

[10] B. Zhang, C. Bilder, B. Biggerstaff, F. Schaarschmidt, and B. Hitt. *binGroup: Evaluation and Experimental Design for Binomial Group Testing*, 2018. R package version 2.2-1. URL: `https://CRAN.R-project.org/package=binGroup`.

A Common methods

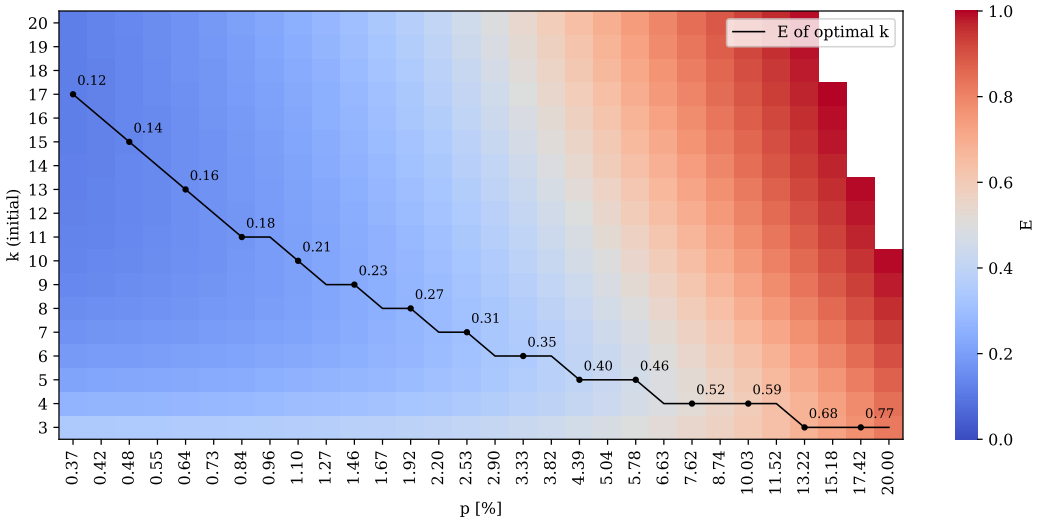

Figure 1: E(D2) for k and p, max k = 20

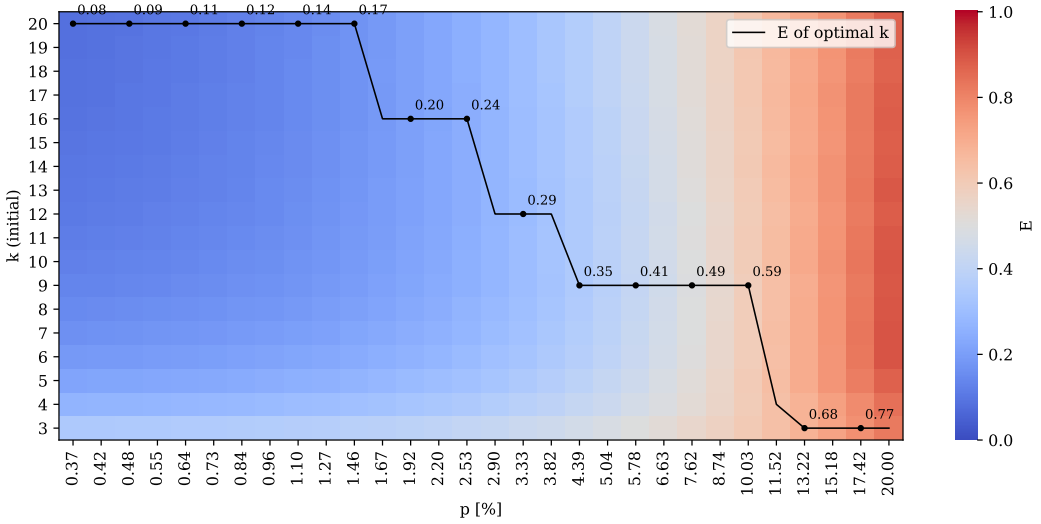

Figure 2: E(D3) for k and p, max k = 20

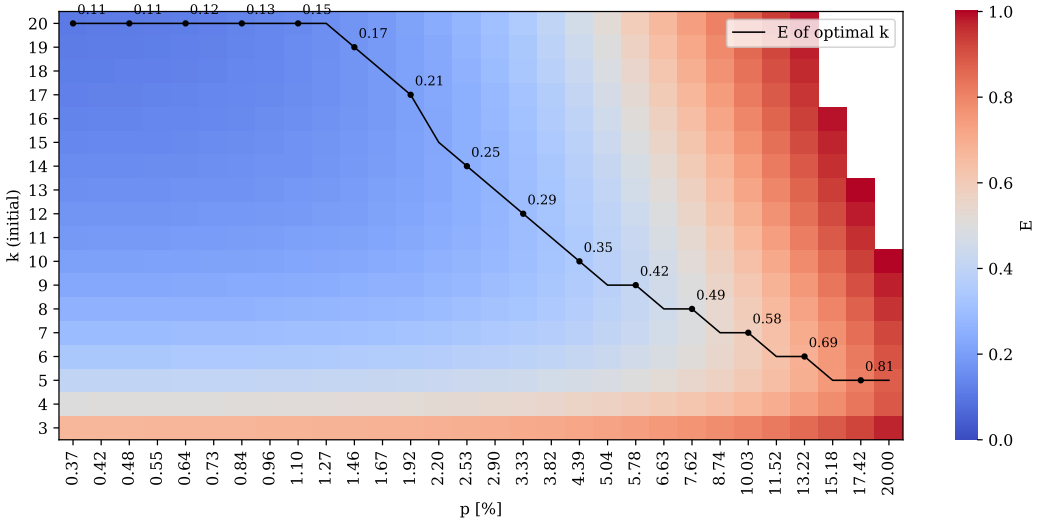

Figure 3: E(A2) for k and p, max k = 20

B Combined

In the following figures, h = 'no. of high-risk individuals per group', k / k_l / k_h = 'optimal group size' for combined / separate-low / separate-high testing, c / s = 'expected no. of tests per individual in combined / separate testing respectively', crossed-out squares denote cases where it is not beneficial to do the combined testing anymore. Combined testing is never beneficial in case of ID2.

