

ON THE GENERIC CUT-POINT DETECTION PROCEDURE IN THE BINOMIAL GROUP TESTING

U. ČIŽIKOVIEŇ¹, V. SKORNIakov^{2*}, §

ABSTRACT. Initially announced by Dorfman in 1943 for medical screening, (Binomial) Group Testing (BGT) was quickly recognized as a useful probabilistic tool in many other fields: quality control, communications and networking, engineering, statistics, etc. To apply any particular BGT procedure effectively, one first of all needs to know an important operating characteristic, the so called Optimal Cut-Point (OCP), describing the limits of its applicability. The determination of the latter is often a complicated task. In this work, we provide a generic algorithm suitable for a wide class of the BGT procedures and demonstrate its applicability by example. The way we do it exhibits independent interest since we link the BGT to seemingly unrelated field — the bifurcation theory.

Keywords: Group Testing, Cut-Point, Bifurcation Theory.

AMS Subject Classification: 37N99, 92F05, 94D99

1. INTRODUCTION

Group Testing (GT) refers to a set of methods for finding defectives in a given cohort of items. The first and likely the best known GT procedure was suggested by Dorfman [4]. During World War II, he sought for a cost-saving screening for Syphilis of US soldiers and proposed the following method. Given n blood samples, pool them and test. In case the pooled sample tests negative, finish; in case it tests positive, retest each individual.

Since appearance of that paper, many GT procedures were proposed. As the name suggests, they all share one feature: in the process of testing (at least in some stages), items are grouped and tested in pools. In other respects there can be significant differences. It is common practice to distinguish between Probabilistic Group Testing (PGT) and Combinatorial Group Testing (CGT). In PGT, one assumes that each tested item can be defective with some probability; in CGT, one assumes that in the tested cohort there are d defectives and the defectiveness generating mechanism is not considered. For a good

^{1,2} Vilnius University, Faculty of Mathematics and Informatics, Naugarduko Str. 24, Vilnius, LT-03225, Lithuania.

e-mail: ugne.cizikoviene@mif.stud.vu.lt; ORCID no. 0000-0001-5107-530X.

e-mail: viktor.skorniakov@mif.vu.lt; ORCID no. 0000-0001-7260-6931.

* corresponding author.

§ Manuscript received: August 30, 2024; accepted: November 11, 2024.

TWMS Journal of Applied and Engineering Mathematics, Vol.15, No.9; © Işık University, Department of Mathematics, 2025; all rights reserved.

review of some of the PGT procedures, we refer to Malinovsky [7]; Aldridge et. al. [1] covers CGT and provides some other classification criteria as well.

In this paper, we focus on the subclass of PGT called the Binomial Group Testing (BGT). In BGT, one restricts PGT by further assuming that:

- (BGT1) each item is independent of the rest;
- (BGT2) the probability of defectiveness $p \in (0, 1)$ is the same for all items in the tested cohort.

An intuition suggests that PGT makes sense only when the defectiveness is rare. In the frame of BGT, this concept is formalized as follows. Given BGT procedure X , denote:

- $T_X(n, p)$ = "the random number of tests utilized per cohort spanning n items";
- $M_X(n, p) = \mathbb{E} T_X(n, p)$ = "average number of tests utilized per cohort spanning n items";
- $t_X(n, p) = \frac{M_X(n, p)}{n}$ = "number of tests spent per single item on average when testing a cohort spanning n items".

One says that, for a fixed p , procedure X makes sense provided $\inf_{n \in \mathbb{N}} t_X(n, p) < 1$. For example, in the above described Dorfman procedure D , one often ends up with a single test whenever the prevalence of the disease p is small; assuming an opposite (i.e., large p), $M_D(n, p) \approx n + 1$ this way making procedure insensible. The same logic applies to other procedures as well. Therefore, one can expect that, for large p 's, any BGT procedure makes no sense. This indeed holds true. Namely, there is a fundamental result due to Ungar [13]. It says that, for $p > \frac{3-\sqrt{5}}{2}$, $\inf_{n \in \mathbb{N}} t_X(n, p) > 1$ for any BGT. In words, under BGT assumptions (BGT1)–(BGT2), it is optimal to test items individually and do not invoke GT if $p > \frac{3-\sqrt{5}}{2}$. On the other hand, in the same paper, it was shown that, for $p < \frac{3-\sqrt{5}}{2}$, individual testing is not optimal and there always exists a BGT procedure with $\inf_{n \in \mathbb{N}} t_X(n, p) < 1$. In the GT literature, the value $\frac{3-\sqrt{5}}{2}$ is termed the Universal Cut-Point (UCP). However, it is often the case that particular procedure has its own Optimal Cut-Point (OCP) $p_{X,c} < \text{UCP}$. The detection of this operational characteristic is an important and usually not a straightforward task. In this paper, we describe a generic algorithm for finding of $p_{X,c}$ suitable for a wide class of BGT procedures.

The remaining part of the paper contains two sections and two appendixes. In Section 2, we state our main result. In Section 3, we provide examples of applications and a short accompanying discussion. In appendixes, we place proofs and figures.

2. RESULTS

2.1. Notions and Assumptions. In the remaining part of the paper, X stands for the unspecified (generic) BGT procedure. Corresponding quantities are sub-scripted: $M_X(n, p), t_X(n, p)$, etc. When turning to the particular procedure, X is replaced by another letter. $\text{UCP} := \frac{3-\sqrt{5}}{2}$ stands for the previously mentioned Universal Cut-Point. For some procedures, $n \in \mathbb{N}$ is not an actual size of the tested cohort $N(n)$ but a parameter. For example, in case of procedure described in Subsection 3.2, the size of the tested cohort $N(n) = n^2$ for $n \in \{2, 3, \dots\}$. Such parametrization redefines $t_X(n, p)$ making it equal to $\frac{M_X(n, p)}{N(n)}$ yet it does not affect our results. Therefore, we retain all previous notions unchanged. When dealing with a particular procedure, the role of n is unambiguous.

