

Optimal testing strategies for infectious diseases

Harris Georgiou¹

Data Science Lab, Dept. of Informatics, University of Piraeus, Greece
<https://www.datastories.org>
hgeorgiou@unipi.gr

Abstract. Screening tests for infectious diseases is a problem typically addressed in the field of Medicine and Epidemics. However, as the SARS-CoV-2 pandemic emerged, it became clear that there is no globally accepted strategy for optimizing such procedures, e.g. in international transportation and border checks, which policy makers can employ. In this study, the general problem of developing optimal testing strategies for infectious diseases is explored under the scope of Game Theory, sampling and estimation methods from classic Statistics, as well as Bayesian methods for the proper treatment of posterior updates, leading to the benefits of employing Machine Learning for data-driven structural risk minimization. Six main guidelines are established by this work, dictating estimated variance of prevalence and associated risk as the main minimization target, in terms of both a criterion for inflow quotas allocation between population groups, as well as optimal posterior updates via classic confidence intervals and Bayesian methods. As a result, it is established that minimum infection risk, not optimal resource allocation, is the real challenge and top priority in formalizing optimal screening strategies for such risk mitigation policies.

Keywords: Epidemics · SARS-CoV-2 · screening methods · testing strategies · Game Theory · Machine Learning · Bayesian methods.

1 Introduction

The SARS-CoV-2 pandemic that characterized 2020 was the most crucial factor in revisiting common practices and re-establishing risk-mitigation policies in terms of population screening for infectious diseases. As of September 2021, world-wide statistics [10] show that a third pandemic surge is at its peak and there are more than 219 million confirmed cases, with more than 4.55 million deaths directly associated with the COVID disease. This is perhaps the most severe health emergency since the Spanish flu a century ago. Although this virus seems to exhibit patterns of droplet rather than aerosol transmission, its rate is very high and it combined with an incubation period of 4-7 days [24]. Due to international flights, this ‘window of opportunity’ is extremely crucial for the virus to spread undetected in asymptomatic carriers. The recent ‘delta’ variant seems to exhibit even longer incubation period, perhaps up to 7-8 days, while

the virus itself is more contagious than the original strain, making it even more dangerous.

Testing strategies is perhaps the single most valuable tool for the containment of the virus spread, especially between countries and when international travel bans are to be lifted. However, there are theoretical and practical aspects that make perfect screening impossible, leaving the decision makers with the crucial but ill-defined challenge of how to introduce risk-mitigation plans, based on imperfect information, time delays and limited resources.

With the SARS-CoV-2 pandemic still surging almost two years after its first appearance in Wuhan (November 2019), the current focus in testing strategies world-wide is towards long-term epidemic monitoring within a country, i.e., as part of mitigation policies for restoring normal economic activities and avoiding strict lockdowns. There are numerous guidelines from the international health organizations like CDC (USA), ECDC (EU) and WHO [12,11,13,14,43], as well as from the research community [6,36,28]. However, there are only few works addressing the challenge of optimizing testing in border checks, i.e., very short-term screening human flows (travellers, within 1 hour at most) using only limited resources (number of test kits per day per entry point). These approaches include mostly adaptations w.r.t. incidence rates [2,21,33], statistical models on prevalence [9,20], bandit formulations for ‘hit’ rate optimization [5,27], etc. There are only few game-theoretic approaches for modelling the evolution of the outbreak, the effects of the individual behaviours and the mitigation policies [7,35,3]. However, none of these approaches address the challenge of combining incidence rates, inherent cost for ‘missed’ cases and the posterior (after testing) probabilities for healthy/infected individuals, specifically for very short-term screening human flows in border crossings between countries, which essentially is the driving factor for turning national-level epidemics into a global pandemic. Moreover, no such approach is available as a well-defined framework based on mathematical foundations and derived strategies.

In this work, the challenge of optimal testing strategies for infectious disease screening is treated in a unified way. Beginning from the problem under the viewpoint of Game Theory, the decision-making authority has to optimize its testing strategy for groups of individuals, partitioned on the basis of some inherent property, e.g. the country of origin, demographics, etc. Assuming a fixed capability on the base task of selecting ‘safe’ and ‘non-safe’ individuals, the game setup of player-against-Nature and the goal is to minimize the loss from improper allocation of increased and decreased rates of flow, which is typically associated to what is happening in screening gates between countries or between regions within the same country.

On top of the game-theoretic formalization of the testing strategies for the decision-making authority, their capability is further explored within the context of Sampling Theory and Estimation Theory. Since all estimations are based on sampled data and not perfect ‘Oracle’ view of the entire population, proper adjustments should be made to the properties that define the game-theoretic optimizations. Moreover, the inherent limitations and ‘static’ nature of the stan-

standard confidence interval methods are compared to the more rigorous and intuitive view of the Bayesian methods for optimal posterior updates of the corresponding probabilistic models.

Aggregating all these aspects of the screening tests during an epidemic, this work presents a proper formalization of each step in a constructive way and clearly defines their constraints. Section 2 defines the main task, which is the problem of optimizing the policies for screening tests; section 3 provides the general game-theoretic framework, solution concepts and evidence-based adaptations; section 4 extends this framework to multiple ‘experts’, also providing solution concepts and application to testing strategies; section 5 describes the related theory for sampling, estimation and evidence-based posterior updates, including point statistics, confidence intervals and Bayesian methods; finally, section 6 discusses further complications from having to cope with time-varying uncertainty in the evidence and counter-intuitive testing strategies for prompt containment of the disease, as well as some concluding remarks in section 7.