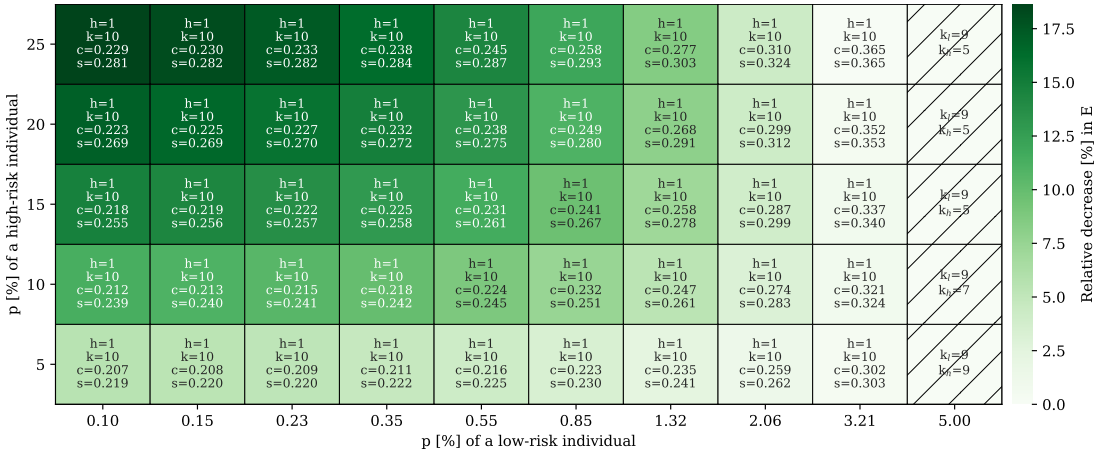

Figure 4: IA2 combined vs. A2 separate group testing, max k = 10

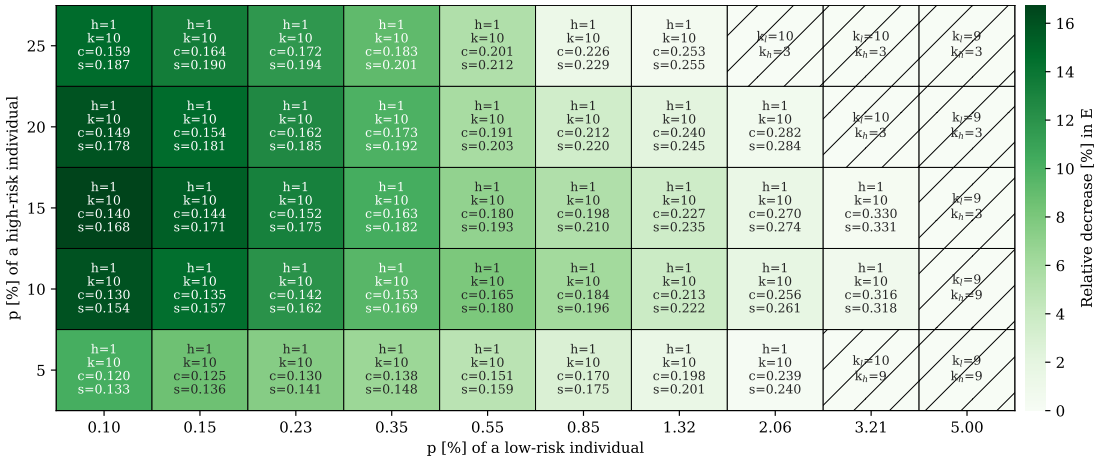

Figure 5: ID3 combined vs. D3 separate group testing, max k = 10

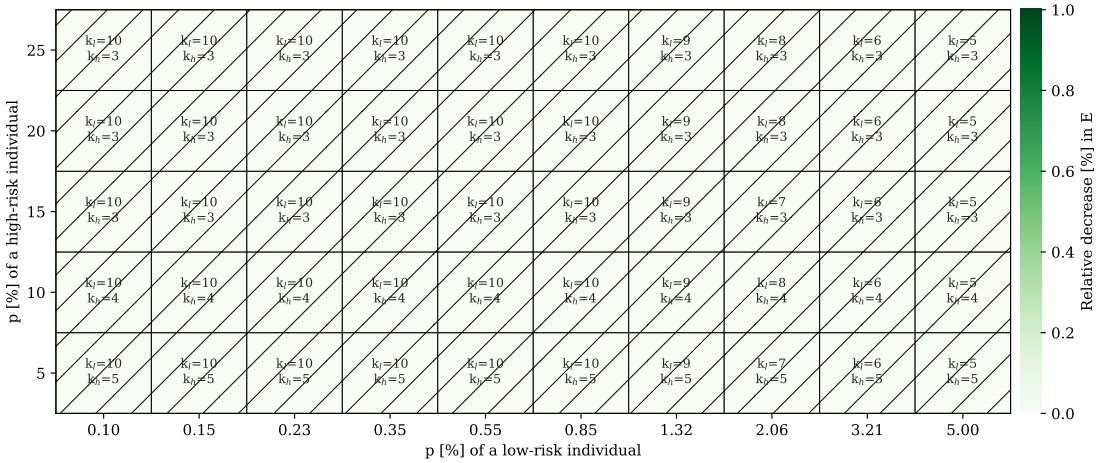

Figure 6: ID2 combined vs. D2 separate group testing, max k = 10

C Non-adaptive