In what follows, we focus on the subset of the BGT procedures with a mean $M_X(n, p)$ satisfying assumptions listed below.

- (M0) $\exists c \geq 2$ s.t. X is a-priori known to be useless for $n \in [1, c)$.

- (M1) Function $\mathbb{N} \times (0, \text{UCP}] \ni (n, p) \mapsto M_X(n, p)$ can be treated as a continuous function on $[c, \infty) \times (0, \text{UCP}]$ differentiable in the whole interior of its domain.
 (M2) $\forall n \in (c, \infty)$ function $(0, \text{UCP}] \ni p \mapsto M_X(n, p)$ is strictly increasing.
 (M3) $\forall n \in (c, \infty) t_X(n, \text{UCP}) > 1$.
 (M4) $\forall n \in (c, \infty) \exists p \in (0, \text{UCP}) : t_X(n, p) < 1$.

(M0) may be viewed as a condition required to restrict the range of the dynamical system we make use of when looking for the OCP $p_{X,c}$. Together with (M3), it also defines the boundary value for determination of $p_{X,c}$ for the case $p_{X,c} = \text{UCP}$ (see examples and discussion in Section 3). In a usual case, one can set $c = 2$ and have it since, in this context, $c = 2$ means that taking one item we do not need any GT procedure: to identify the defectiveness of a single item, one always needs one test.

Constraint (M1) is the most restrictive since not all BGT procedures can be naturally extended to have a differentiable mean with respect to both arguments (in case of p this always holds true). However, this particular assumption is the one we rely on. It also enables us to draw the connection with the bifurcation theory.

Other constraints can be justified naturally and attributed to many of the BGT procedures in general.

(M2) states that an average number of tests per batch spanning n items should increase together with the rate of defectiveness. For justification, we mention another fundamental result due to Yao and Hwang [15] who have demonstrated that, $\forall n \in \mathbb{N}$, function $(0, \text{UCP}] \ni p \mapsto \inf_X M_X(n, p)$ is strictly increasing (with an infimum being taken over all possible BGT procedures).

Finally, (M4) in technical terms states that we focus on the procedures applicable to any number of tested items at least for some p 's in the range of their sensibility. This is very often the case since many procedures are suitable for large scale testing when the rate of defectiveness is small.

2.2. Statement. Our results are given in two propositions stated below. The first one characterizes properties of the OCP.

Proposition 2.1. *Assume (M0)–(M4). Let $p_{X,c} = \sup\{p \in (0, \text{UCP}) \mid \exists n \in (c, \infty) : t_X(n, p) < 1\}$. Then $\forall p \in (0, p_{X,c})$ procedure X makes sense on the continuous scale; $\forall p \in (p_{X,c}, \text{UCP}]$ it makes no sense at all, that is,*

$$(n, p) \in (c, \infty) \times (p_{X,c}, \text{UCP}] \Rightarrow t_X(n, p) > 1. \quad (1)$$

The second proposition demonstrates that, under (M0)–(M4), there exists a generic procedure for finding $p_{X,c}$, and it can be naturally cast in terms of the bifurcation theory as follows. Treating $p \in (0, \text{UCP}]$ as a control parameter and $n \in (c, \infty)$ as a function of some latent continuous argument, consider autonomous dynamical system

$$\dot{n} = t_X(n, p) - 1. \quad (2)$$

Proposition 2.2. *Assume (M0)–(M4). $p_{X,c}$ is a bifurcation point of the system (2) and one can distinguish between three types of possible bifurcations.*

(b0) $p_{X,c}$ is the only value of the control parameter for which (2) admits fixed points in (c, ∞) . In this case $p_{X,c} < \text{UCP}$ and all $n \in (c, \infty)$ solve $t_X(n, p_{X,c}) = 1$.

If there exists $p_l \in (0, p_{X,c})$ for which (2) admits a fixed point $n \in (c, \infty)$, there are two possibilities:

(b1) (2) has fixed points in (c, ∞) for all $p \in [p_l, p_{X,c})$ yet there are no fixed points corresponding to $p_{X,c}$;

(b2) (2) has fixed points in (c, ∞) for all $p \in [p_l, p_{X,c}]$ including $p_{X,c}$ which then is necessary smaller than UCP.

In all cases, bifurcation curve induces a differentiable map $(c, \infty) \ni n \mapsto p_n \in (0, p_{X,c}]$. Therefore, $p_{X,c}$ can be determined by finding its maximum. For bifurcations of types (b0) and (b2), this amounts to solving a system

$$\begin{cases} t_X(n, p) = 1, \\ \frac{\partial}{\partial n} t_X(n, p) = 0 \end{cases} \quad (3)$$

(with respect to both n and p) and then picking up a largest p value from the set $S = \{(n, p) \in (c, \infty) \times (0, \text{UCP}] \mid (n, p) \text{ solves (2)}\}$. For the bifurcation of type (b1), $p_{X,c} = \max(\lim_{n \rightarrow c+} p_n, \lim_{n \rightarrow \infty} p_n)$. In particular, this holds true when (3) has no solution lying in $(c, \infty) \times (0, \text{UCP}]$.

Before proceeding to examples, we provide several remarks.

Remark 2.1. Prop. 2.2 establishes the procedure for finding OCP on the continuous scale (COCP). In practice, one operates on the discrete one since the number of tested items n is integer. As a rule, discrete scale OCP (DOCP) is lower than $\text{COCP} = p_{X,c}$ of Prop. 2.1. However, the difference is usually small (see examples in Section 3) whereas the determination of DOCP is often times quite involved (check the references provided in Section 3). Moreover, in case of (b2), DOCP can often be recovered as follows:

- take n_c s.t. $(n_c, p_{X,c})$ solves (3);
- put $\text{DOCP} = \max(p_{\lfloor n_c \rfloor}, p_{\lceil n_c \rceil})$.