2 Problem statement

First of all, let us define the exact optimization task, which in general terms can be described with the two equivalent statements:

- *Minimize* the *risk* margin of *not identifying* infected individuals in a target group, i.e., the *cost* expectancy.
- *Maximize* the *safety* margin of *identifying* infected individuals in a target group, i.e., the *gain* expectancy.

In the first statement, ‘cost’ is referring to the probabilistic expectancy of the overall negative impact for the group from missed cases of infected individuals, which is associated to a risk margin, i.e., a ‘miss’ probability. Similarly, in the second statement, ‘gain’ is referring to the probabilistic expectancy of the overall positive impact for the group from detected cases of infected individuals, which is associated to a safety margin, i.e., a ‘hit’ probability. The two definitions can be considered equivalent in the sense that detecting and isolating infected individuals is always beneficial for the group. Hence, in the following analysis they are used interchangeably as needed, with complementary probabilities and outcomes.

3 Game-theoretic formalization

A very generic approach in formalizing the definition of this optimization setup is via Game Theory [34,31,18]. Specifically, a zero-sum game can be designed between ‘Nature’ that defines the (unknown) infected individuals and the checking ‘authority’ that tries to identify and isolate them, i.e., mitigate the negative impact of missed cases. In this setup, ‘Nature’ is typically the ‘negative’ player and the ‘authority’ is the ‘positive’ player; since ‘Nature’ is the stochastic factor out of any immediate control, ‘authority’ is of the main interest here and is

typically associated to the positive-valued outcomes of the game. Hence, the second statement in section 2 is the one that is preferred by default when defining probabilities ('hits') and outcomes ('gain').

Let C be the zero-sum 'checking' game according to the previous generic description of section 2. Let N be the total number of individuals, either 'safe' (not infected) or 'non-safe' (infected), partitioned according to $L = \langle \ell_k \rangle$, $N = |\bigcup \ell_k|$, $k \in \{1, \dots, |L|\}$, and with corresponding subset sizes n_k , $N = \sum_{k=1}^{|L|} n_k$. If p_k^s is the probability of an individual in group k being 'safe', then $p_k^{ns} = (1 - p_k^s)$ is the associated probability of an individual being 'non-safe' in that same group. With n_k being the size of group k , $c_k^s \geq 0$ is the marginal gain from each 'safe' individual and $-c_k^{ns} \leq 0$ is the marginal cost from each 'non-safe' individual, then the probabilistic expectancy (game value) from each group outcome is:

$$v_k(C) = n_k(p_k^s c_k^s - (1 - p_k^s) c_k^{ns}) \quad (1)$$

where c_k^s and c_k^{ns} can be considered as common for all subsets in L , thus can be used as c^s and c^{ns} , respectively, with k omitted.

Besides the p_k^s probability, n_k is the other crucial factor for determining the overall outcome of the game. In practice, there are two extreme options for the checking authority: (a) admitting all the n_k individuals in the group or (b) not admitting anyone of them. If subsets ℓ_k are fixed and cannot be partitioned further, then only these two extreme options are available and the task becomes discrete (combinatorial), i.e., selecting or discarding each ℓ_k from L . Option (b) is the pure strategy that always guarantees a non-negative outcome, but in practice this also comes with an associated cost of not admitting the 'safe' individuals. In terms of game C , there are two pure strategies for each player and four outcomes in total, as illustrated in Table 1. However, since the main interest here is to estimate the optimal strategy for the checking authority (rows) and the two extreme options can be merged taking $n_k \geq 0$, the corresponding target function for maximization is Eq.1 restated as:

$$v_k(C) = n_k(p_k^s c^s - (1 - p_k^s) c^{ns}) \quad (2)$$

Table 1. Zero-sum 'checking' game C in complete 2-by-2 form.

	p_k^s	$1 - p_k^s$
$n_k > 0$	$+c^s$	$-c^{ns}$
$n_k = 0$	$-c^s$	$+c^{ns}$

The optimization task defined by Eq.2 is considered against all the subsets ℓ_k in partitioning L , i.e., for some fixed total sum $N = \sum_{k=1}^{|L|} n_k$ and the combined expectancy $v(C) = \sum_{k=1}^{|L|} v_k(C)$, regarding the partitioning into subsets $\ell_k \in L$ and their corresponding sizes n_k . The exact aggregated maximization of $v(C)$ against partitioning L with subsets of sizes n_k is stated in Definition 1 for arbitrary sizes n_k and in Definition 2 for fixed sizes n_k (selection of indices k).