The following figures show $E(\text{A1})$ for different group sizes k, infection rates p, and number of partitions $n \in \{2, 3, 4\}$. The maximal k was constrained to 20 and cells with $PPV(\text{A1})$ lower than a given threshold were masked.

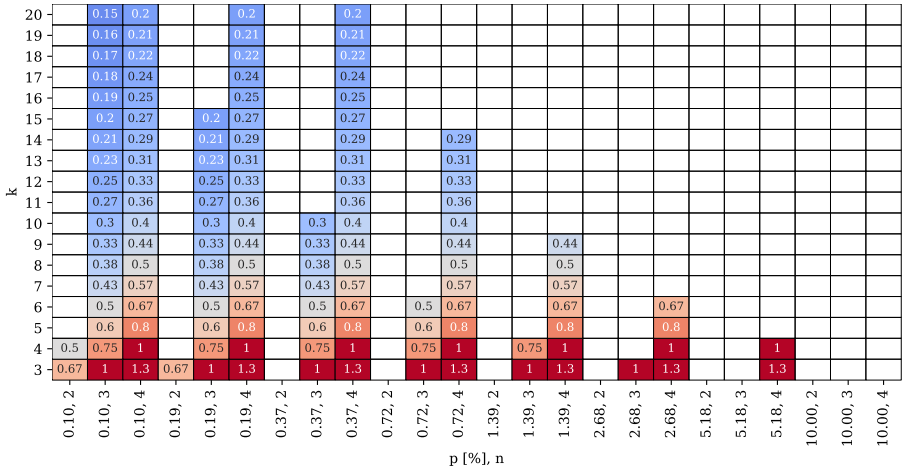

Figure 7: $E(\text{A1})$ for k, p, and $n \in \{2, 3, 4\}$ with $S_e = S_p = 1$ restricted to $PPV(\text{A1}) > 0.99$