In case of (b1), DOCP is very likely to coincide with COCP.

Remark 2.2. We are inclined to think that bifurcations (b1)–(b2) are the prevalent ones since we are unaware about practical examples of (b0) satisfying our assumptions. Yet Subsection 3.5 contains an example showing that a counterpart of (b0) may occur on the discrete scale. We were unable to exclude this type theoretically. Thus, appealing to the mentioned example, we are inclined to think that (b0) is not a redundant case but an exceptional one, corresponding to optimal procedures (see the discussion Subsection 3.6).

Remark 2.3. We did not analyze the behaviour of the lower cut-point $\text{LCP} := p_l$ of Prop. 2.2 because our intent was to consider procedures making sense for all p 's in $(0, \text{OCP})$. In certain cases, the behaviour of p_l can be of primary importance. For example, Yao and Hwang [16] investigated pairwise BGT algorithm which was known to perform optimally only for p 's lying sufficiently far away from origin. Hence, in their case, p_l (not $p_{X,c}$) was the primary quantity of interest.

3. EXAMPLES AND DISCUSSION

In this section, we provide several examples demonstrating applications of Prop. 2.2. We also provide a couple of examples of the procedures violating our conditions. For the sake of convenience, we put $q := 1 - p$ (and parameterize by this parameter as well).

3.1. Dorfman procedure D. We have already described Dorfman procedure in the Introduction 1. Its description implies that

$$\begin{aligned} M_D(n, p) &= 1 \cdot q^n + (n+1)(1 - q^n) = n+1 - nq^n, \\ t_D(n, p) &= \frac{M_D(n, p)}{n} = 1 + \frac{1}{n} - q^n. \end{aligned} \quad (4)$$

We put $c = 2$ in (M0). (M1) obviously holds. Since $\frac{\partial}{\partial p} t_D(n, p) = nq^{n-1} > 0$ for all $n \in (2, \infty)$, (M2) holds as well. As for (M3), note that

$$\frac{d}{dn} t_D(n, \text{UCP}) = \frac{d}{dn} \left(1 + \frac{1}{n} - \left(\frac{\sqrt{5}-1}{2} \right)^n \right) = -\frac{1}{n^2} - \left(\frac{\sqrt{5}-1}{2} \right)^n \ln \left(\frac{\sqrt{5}-1}{2} \right).$$

Equating this to zero (and solving numerically¹) one finds out that this function has a unique minimum at $n_{\min} \approx 2.888$ and $\min_{n>2} t_D(n, \text{UCP}) \approx t_D(2.888, \text{UCP}) = 1.097$. Moreover, it has maximum at $n_{\max} \approx 5.75$ and then decreases to $\lim_{n \rightarrow \infty} t_D(n, \text{UCP}) = 1$. Finally, from (4) it follows that

$$\forall n \in (2, \infty) \quad \lim_{p \rightarrow 0+} t_D(n, p) = \frac{1}{n}.$$

Therefore, (M4) holds as well.

Figure 1 shows a plot of the inverted bifurcation map $n \mapsto p_n$ described in Prop. 2.2. In this case, it admits analytical expression: $p_n = 1 - \left(\frac{1}{n}\right)^{\frac{1}{n}}$. The latter follows directly from (2) by equating it to zero and solving w.r.t. $p = 1 - q$. System (3) is given by

$$\begin{cases} \frac{1}{n} = q^n, \\ -\frac{1}{n^2} = q^n \ln q, \end{cases} \quad (5)$$

and can be solved analytically too. Its solution is $(n^*, p_{D,c}) = (e, 1 - e^{-e^{-1}})$. Samuels [9] investigated procedure D on the discrete scale. His analysis revealed that DOCP = $1 - 3^{-3^{-1}}$. Our COCP is quite close. Moreover, we can recover DOCP by applying the method described in the Remark 2.1. In this particular case the method works and affirms Samuels' result.

3.2. Squared Array procedure A2. Squared Array (A2) procedure was introduced by Phatarfod and Sudbury [8] (and later generalized by Berger, Mandell and Subrahmanya [2]). To apply A2, one places $n^2, n \in \{2, 3, \dots\}$ items on $n \times n$ matrix and applies the following tests:

- n tests of pooled samples corresponding to row pools;
- n tests of pooled samples corresponding to column pools;
- individual tests on items having positive both row and column tests.

This way all defectives get identified.

It was shown by Phatarfod and Sudbury [8] that

$$M_{A2}(n, p) = 2n + n^2 (1 - 2q^n + q^{2n-1}). \quad (6)$$

Noting that the tested cohort has $N(n) = n^2$ items in total, we therefore have that

$$t_{A2}(n, p) = \frac{2}{n} + 1 - 2q^n + q^{2n-1} = \frac{2}{n} + (1 - q^n)^2 + pq^{2n-1}. \quad (7)$$

From the latter expression, it follows that A2 makes sense only for $n > 2$. Bearing in mind practical aspect (i.e., the fact that cohort sizes are integers) we therefore set $c = 3$ in (M0). Note that, due to the design of the procedure, this truncation actually restricts

¹calculations were accomplished by making use of SciPy [14] — a package for Scientific Computing in Python

sizes of the tested cohorts to start from $9 = 3 \times 3$ (not 3). It is obvious that (M1) holds. Since $\forall p \in (0, \text{UCP}]$

$$\frac{\partial}{\partial p} M_{A2}(n, p) = n^2 (2nq^{n-1} - (2n-1)q^{2n-2}) = n^2 (q^{2n-2} + 2nq^{n-1}(1 - q^{n-1})) > 0,$$

(M2) holds as well.