Definition 1 (Optimal partitioning in game C with subsets of arbitrary sizes). Let C be a zero-sum ‘checking’ game as described in Eq.2, with $0 \leq p_k^s \leq 1$, $\{c^s, c^{ns}\} \geq 0$, $L = \langle \ell_k \rangle \neq \emptyset$, $N = |\bigcup \ell_k| = \sum_{k=1}^{|L|} n_k$. Then the combined expectancy $v(C) = \sum_{k=1}^{|L|} v_k(C)$, regarding the partitioning L into subsets ℓ_k of corresponding sizes n_k , is maximized with a specific set $\langle n_k \rangle = \{n_1, \dots, n_{|L|}\}$:

$$\langle n_k \rangle : \arg \max_{n_k} \sum_{k=1}^{|L|} v_k(C) \quad (3)$$

If the partitioning L defines subsets of fixed size, then the problem becomes a discrete optimization task, with the target being the selection of a combination of ℓ_k in L that exhibit the maximum overall payoff, as Definition 2 states:

Definition 2 (Optimal partitioning in game C with subsets of fixed sizes). Let C be a zero-sum ‘checking’ game as described in Eq.2, with $0 \leq p_k^s \leq 1$, $\{c^s, c^{ns}\} \geq 0$, $L = \langle \ell_k \rangle \neq \emptyset$, $N = |\bigcup \ell_k| = \sum_{k=1}^{|L|} n_k$. Then the combined expectancy $v(C) = \sum_{k=1}^{|L|} v_k(C)$, regarding the partitioning L into subsets ℓ_k of corresponding fixed sizes $n_k = \ell_k$, is maximized with a specific combination of indices k in the set $\langle n_k \rangle \in \{n_1, \dots, n_{|L|}\}$:

$$\langle k \rangle : \arg \max_k \sum_{k=1}^{|L|} v_k(C) \quad (4)$$

Definitions 1 and 2 formalize the optimization task for the checking authority (‘max’ player) regarding the game C . In the first case the authority is free to adjust $\langle n_k \rangle$ freely with the only constraint being $n_k \geq 0$, i.e., enlarge or reduce each subset in the partitioning L . In the second case the subset sizes n_k are fixed and the authority can only adjust the selection of indices in $\langle k \rangle \in \{1, \dots, |L|\}$. In both cases, the subset sizes n_k are constrained by the partitioning L , i.e., $N = \sum_{k=1}^{|L|} n_k$.

3.1 Solution concepts

As stated previously, the obvious strategy for the checking authority in order to ensure $v(C) \not\prec 0$ is to set $n_k = 0, \forall k \in \{1, \dots, |L|\}$, if there is such a valid option available. However, the most usual case is that the total sum $N = \sum_{k=1}^{|L|} n_k > 0$ is a fixed constraint, i.e., cannot be avoided or maybe not even reduced. This essentially makes the overall task for $v(C)$ against $\langle n_k \rangle$ a convex optimization problem in a continuous (Definition 1) or a discrete (Definition 2) space.

In order to see how $v(C)$ is maximized against n_k given p_k^s taking into account all k , Eq.2 is applied to Eq.3 and Eq.4 for the arbitrary or fixed sizes n_k , respectively. Intuitively, we expect that for $p_1^s \geq p_2^s \geq \dots \geq p_{|L|}^s$, the obvious choice for $k \in \{1, \dots, |L|\}$ is such that it maximizes the corresponding subset sizes, i.e., $n_1 \geq n_2 \geq \dots \geq n_{|L|}$ while satisfying the constraint $N = \sum_{k=1}^{|L|} n_k > 0$. Lemma 1 and Theorem 1 provide the formal proofs for this optimizer.

Lemma 1. *Let C be a zero-sum ‘checking’ game as described in Definition 1, with $0 \leq p_k^s \leq 1$, $\{c^s, c^{ns}\} \geq 0$, $L = \langle \ell_k \rangle \neq \emptyset$, $N = |\bigcup \ell_k| = \sum_{k=1}^{|L|} n_k$. Let Θ be the indices defining the descending ordering of $\langle p_k^s \rangle$, that is $p_{\theta_1}^s \geq p_{\theta_2}^s \geq \dots \geq p_{\theta_{|L|}}^s$. Then Θ also defines the same descending ordering for $\langle \gamma_k \rangle$, where $\gamma_k = p_k^s c^s - (1 - p_k^s) c^{ns}$ as in Eq.2.*

Proof. See [19].

Theorem 1 (Optimal mixture of arbitrary-size subsets in game C). *Let C be a zero-sum ‘checking’ game as described in Definition 1, with $0 \leq p_k^s \leq 1$, $\{c^s, c^{ns}\} \geq 0$, $L = \langle \ell_k \rangle \neq \emptyset$, $N = |\bigcup \ell_k| = \sum_{k=1}^{|L|} n_k$. Let $q_k \geq n_k$ be the upper limit (quota) for the size of the corresponding subset ℓ_k . Let Θ be the indices defining the descending ordering of $\langle p_k^s \rangle$, that is $p_{\theta_1}^s \geq p_{\theta_2}^s \geq \dots \geq p_{\theta_{|L|}}^s$. Then the combined expectancy $v(C) = \sum_{k=1}^{|L|} v_k(C)$, regarding the partitioning into subsets $\ell_k \in L$ of corresponding sizes $n_k = |\ell_k|$, is maximized with:*

$$r : N_{\Theta} = \sum_{j=\theta_1}^{\theta_r} q_j \leq N < \sum_{j=\theta_1}^{\theta_{r+1}} q_j \quad (5)$$

$$\langle n_k \rangle = \{q_{\theta_1}, \dots, q_{\theta_r}, N - N_{\Theta}, 0, \dots, 0\} \quad (6)$$

Proof. See [19].

What Theorem 1 provides is a proof of the intuitive approach from everyday practice: when facing a set of $|L|$ choices associated with different rewards and a total sum N of selections, it is normal that the most logical thing to do is maximize the selections from the top rewards, then do the same for the second-best rewards, etc, until N is completed.