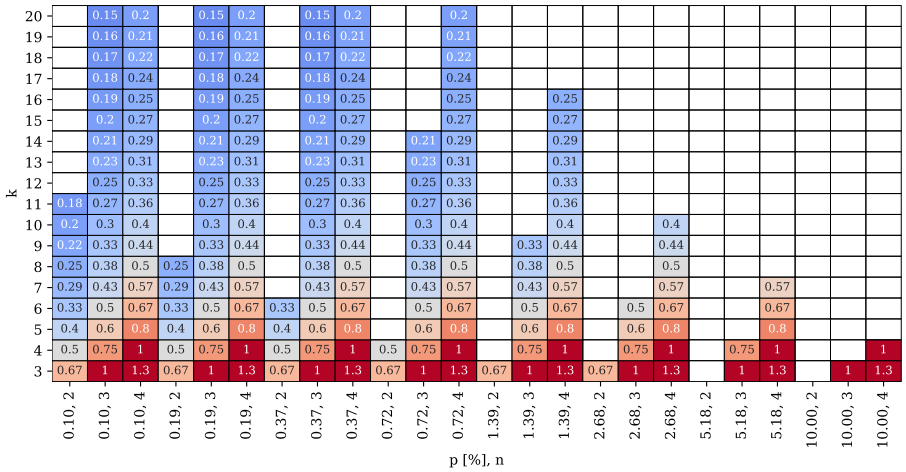

Figure 8: $E(\text{A1})$ for k, p, and $n \in \{2, 3, 4\}$ with $S_e = S_p = 1$ restricted to $PPV(\text{A1}) > 0.9$

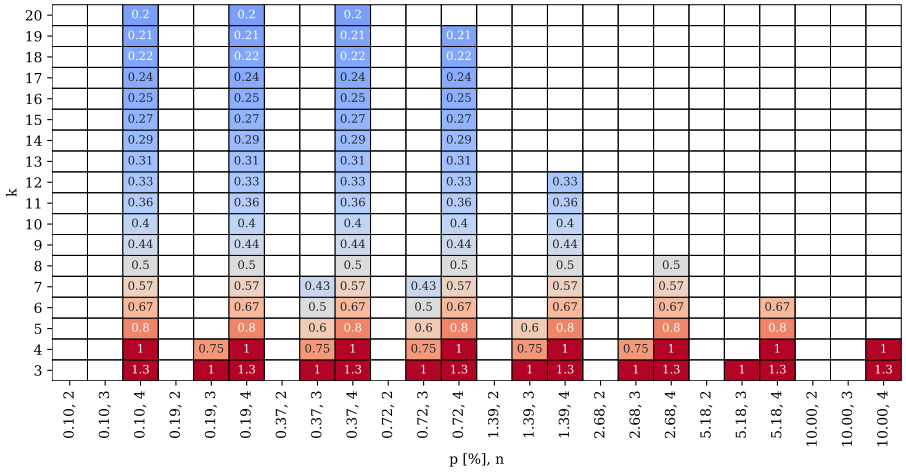

Figure 9: $E(\text{A1})$ for k, p, and $n \in \{2, 3, 4\}$ with $S_e = S_p = 0.95$ restricted to $PPV(\text{A1}) > 0.9$

D Array partitions

Example of 4 different ways of partitioning 25 many persons into different groups of size 5, such that every person is the unique intersection of all the groups they belong to. Note that taking rows, columns, and diagonals (i.e. $n = 3$) in this manner extends to all group sizes k, while taking rows, columns, diagonals, and counterdiagonals only extends to odd k (there may of course be other choices of partitions to achieve $n = 4$ for even k). Each of matrices below indicates one way to partition the persons and a given position in the matrix always corresponds to the same person. For each of the matrices (i.e. ways to partition) persons which have the same number at their position (in this matrix) are grouped together.
More concretely, let $x_{i,j}$ refer to the person corresponding to the entries in the i-th row and j-th column. Then $x_{1,1}$ is

- grouped together with $x_{1,2}, x_{1,3}, x_{1,4}, x_{1,5}$ in the 'rows' partition,
- grouped together with $x_{2,1}, x_{3,1}, x_{4,1}, x_{5,1}$ in the 'columns' partition,
- grouped together with $x_{2,2}, x_{3,3}, x_{4,4}, x_{5,5}$ in the 'diagonals' partition
- grouped together with $x_{5,2}, x_{4,3}, x_{3,4}, x_{2,5}$ in the 'counterdiagonals' partition.

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 & 4 \\ 5 & 5 & 5 & 5 & 5 \end{bmatrix}$$ 'rows'

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$ 'columns'

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 2 & 3 & 4 \\ 4 & 5 & 1 & 2 & 3 \\ 3 & 4 & 5 & 1 & 2 \\ 2 & 3 & 4 & 5 & 1 \end{bmatrix}$$ 'diagonals'

$$\begin{bmatrix} 5 & 4 & 3 & 2 & 1 \\ 4 & 3 & 2 & 1 & 5 \\ 3 & 2 & 1 & 5 & 4 \\ 2 & 1 & 5 & 4 & 3 \\ 1 & 5 & 4 & 3 & 2 \end{bmatrix}$$ 'counterdiagonals'