Justification of (M3) can be done by accomplishing the following steps:

- check that $\frac{\partial}{\partial n} t_{A2}(n, p) = -\frac{2}{n^2} - 2q^n \ln q(1 - q^{n-1})$ and $\frac{\partial^2}{\partial n^2} t_{A2}(n, p) = \frac{4}{n^3} - 2q^n \ln^2 q(1 - 2q^{n-1})$;
- numerically solve $\frac{\partial^2}{\partial n^2} t_{A2}(n, p) \Big|_{p=\text{UCP}} = 0$ and obtain two roots: $n_1 \approx 5.278, n_2 \approx 9.448$;
- verify that n_1 corresponds to maximum whereas n_2 corresponds to minimum of $n \mapsto \frac{\partial}{\partial n} t_{A2}(n, \text{UCP})$ and that $\frac{\partial}{\partial n} t_{A2}(n_1, \text{UCP}) < -0.0055 < 0$;
- conclude that $n \mapsto t_{A2}(n, \text{UCP})$ is decreasing and (M3) holds since

$$\lim_{n \rightarrow \infty} t_{A2}(n, \text{UCP}) = 1.$$

Finally, from the last expression given in (7), it follows that

$$t_{A2}(n, p) \xrightarrow{p \rightarrow 0+} \frac{2}{n} \leq \frac{2}{3}.$$

Hence (M4).

Figure 2 shows the inverted bifurcation curve $(3, \infty) \ni n \mapsto p_n$ of Prop. 2.2. System (3) is given by

$$\begin{cases} \frac{2}{n} - 2q^n + q^{2n-1} = 0, \\ -\frac{1}{n^2} - q^n(1 - q^{n-1}) \ln q = 0. \end{cases}$$

As can be seen from the curve, it has a unique solution $(n^*, p_{A2,c}) \approx (4.454, 0.252)$. Kim and Hudgens [5] have investigated A2 on the discrete scale. They have proved that $\text{DOCP} = 0.2498$. This point precisely coincides with $p_{\lceil n^* \rceil}$. Thus, the method described in 2.1 again led to the recovery of the DOCP.

3.3. Modified Dorfman procedure MD. The original Dorfman procedure D is inconsistent: if the pooled sample tests positive and all but the last individual items test negative, it still tests the last one. In such case, Sobel and Groll [10] suggested not to test the last item since test result can be inferred. We call their procedure Modified Dorfman (MD). For MD,

$$t_{MD}(n, p) = 1 - q^n + \frac{1 - pq^{n-1}}{n}. \quad (8)$$

We set $c = 2$ and (M0) is satisfied. As usually, (M1) is obvious. Since

$$\frac{\partial}{\partial p} t_{MD}(n, p) = nq^{n-1} - \frac{1}{n} (q^{n-1} - (n-1)pq^{n-2}) = q^{n-1} \left(n - \frac{1}{n} \right) + \frac{n-1}{n} pq^{n-2} > 0$$

for any fixed $n \in (2, \infty)$, (M2) holds. $\forall n \in (2, \infty) \lim_{p \rightarrow 0+} t_{MD}(n, p) = \frac{1}{n} < 1$. Hence (M4).

Finally, verification of (M3) can be done in the same as in case of procedure A2. Since exercise is quite lengthy and tedious, we omit the details as well as checking that system (3) does not admit solution lying in $(2, \infty) \times (0, \text{UCP}]$. The latter means that we have bifurcation of type (b1). Since $t_{MD}(2, \text{UCP}) = 1$, we conclude that $p_{MD,c} = \lim_{n \rightarrow 2+} p_n = \text{UCP}$. Figure 3 provides graphical illustration of the said.

3.4. Sterrett procedure S. Sterrett [11] suggested to modify the modified Dorfman's procedure as follows:

- (s1) test pool consisting of all items; in case it tests negative, finish; otherwise proceed to the step (s2).
- (s2) test items one-by-one until the first positive appears; consider the remaining untested set as an initial one and go back to the step (s1).

Sobel and Groll [10] demonstrated that

$$t_S(n, p) = 2 - q + \frac{2q - (1 - q)^{-1}(1 - q^{n+1})}{n}. \quad (9)$$

We set $c = 2$ in (M0). (M1) then readily holds. An equivalent way to verify (M2) lies in showing that $(1 - \text{UCP}, 1) \ni q \mapsto t_S(n, p)$ is decreasing. Since

$$\frac{\partial}{\partial q} t_S(n, p) = \left(\frac{2}{n} - 1 \right) - \frac{1}{n} \left(\frac{1 - q^{n+1}}{1 - q} \right)'_q,$$

one sees that the term $2/n - 1 < 0$ for any $n > 2$ and it suffices to show that the derivative

$$\left(\frac{1 - q^{n+1}}{1 - q} \right)'_q = \frac{1 - (n + 1)q^n + nq^{n+1}}{(1 - q)^2}$$

is positive. We omit the details of this exercise. (M4) follows by noting that, for any fixed $n > 2$,

$$\lim_{p \rightarrow 0+} t_S(n, p) = \lim_{q \rightarrow 1-} t_S(n, p) = 1 + \frac{2 - \lim_{q \rightarrow 1-} \frac{(1 - q^{n+1})'_q}{(1 - q)'_q}}{n} = 1 + \frac{2 - \frac{n+1}{1}}{n} = \frac{1}{n} < 1.$$

As in the previous example, we omit verification of (M3). It amounts to careful analysis of the derivative of $(2, \infty) \ni n \mapsto t_S(n, \text{UCP})$. One can also show that the system (3) does not admit solutions (n, p) lying in $(2, \infty) \times (0, \text{UCP}]$. Since UCP solves $t_S(2, p) = 1$ (w.r.t. p), we again have that $p_{S,c} = \lim_{n \rightarrow 2+} p_n = \text{UCP}$ as in the previous example. Figure 4 demonstrates that the bifurcation curve qualitatively exhibits the same behaviour too.