The same approach for the optimal strategy can be applied when the selections are of fixed subsets, which essentially makes the overall problem a discrete optimization task. Based on Theorem 1, Theorem 2 proves that it reduces to selecting the subset of the best-reward options.

Theorem 2 (Optimal mixture of fixed-size subsets in game C). *Let C be a zero-sum ‘checking’ game as described in Definition 2 with $0 \leq p_k^s \leq 1$, $\{c^s, c^{ns}\} \geq 0$, $L = \langle \ell_k \rangle \neq \emptyset$, $N = |\bigcup \ell_k| = \sum_{k=1}^{|L|} n_k$. Let $n_k = q_k = |\ell_k|$ be the size and its upper limit (quota) for each subset $\ell_k \in L$, i.e., fixed for each $k \in \{1, \dots, |L|\}$. Let $\gamma_k = p_k^s c^s - (1 - p_k^s) c^{ns}$ as in Lemma 1. Let Θ be the indices defining the descending ordering of $\langle n_k, \gamma_k \rangle$, that is $n_{\theta_1} \gamma_{\theta_1} \geq n_{\theta_2} \gamma_{\theta_2} \geq \dots \geq n_{\theta_{|L|}} \gamma_{\theta_{|L|}}$. Then the combined expectancy $v(C) = \sum_{k=1}^{|L|} v_k(C)$, regarding the partitioning into subsets $\ell_k \in L$ of corresponding sizes $n_k = |\ell_k|$, is maximized with:*

$$r : N_{\Theta} = \sum_{j=\theta_1}^{\theta_r} q_j \leq N < \sum_{j=\theta_1}^{\theta_{r+1}} q_j \quad (7)$$

$$\langle k \rangle \subseteq \Theta : \langle k \rangle = \{\theta_1, \dots, \theta_r, \widehat{\theta_{r+1}}, 0, \dots, 0\} \quad (8)$$

where $q_{\widehat{\theta}_{r+1}} = N - N_{\Theta} \leq q_{\theta_{r+1}}$.

Proof. See [19].

Remark 1. For the last subset in position $\theta_{r+1} \in \Theta$ the value $n_{\theta_{r+1}} = N - N_{\Theta}$ may not be a valid option if it is not compatible with the fixed-size $\ell_{\theta_{r+1}}$ or, equivalently, the total sum may not be satisfied, i.e., $\sum_{k=1}^{|L|} n_k < N$. This depends on the exact partitioning L in relation to N and the defined quotas $\langle q_k \rangle$; this does not invalidate the general solution provided by Theorem 2.

It is worth noting that the descending ordering Θ in Theorem 2 is now strictly defined for $\langle n_k \gamma_k \rangle$ and cannot be deferred to $\langle p_k^s \rangle$ according to Lemma 1. This is because $n_k = q_k = |\ell_k|$ is now fixed and, thus, cannot be arbitrarily set to zero for the lower-ranked positions in $\langle n_k \rangle$. In that sense, Eq.2 can be rewritten as:

$$v_k(C) = \delta_k q_k (p_k^s c^s - (1 - p_k^s) c^{n_s}) \quad (9)$$

where $\delta_k \in \{0, 1\} \forall k \in \{1, \dots, |L|\}$.

It should also be noted that the discrete case as stated in Definition 2 it is loosely related to the *subset sum* problem [25], which explores the different ways that a positive integer can be expressed as the sum of other positive integers. Another way to state the subset sum problem is: given a set of positive integers and a target sum N , does any subset of the numbers sum to precisely N ; or more loosely, find a subset whose sum is as close as possible to N - this is precisely what Eq.7 in Theorem 2 indicates regarding the selection of $\ell_k \in L$. However, the main difference here is that there is only one $|\ell_k|$ positive integer to use from each ‘class’ k in the sum, i.e., it is purely a matter of selection of singletons rather than combination of (possible) repetitions of numbers.

Based on Theorem 1, Algorithm 1 provides a baseline constructive procedure for calculating the optimal solution of Eq.5 and Eq.6. Input $|L|$ defines compartments ℓ_k but not their sizes and inputs c^s, c^{n_s} are not strictly necessary, they are only referenced for proper definition of the arg max equation. This algorithm also applies to the discrete case as described by Theorem 2, i.e., optimizing against $\langle k \rangle$ instead of $\langle n_k \rangle$, provided that: (a) the sorting statement for getting Θ is applied to $\langle n_k \gamma_k \rangle$; and (b) the last *if* statement also includes a validity check with regard to the value $N - N_{\Theta}$.

From Algorithm 1 it is obvious that the constructive procedure outlined is quite acceptable in terms of complexity. In fact, the most complex part is the sorting step, typically in the order of $O(n \log n)$. The iteration in the main loop is clearly linear at $O(n)$, since the two summations inside the loop are actually temporary variables stepwise-increased (no loops). This low-complexity property of the solution is particularly important for Theorem 2, showing that even though the discrete optimization task is combinatorial, the optimal solution can be constructed generally in $O(n \log n)$, or even $O(n)$ if the input vectors are already sorted.