3.5. Examples violating our assumptions. We have already mentioned the work by Yao and Hwang [16] in Remark 2.3. Without digging into details, the pairwise testing (PT) procedure investigated there has a mean

$$M_{PT}(n, p) = n \frac{2 - q^2}{1 + q} + \frac{q^2 + q - 1}{(1 + q)^2} (1 - (-q)^n).$$

Clearly, it can not be extended to the continuously differentiable function. Hence, (M1) does not hold. (M3) does not hold as well. Indeed, UCP solves $q^2 + q - 1 = 0$; also $\frac{2 - q^2}{1 + q} \Big|_{p=\text{UCP}} = 1$ and

$$\frac{2 - q^2}{1 + q} < 1 \iff 0 < q^2 + q - 1 \iff p \in (0, \text{UCP}). \quad (10)$$

Another example of this kind is the Halving procedure (H) suitable for testing cohort having $2^n, n \in \mathbb{N}$ items. Its mean

$$M_H(n, p) = 1 + 2^{n+1} \sum_{k=1}^n \frac{1 - q^{2^k}}{2^k}$$

again violates (M1). We were unable to trace back the roots of this procedure with certainty yet Johnson [6] provides a full chapter (see Chapter 10) devoted to an in-depth analysis of its generalization allowing for an imperfect testing. It seems, however, that idea of halving was known long ago because it appears already in [10], Appendix C.

3.6. Discussion. The main limitation of our method is the assumption (M1). All other assumptions are much more likely to be true when talking about a typical GT procedure. At a first glance, it may seem that (M3) rules out procedures having $p_{X,c} = \text{UCP}$. As demonstrated by example, this is not the case. In fact, it restricts the subset of GT procedures to those having $p_{X,c}$ on the boundary of the domain of the bifurcation curve. We are inclined to think that this way we rule out optimal procedures, i.e. those which are best performing in certain classes. However, we do not treat this as a drawback since for optimal procedures one generally expects that $p_{X,c} = \text{UCP}$. Expanding in this direction, we note that relationships (10) uncover an interesting fact mentioned previously:

$$\forall n \in (2, \infty) t_{PT}(n, \text{UCP}) = 1$$

which means that PT procedure is of type (b0). Yao and Hwang [16] have proved that PT is an optimal nested testing procedure² if and only if $p \in \left[1 - \frac{\sqrt{2}}{2}, \frac{3-\sqrt{5}}{2}\right]$. This suggests that procedures of type (b0) are likely to be those which are optimal in some region.

In case (M0)–(M4) hold, our algorithm appears to be efficient. There is a word of caution: one has to choose c in (M0) carefully. The point is that the dynamical system (2), when viewed on a wider domain, may exhibit more complicated bifurcations. Figure 5 provides a convincing graphical illustration. Turning to the types of bifurcations, one sees that, in terms suggested by Strogatz [12], (b1) is usually the saddle point bifurcation, whereas for (b2) the system admits fixed points for all but boundary value of the control parameter $p \in (0, \text{UCP}]$. We are inclined to think that the dynamical system approach taken by us could be extended to GT procedures violating (M1) and successfully used to investigate other general properties of GT procedures yet one needs to work on the discrete scale.

APPENDIX A. PROOFS

Proof of Proposition 2.1. Take $p \in (0, p_{X,c})$. By definition of $p_{X,c}$, there exists $n \in (c, \infty) : t_X(n, p) < 1$. Hence, X makes sense for that p . Since $p \in (0, p_{X,c})$ was arbitrary, it holds true for all $p \in (0, p_{X,c})$.

Next, assume that $p_{X,c} < \text{UCP}$ (otherwise implication (1) is obvious) and take $p \in (p_{X,c}, \text{UCP}]$. Case " $\exists n \in (c, \infty) : t_X(n, p) < 1$ " contradicts the definition of $p_{X,c}$. Hence, $\forall n \in (c, \infty) t_X(n, p) \geq 1$. Assuming that $t_X(n, p) = 1$ for some $n \in (c, \infty)$ again leads to contradiction. Indeed, take $p' \in (p_{X,c}, p)$ and employ (M2) to deduce that

$$t_X(n, p') < t_X(n, p) = 1 \Rightarrow p_{X,c} = \sup\{p \in (0, \text{UCP}) \mid \exists n \in (c, \infty) : t_X(n, p) < 1\} \geq p' > p_{X,c}. \quad \square \quad (11)$$

Proof of Proposition 2.2. First, note that $t_X(n, p_{X,c}) \geq 1$ for all $n \in (c, \infty)$ provided $p_{X,c} < \text{UCP}$. Assuming an opposite leads to the same contradiction as in (11). With this in view, we proceed to the analysis of the distinct types of possible bifurcations.

²we do not define the class of the nested procedures here and refer an interested reader to [16]; for $p = \text{UCP}$, they treat PT procedure as an optimal one; this is a matter of opinion since $t_{PT}(n, \text{UCP}) = 1$ and the same result holds for one-by-one testing

Case (b0). Since $p_{X,c}$ solves $t_X(n, p_{X,c}) = 1$ for some $n \in (c, \infty)$, (M3) implies that $p_{X,c} < \text{UCP}$. Further, note that

$$\forall (n, p) \in (c, \infty) \times (0, p_{X,c}) \quad t_X(n, p) < 1. \quad (12)$$

must hold since existence of $(n, p) \in (c, \infty) \times (0, p_{X,c})$ s.t. $t_X(n, p) \geq 1$ contradicts the premise³ " $p_{X,c}$ is the only value of the control parameter for which dynamical system (3) admits fixed points in (c, ∞) ".

Fix arbitrary $(n_1, p_1) \in (c, \infty) \times (0, p_{X,c})$ and take $p_2 \in (p_{X,c}, \text{UCP})$. By Prop. 2.1, $t_X(n_1, p_2) > 1$. By (M2), $[p_1, p_2] \ni p \mapsto t_X(n_1, p)$ is strictly increasing. Therefore, there exists unique $p_0 \in (p_1, p_2)$ s.t. $t_X(n_1, p_0) = 1$. By the premise, $p_0 = p_{X,c}$. Since n_1 was arbitrary, it follows that

$$\forall n \in (c, \infty) \quad t_X(n, p_{X,c}) = 1 \quad (13)$$

and $p_{X,c}$ is the unique value of the control parameter $p \in (0, \text{UCP})$ obeying this property. Differentiating both sides of (13) with respect to n one finds out that (3) holds as well.

Case (b1)–(b2). Assume that there exists $(n_l, p_l) \in (c, \infty) \times (0, p_{X,c})$ s.t. $t_X(n_l, p_l) = 1$. Take arbitrary $p \in (p_l, p_{X,c})$. Since $p > p_l$, it follows that $t_X(n_l, p) > 1$ because of (M2). On the other hand, by the definition of $p_{X,c}$, there exists $n_p \in (c, \infty)$ s.t. $t_X(n_p, p) < 1$. Therefore, by the Intermediate Value Theorem and continuity of $n \mapsto t_X(n, p)$, there exists $n_1 \in (\min(n_p, n_l), \max(n_p, n_l))$ s.t. $t_X(n_1, p) = 1$. Since this holds for any $p \in [p_l, p_{X,c})$, we have (b1) provided $\forall n \in (c, \infty) \quad t_X(n, p_{X,c}) > 1$. Otherwise, we have (b2), and $p_{X,c}$ then can't be equal to UCP because of (M3).

Inversion of the bifurcation curve. When dealing with (b0), we have already shown that the bifurcation curve defines a constant map $(c, \infty) \ni n \mapsto p_n \equiv p_{X,c}$. As for (b1)–(b2), note that, for any fixed $n \in (c, \infty)$, the following applies:

- by the said in the very beginning of the proof and (M3),

$$\forall n \quad t_X(n, p_{X,c}) \geq 1; \quad (14)$$

- by (M4),

$$\forall n \exists p \in (0, p_{X,c}) : t_X(n, p) < 1; \quad (15)$$

- (14)–(15) and (M2) imply existence of a unique $p_n \in (0, p_{X,c}]$ s.t. $t_X(n, p_n) = 1$.

Therefore, we have a well defined map $(c, \infty) \ni n \mapsto p_n \in (0, p_{X,c}]$. By (M1)–(M2), $p \mapsto t_X(n, p)$ is differentiable and increasing for any $n \in (c, \infty)$. Therefore, $\forall n \in (c, \infty) \quad \frac{\partial}{\partial p} t_X(n, p) > 0$ and one can apply the Implicit Function Theorem to $\varphi(n, p) = t_X(n, p) - 1$ to deduce that $n \mapsto p_n$ is differentiable as well. Moreover, differentiating both sides of $t_X(n, p_n) = 1$ and applying the chain rule, we have that

$$\frac{\partial}{\partial n} t_X(n, p_n) + \frac{\partial}{\partial p} t_X(n, p_n) \frac{\partial}{\partial n} p_n = 0 \Rightarrow \frac{\partial}{\partial n} p_n = - \frac{\frac{\partial}{\partial n} t_X(n, p_n)}{\frac{\partial}{\partial p} t_X(n, p_n)}. \quad (16)$$

Therefore, looking for extremes of $n \mapsto p_n$ one has to solve

$$\frac{\partial}{\partial n} t_X(n, p_n) = 0 \iff \frac{\partial}{\partial n} p_n = 0$$

with respect to n . Since p_n also solves $t_X(n, p) = 1$, extremes and corresponding values can be obtained by solving (3). For bifurcations of type (b0) and (b2), maximal value $p_{X,c}$ is attained at some inner point(s) $n_c \in (c, \infty)$; for the bifurcation of type (b1), the maximal value lies on the boundary of its domain. \square

³ $t_X(n, p) = 1$ is clearly impossible; assuming strict inequality $t_X(n, p) > 1$, to reach contradiction, employ (M2) and (M4)

APPENDIX B. FIGURES

The figures were produced by making use of Desmos Graphing Calculator [3].

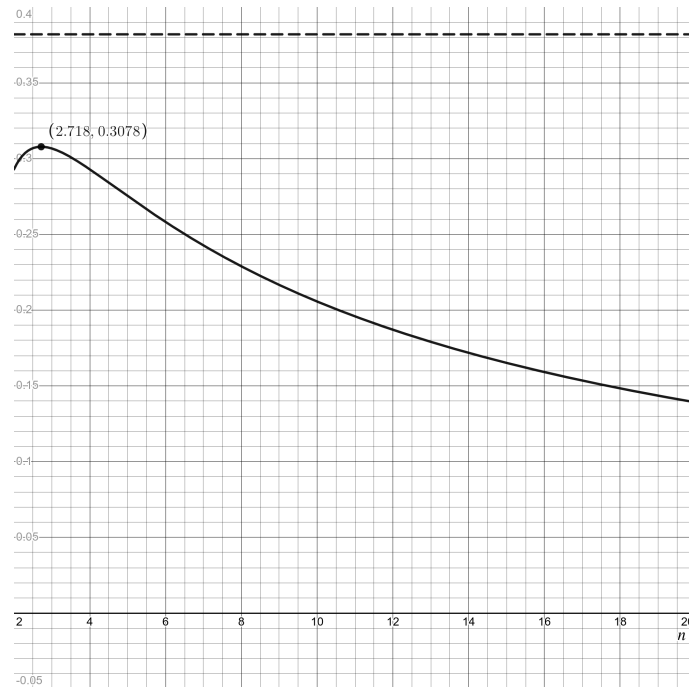


FIGURE 1. Plot of $n \mapsto p_n$ (solid line) for the procedure D; dashed line shows constant line $p = \text{UCP} = \frac{3-\sqrt{5}}{2}$; maximal value yields $\text{COCP} = 1 - e^{-e^{-1}} \approx 0.3078$; value $p_{[e]} = 1 - 3^{-3^{-1}}$ is equal to DOCP .