Algorithm 1: Optimal mixture of partitioned pool

Result: Maximize the expected gain in mixing partitions of different success rates:

$$\langle n_k \rangle : \arg \max_{n_k} \sum_{k=1}^{|L|} n_k (p_k^s c^s - (1 - p_k^s) c^{ns}) \quad (10)$$

Input:

- partitioning: $L = \langle \ell_k \rangle \neq \emptyset, k = \{1, \dots, |L|\}$
- constraint: $N = |\bigcup \ell_k| = \sum_{k=1}^{|L|} |\ell_k|$
- success rates: $\langle p_k^s \rangle = \{p_1^s, \dots, p_{|L|}^s\}, 0 \leq p_k^s \leq 1$
- subset quotas: $\langle q_k \rangle = \{q_1, \dots, q_{|L|}\}, |\ell_k| \leq q_k$
- marginal gain/cost: $\{c^s, c^{ns}\} \geq 0$

Output:

- subset sizes: $\langle n_k \rangle = \{n_1, \dots, n_{|L|}\}$

```

1 sort  $\langle p_k^s \rangle \rightarrow \Theta : p_{\theta_1}^s \geq p_{\theta_2}^s \geq \dots \geq p_{\theta_{|L|}}^s$  ;
2  $\langle n_k \rangle \leftarrow \mathbf{0}$  ;
3  $r \leftarrow 0$  ;
4  $finished \leftarrow False$  ;
5 while  $\neg finished$  do
6    $r \leftarrow r + 1$  ;
7    $n_{\theta_r} \leftarrow q_{\theta_r}$  ;
8    $finished \leftarrow \left( \sum_{k=\theta_1}^{\theta_r} n_k \leq N \right) \wedge \left( \sum_{k=\theta_1}^{\theta_{r+1}} n_k > N \right)$  ;
9 end
10 if  $\sum_{k=\theta_1}^{\theta_r} n_k < N$  then
11    $n_{\theta_{r+1}} \leftarrow N - \sum_{k=\theta_1}^{\theta_r} n_k$  ;
12 end
13 return  $\langle n_k \rangle$ ;

```

3.2 Evidence-based soft partitioning

The game-theoretic formalization explored previously addresses a setup of ‘compartmentalized’ pool of N individuals according to some partitioning L , while having only one ‘expert’ with success rate p_k^s per subset $\ell_k \in L$. This essentially reduces the problem, in both continuous and discrete cases, to a convex optimization task as defined in Definitions 1 and 2, respectively. However, in the general case there may be no strict partitioning L and, hence, N may be one single pool of individuals - this is actually the generalization of having L defined by some *evidence* that updates the inclusion to some $\ell_k \in L$ and exclusion from the other $|L| - 1$ partitions of each individual, according to a posterior probability that is not strictly equal to unity. Additionally, there may be more than one ‘experts’ per such partition ℓ_k , each evaluating the individuals with a different success rate p_k^s and, hence, an aggregation scheme has to be applied for the final evaluation of ‘safe’ or ‘non-safe’ labelling. Eq.11 describes the Bayesian posterior probability of O_j of m mutually exclusive outcomes given an observed evidence E :

$$P(O_j|E) = \frac{P(E|O_j)P(O_j)}{\sum_{j=1}^m P(E|O_j)P(O_j)} = \frac{P(E|O_j)P(O_j)}{P(E)} \quad (11)$$

Exploiting evidence for posterior updates and combining aggregated experts is a highly challenging and multi-aspect research area that has been explored for many years from different disciplines, ranging from Evolutionary Biology and Sociology to Game Theory and Ensemble Learning [23,42,17]. In section 4 the problem of optimal testing strategies for infectious diseases is restated in this general context, providing a proper analytical solution and proofs for combining multiple experts and with arbitrary posterior updates.

4 Weighted Majority Games

In collective decision-making, the individual outputs in an ensemble of experts with moderate performance levels are aggregated in order to produce a collective decision in an optimal way [18]. According to the Condorcet Jury Theorem [8], if the experts’ individual decisions are independent and their corresponding estimations are more likely to be correct than incorrect ($p > 0.5$), then an increase in the collective performance of the ensemble is guaranteed when the individual estimations are combined with a plurality voting scheme. Moreover, this increase in performance continues to increase asymptotically as the size of the ensemble increases, i.e., as more independent experts are added.

Although the experts within such an ensemble can be viewed as competitive players each trying to impose its own choice upon the output of the ensemble, in reality this depends on the *coalitions* that each choice forms regarding these sub-groups. In other words, each expert is relevant to the collective output only if it participates in the sub-group that dictates this final decision, e.g., via majority voting. This cooperative type of games is well-studied within the context of Coalitional Gaming [31], a natural extension of the non-cooperative setups

in classic Game Theory. In coalitional games, the optimality of the collective decision-making depends on the exact aggregation rule and the parameters or *weights* that are assigned to each member of the ensemble, typically in association to each competency level for the specific task at hand [23,17].

The case of a having K options to choose from, using (weighted) plurality voting as aggregation rule, is defined as a Weighted Majority Game (WMG) [34]. It has been proven by Nitzan and Paroush (1982) [32] and Shapley and Grofman (1984) [38] that the optimal decision rules, in terms of collective performance, are the Weighted Majority Rules (WMR); this is in fact a different name for the well-known weighted majority voting schemes [26], which are often used in Machine Learning for combining hard-output classifiers. The same assertion has also been verified by Ben-Yashar and Nitzan [4] as the optimal aggregation rule for committees under the scope of informative voting in Decision Theory.