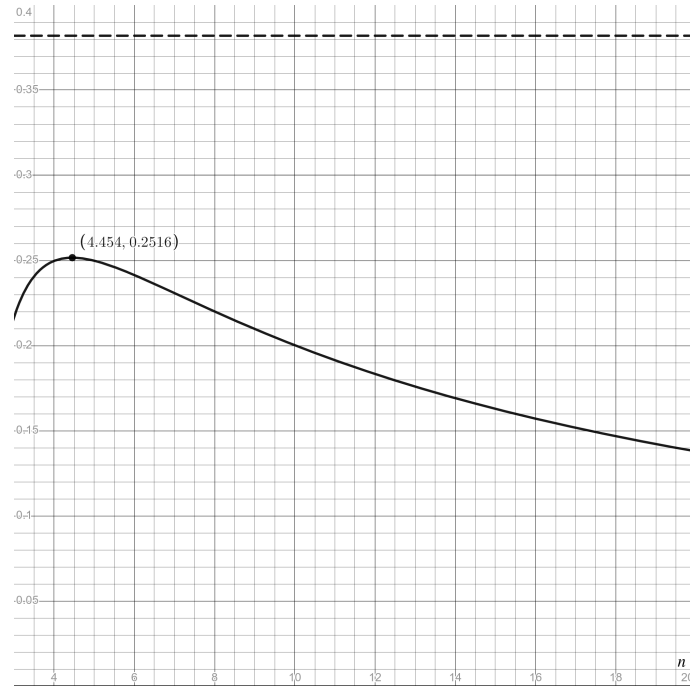


FIGURE 2. Plot of $n \mapsto p_n$ (solid line) for the procedure A2; dashed line shows constant line $p = \text{UCP} = \frac{3-\sqrt{5}}{2}$; maximal value yields COCP ≈ 0.2516 ; $p_5 \approx 0.2498$ is equal to DOCP.

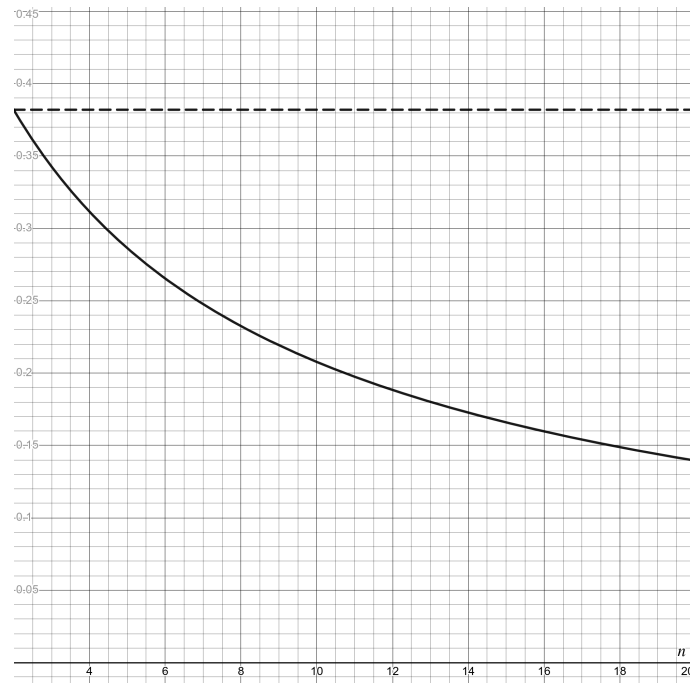


FIGURE 3. Plot of $n \mapsto p_n$ (solid line) for the procedure MD; dashed line shows constant line $p = \text{UCP} = \frac{3-\sqrt{5}}{2}$.

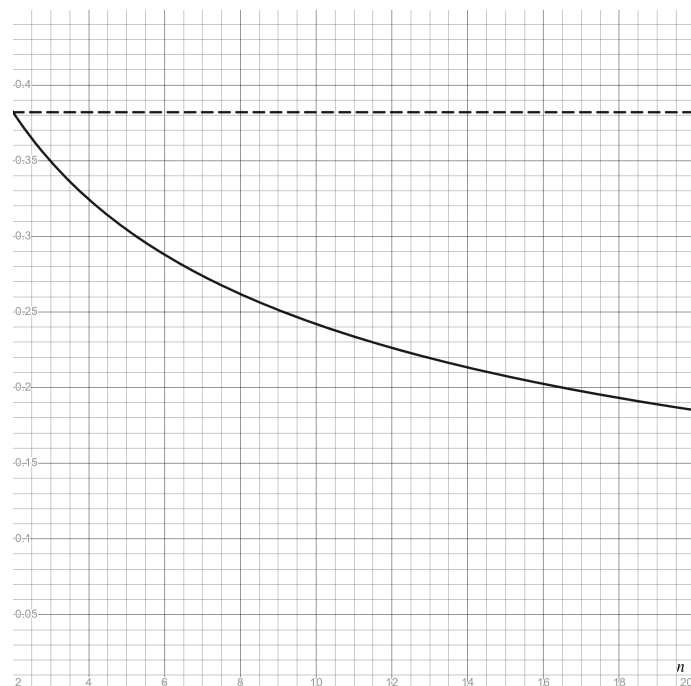


FIGURE 4. Plot of $n \mapsto p_n$ (solid line) for the procedure S; dashed line shows constant line $p = \text{UCP} = \frac{3-\sqrt{5}}{2}$.

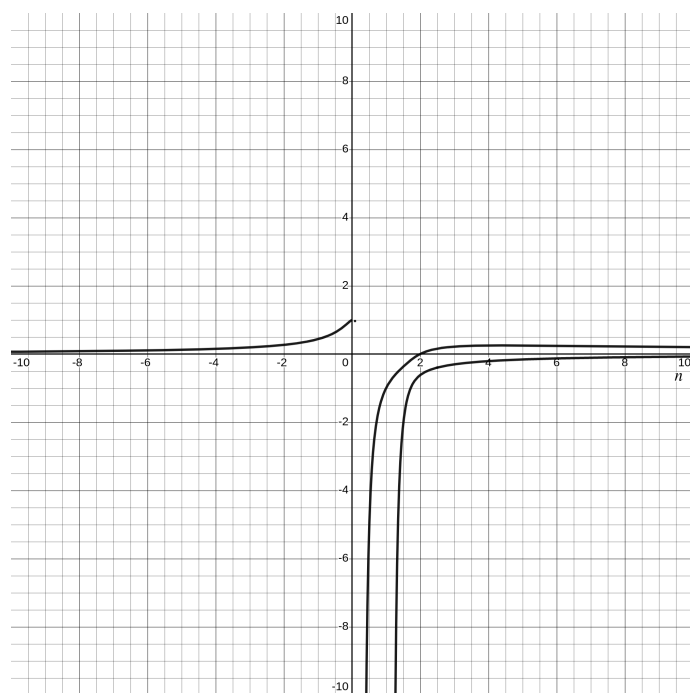


FIGURE 5. Plot of bifurcation curve for procedure A2 when the domain of the dynamical system (2) is extended to the whole real line.

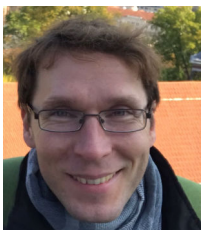
Acknowledgement. The authors would like to thank anonymous referee for pointing out some bibliographical information which was incorporated into the final version of the manuscript.

REFERENCES

- [1] Aldridge, M., Johnson, O., and Scarlett, J., (2019), Group testing: an information theory perspective, Foundations and Trends in Communications and Information Theory Series, Now Publishers.
- [2] Berger, T., Mandell, J. W., Subrahmanya, P., (2000), Maximally efficient two-stage screening, *Biometrics*, 56(3), pp. 833-840.
- [3] PBC Desmos Studio, (2023), Desmos graphing calculator, <https://www.desmos.com/calculator>, On-line, Accessed March 2023.
- [4] Dorfman, R., (1943), The detection of defective members of large populations, *The Annals of Mathematical Statistics*, 14(4), pp. 436-440.
- [5] Hudgens, M. G., Kim, H. Y., (2011), Optimal configuration of a square array group testing algorithm, *Communications in Statistics - Theory and Methods*, 40(3), pp. 436-448.
- [6] Johnson, N. L., Kotz, S., Xizhi, W., (1991), *Inspection errors for attributes in quality control*, Springer US, Boston, MA.
- [7] Malinovsky, Y., Albert, P. S., (2019), Revisiting nested group testing procedures: new results, comparisons, and robustness, *The American Statistician*, 73(2), pp. 117-125.
- [8] Phatarfod, R. M., Sudbury, A., (1994), The use of a square array scheme in blood testing, *Statistics in Medicine*, 13(22), pp. 2337-2343.
- [9] Samuels, S. M., (1978), The exact solution to the two-stage group-testing problem, *Technometrics*, 20(4), pp. 497-500.
- [10] Sobel, M., Groll, P. A., (1959) Group testing to eliminate efficiently all defectives in a binomial sample, *Bell System Technical Journal*, 38(5), pp. 1179-1252.
- [11] Sterrett, A., (1957), On the detection of defective members of large populations, *The Annals of Mathematical Statistics*, 28(4), pp. 1033-1036.
- [12] Strogatz, S. H., (2015), *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*, Second edition, Westview Press.
- [13] Ungar, P., (1960), The cutoff point for group testing, *Communications on Pure and Applied Mathematics*, 13(1), pp. 49-54.
- [14] Virtanen, P., et al., (2020), SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nature Methods*, 17(3), pp. 261-272.
- [15] Yao, Y. C., Hwang, F. K., (1988), A fundamental monotonicity in group testing, *SIAM Journal on Discrete Mathematics*, 1(2), pp. 256-259.
- [16] Yao, Y. C., Hwang, F. K., (1990), On optimal nested group testing algorithms, *Journal of Statistical Planning and Inference*, 24(2), pp. 167-175.



Ugnė Čižikovienė received bachelor's degree in Econometrics from Vilnius University (VU), Vilnius, Lithuania, in 2016. After that she continued master studies in VU and ended up with master's degree in Economics in 2018. Two years later, in 2020, she again entered VU and started doing PhD under supervision of the second author of the present paper. The foreseen graduation date is 2025. During her studies, she worked both in and outside the academy. Outside the academy, she was employed as the data analyst in several private Lithuanian companies. During her PhD studies, she served as a junior assistant in VU simultaneously doing her research.



Viktor Skorniakov received all his degrees from Vilnius University (VU), Vilnius, Lithuania: bachelor's degree (Mathematical Statistics) in 2004, master's degree (Mathematical Statistics) in 2006, and PhD (Mathematics) in 2010. After obtaining bachelor's degree he served as a statistician/programmer in "Statistics Lithuania" (for approximately two years) and later in VU Hospital "Santaros klinikos" (for approximately five years). During that period, he also did PhD related research and taught various courses in VU. After that, he completely moved to the academy and continues here up to date in a capacity of associate professor. Since 2022, he is also a head of Institute of Applied Mathematics in VU, Faculty of Mathematics and Informatics.