Within this context, Definition 3 and 4 formalize the Weighted Majority Voting (WMV) as the aggregation scheme for WMG, respectively. Furthermore, Definition 5 formalize the WMR as the optimal aggregation rule for such WMG setups.

Definition 3 (Weighted Majority Voting (WMV)). *Let G be an ensemble of K decision-making ‘experts’ $\{D_1, \dots, D_K\}$ with individual outputs $s_i \in \Omega$ against labels $\omega_j \in \Omega$ and corresponding accuracies $\{p_1, \dots, p_K\}$. Then the voting support $g_j(\mathbf{x})$ for label ω_j given the input \mathbf{x} is defined as:*

$$g_j(\mathbf{x}) = \sum_{i \in K_j} w_i, K \supseteq K_j : \{s_i = \omega_j\} \quad (12)$$

where the voting weights w_i are typically defined as $0 \leq w_i \leq 1$ and $\sum_{i=1}^K w_i = 1$.

Definition 4 (Weighted Majority Game (WMG)). *Let G be an ensemble of K decision-making ‘experts’ $\{D_1, \dots, D_K\}$ in a WMV setup according to Definition 3. Then the Weighted Majority Game (WMG) of the ensemble defines the winning coalitions $K_{win} \subseteq K$ with regard to their corresponding output label as the ones having support $g_j(\mathbf{x})$ no less than a pre-defined lower threshold or decision quota q_{win} :*

$$i \in K_{win} : g_j(\mathbf{x}) \geq q_{win}, K_{win} \subseteq K \quad (13)$$

Definition 5 (Weighted Majority Rule (WMR)). *Let G be an ensemble of K decision-making ‘experts’ $\{D_1, \dots, D_K\}$ in a WMG setup according to Definition 4, but with $q_{win} = 0$ and the largest support $g_j(\mathbf{x})$ always defining a single winning coalition. Then the Weighted Majority Rule (WMR) of the ensemble under these constraints is according to the output label with the maximum support $g_j(\mathbf{x})$:*

$$i \in K : \arg \max_j g_j(\mathbf{x}) \quad (14)$$

In the case where there are $|\Omega|$ available choices for each expert, it is sufficient to define the decision threshold as $q_{win} = 1/|\Omega|$ in order to ensure well-defined

collective decisions in the sense of both Definition 4 and 5, i.e., selecting the maximum-support option *and* satisfying the decision threshold at the same time.

4.1 Solution concepts

As mentioned previously, WMR has been proven [32,38,4] as the theoretically optimal aggregation rule for WMG. This means that in any ensemble with K voting experts a set of voting weights $\langle w_i \rangle$ can be defined so that the corresponding WMR produces the optimal aggregation producing their collective decision. This is a particularly important result, since it only depends on the aggregation itself and not the design or the internal properties of each individual expert in the ensemble. Hence, the complexity of defining the optimal aggregation completely is reduced to the convex (see Definition 3) optimization task of defining vector $\langle w_i \rangle$, $i = \{1, \dots, K\}$.

In the restricted case of having independent experts and (possibly) fractional weights $w_i \in \mathbb{R}$, the optimal design of WMR has been studied extensively and independently in a wide range of disciplines. Specifically, instead of the intuitive linear mapping of the experts' competencies $\langle p_i \rangle$ to corresponding voting weights $\langle w_i \rangle$ in WMV, the logarithm of the odds or *logodds* rule is applied. According to [38] and [22], the logodds rule has been proposed for mixtures of experts as early as 1961 in [37]. In Machine Learning, the logodds rule re-appeared in the formulation of the Adaptive Boosting or 'Adaboost' algorithm [15] for robust ensemble meta-learning, which gave its creators Yoav Freund and Robert Schapire the 2003 Gödel Prize. In the last two decades the logodds method has used repeatedly in various meta-learning approaches as an analytical solution for optimal weighting in ensembles of experts, e.g. in classifier combination [23,44], with proven performance over real-world problems and experimental datasets, very close and sometimes even higher than fully trained (non-analytical) weighting approaches [22,16,1].

Theorem 3 formally defines the logodds rule optimality for WMR weighting profiles, given conditionally independent decision-makers, and provides a short proof via Bayesian formulation. Next, Lemma 2 specializes it for the simple case of dichotomy choice situations.

Theorem 3 (Log-odds Optimality (general)). *Let G be an ensemble of K decision-making 'experts' $\{D_1, \dots, D_K\}$ in a WMR setup according to Definition 5 and being conditionally independent, i.e., $P(\mathbf{s}|\omega_j) = \prod_{i=1}^K P(s_i|\omega_j)$, $\omega_j \in \Omega$. Then the accuracy of the ensemble (P_{maj}^w) is maximized by assigning weights:*

$$w_i \propto \log \frac{p_i}{1 - p_i}$$

Proof. See [19].

Lemma 2 (Log-odds Optimality (dichotomous choice)). *Let G be an ensemble of K decision-making 'experts' $\{D_1, \dots, D_K\}$ in a WMR setup according to Definition 5 and being conditionally independent, i.e., $P(\mathbf{s}|\omega_j) =$*

$\prod_{i=1}^K P(s_i|\omega_j)$, with $s_i, \omega_j \in \Omega = \{-1, +1\}$. Then the accuracy of the ensemble (P_{maj}^w) is maximized by assigning weights:

$$w_i = \log \frac{p_i}{1 - p_i}$$

and the optimal binary discriminator of the ensemble is:

$$g_{\pm}(\mathbf{x}) = (-\log P(\omega_-) + \log P(\omega_+)) + \sum_{i=1}^K s_i w_i \quad (15)$$

Proof. See [19].

Taking into account the definition of $\langle w_i \rangle$ from Lemma 2 for the binary choice ‘safe’ or ‘non-safe’, target range $w'_i \in [0, 1]$ via Eq.?? and the convexity constraint $\sum_{i=1}^K \widehat{w}_i = 1$, Eq.16 presents the final definition for the weighting profile $\langle \widehat{w} \rangle$ in WMR:

$$\langle \widehat{w} \rangle : \widehat{w}_i = w'_i / \sum_{i=1}^K w'_i = \frac{w_i - w_{min}}{w_{max} - w_{min}} / \sum_{i=1}^K \frac{w_i - w_{min}}{w_{max} - w_{min}} \quad (16)$$

What the WMG approach provides is a generalized approach to formulate the combination of K decision-makers, perhaps pooled into soft partitions by a Bayesian posterior update based on some membership evidence according to Eq.11. The WMR is the theoretically optimal way to do this and the optimal combination weights for the decision-makers can be analytically defined based on their individual competencies according to Lemma 2 and Eq.16. Given the fact that in the WMG approach the partitions are soft and not distinct as in section 3, the game value is now defined across the entire set of N individuals and for all the decision-makers. Again, for the dichotomous choice of ‘safe’ or ‘non-safe’ individuals, this can be defined as:

$$v(G) = N \sum_{i=1}^K (p_i^s c^s - (1 - p_i^s) c^{n_s}) \widehat{w}_i \quad (17)$$

where p_i^s is the competency of expert i on choice ‘safe’ and may be an updated Bayesian posterior according to Eq.11.

4.2 Application to ‘checking’ games

Following the problem definition of section 3, the optimization task here is how the entire set of N may be partitioned into $|L|$ subsets, where each expert may exhibit significantly increased competency and, hence, get a larger WMR weight than the others in the ensemble. In other words, instead of assigning each partition to a single expert as in section 3, enable a combined decision according to WMG, but with Bayesian posterior updates that effectively introduce soft partitioning via the corresponding competency updates and, hence, the WMR weighting profile *langlew*).

For any such partition number $k \in \{1, \dots, |L|\}$, Eq.17 can be redefined as:

$$v_k(G) = n_k \sum_{i=1}^K (p_{i,k}^s c^s - (1 - p_{i,k}^s) c^{ns}) \widehat{w}_{i,k} \quad (18)$$

where $p_{i,k}^s$ is the competency of expert i in partition number k on choice ‘safe’, an updated Bayesian posterior according to Eq.11 that effectively defines the soft partitioning into (non-distinct) subsets of sizes n_k with $\sum_{k=1}^K n_k = N$.

The difference of Eq.18 with Eq.2 is that the game value is estimated for each k of the partitions, employing $K > 1$ instead of only a single expert as in section 3. Thus, for each partition number k , if only expert o (‘oracle’) is assigned with $\widehat{w}_o = 1$ and zero weight for every other expert in the ensemble, Eq.18 reduces to Eq.2.

Eq.18 can be explored in terms of optimality conditions against both competencies $\langle p_{i,k}^s \rangle$ and weights $\langle \widehat{w}_{i,k} \rangle$. The following Lemma 3 proves that the ordering of the competencies of the decision-makers also define the ordering of the expected payoffs for any partition.

Lemma 3. *Let G be a WMG ‘checking’ game of K decision-makers as described in Definition 4, with $0 \leq p_{i,k}^s \leq 1$, $i \in \{1, \dots, K\}$, $\{c^s, c^{ns}\} \geq 0$, $L = \langle \ell_k \rangle \neq \emptyset$, $N = |\bigcup \ell_k| = \sum_{k=1}^{|L|} n_k$. Let Θ_k be the indices defining the descending ordering of $\langle p_{i,k}^s \rangle$ for each k , that is $p_{\theta_{1,k}}^s \geq p_{\theta_{2,k}}^s \geq \dots \geq p_{\theta_{K,k}}^s$. Then, if $p_{i,k}^s \geq \max\{\frac{c^{ns}}{c^s + c^{ns}}, 1/2\}$ and $\widehat{w}_{i,k}$ as defined in Eq.16, Θ_k also defines the same descending ordering for $\langle \zeta_{i,k} \rangle$, where $\zeta_{i,k} = (p_{i,k}^s c^s - (1 - p_{i,k}^s) c^{ns}) \widehat{w}_{i,k}$.*

Proof. See [19].

Lemma 3 is a generalization of Lemma 1 with the inclusion of WMR weighting. Based on this, Theorem 1 can now be reformulated accordingly for K decision-makers in an ensemble:

Theorem 4 (Optimal mixture of arbitrary-size subsets via WMG G). *Let G be a WMG ‘checking’ game of K decision-makers as described in Definition 4, with $0 \leq p_{i,k}^s \leq 1$, $i \in \{1, \dots, K\}$, $\{c^s, c^{ns}\} \geq 0$, $L = \langle \ell_k \rangle \neq \emptyset$, $N = |\bigcup \ell_k| = \sum_{k=1}^{|L|} n_k$. Let $q_k \geq n_k$ be the upper limit (quota) for the size of the corresponding subset ℓ_k . Let Θ be the indices defining the descending ordering of $\langle \max_i p_{i,k}^s \rangle$ against k , that is $p_{\theta_1}^s \geq p_{\theta_2}^s \geq \dots \geq p_{\theta_{|L|}}^s$, with $p_{i,k}^s \geq \max\{\frac{c^{ns}}{c^s + c^{ns}}, 1/2\}$. Then the combined expectancy $v(G) = \sum_{k=1}^{|L|} v_k(G)$, regarding the partitioning into subsets $\ell_k \in L$ of corresponding sizes $n_k = |\ell_k|$, is maximized with:*

$$r : N_\Theta = \sum_{j=\theta_1}^{\theta_r} q_j \leq N < \sum_{j=\theta_1}^{\theta_{r+1}} q_j \quad (19)$$

$$\langle n_k \rangle = \{q_{\theta_1}, \dots, q_{\theta_r}, N - N_\Theta, 0, \dots, 0\} \quad (20)$$

Proof. See [19].

What Theorem 4 states is that the logic of Theorem 1 is still valid for maximizing the overall payoff against a partitioned pool of N individuals to be tested. That is, even when an ensemble of K decision-makers is available for each partition, the subset sizes can still be maximized towards their corresponding quotas taking into account the maximum-position in the descending ordering of logodds-weighted marginal payoffs from the ensemble, instead of the single-expert assignment per partition. Of course, this approach is valid only when n_k is to be maximized against a single decision-maker that is to be selected as ‘active’ from the logodds-weighted ensemble. Instead, if all K decision-makers are considered ‘active’ in parallel and for every partition, then the generalized WMR in Definition 5 and the optimality proof from Theorem 3 are applied. In practice, this means that in Theorem 4 the descending ordering must be taken against the (generic) $\langle \zeta_{i,k}^s \rangle$ instead of replacing it with $\langle p_{i,k}^s \rangle$, i.e., optimizing each n_k in Eq.18 for all $k \in \{1, \dots, |L|\}$.

5 Sampling, estimations and posterior updates

In the previous sections, the problem of optimizing the allocation of N individuals to be tested to K decision-makers was investigated under the assumption of distinct, soft or no partitioning of the pool, namely in section 4 for the second (more generic) case and in section 3 for the others. Based on the analysis above, it was proven that the optimal allocation for maximum payoff, i.e., maximum expectancy of selecting ‘safe’ individuals, depends on the ranking of the competency (or the logodds-weighted transformation of it) of the decision makers regarding this task. In other words, the performance of the members in the ensemble is the criterion upon which this optimal allocation is defined. Furthermore, this optimal allocation can be constructed analytically by employing Theorems 1, 2 or 4, according to the specific setup of the task with regard to the partitioning and the decision-makers.

Although the aforementioned approach is solid and constructive, it has a severe limitation in terms of actual real-world application. Namely, it assumes perfect knowledge of the decision makers’ competencies, i.e., the corresponding vectors $\langle p_{i,k}^s \rangle$. This is rarely the case, since almost always the competencies are simply the best estimations we can get for the corresponding empirical success rates based on finite sample sets. In other words, each of the $p_{i,k}^s$ elements is an estimate, with a specific confidence interval and statistical significance value. Moreover, new sampling data should be incorporated in these estimations to provide a better result, i.e., a narrower and/or shifted confidence interval, in the sense of iteratively updating the corresponding posterior probabilities. Finally, the measurements upon the sample set may not be perfect, hence the estimation should also take into account this uncertainty. All these issues introduce factors of progressive complexity to the estimation of the competencies, not always easy to implement or even formulate as models.

5.1 Point statistics and confidence intervals

The easiest option is to estimate the competencies $\langle p_{i,k}^s \rangle$ via standard sampling theory. From the problem definition and Table 1 in section 2, it is established that the task at hand can be modelled as subsequent independent checks in a pool of ‘safe’ and ‘non-safe’ individuals, i.e, a series of Bernoulli trials. Hence, the proper probabilistic formulation of the corresponding random variable X of selecting ‘safe’ individuals (‘hits’) is via a Binomial distribution [40]:

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{1-x} = \frac{n!}{x!(n-x)!} p^x q^{1-x} \quad (21)$$

where n is the number of trials, x is the number of ‘hits’, p is the competency for ‘hits’ and $q = 1 - p$ the complementary probability for the Bernoulli trials. According to these definitions, the mean value and standard deviation can also be defined as $\mu = np$ and $\sigma = \sqrt{npq}$, respectively.

The sampled values of mean μ and variance σ^2 are known to be *unbiased* estimators, i.e., they both converge to the actual values as the sample size increases. For the first this is true, for the second only approximately for large n :

$$E[\bar{X}] = \mu_{\bar{X}} = \mu \quad (22)$$

$$E[(\bar{X} - \mu)^2] = \sigma_{\bar{X}}^2 = \left(\frac{N}{N-1}\right) \left(\frac{n-1}{n}\right) \frac{\sigma^2}{n} \approx \frac{\sigma^2}{n} \quad (23)$$

where the approximation in Eq.23 is valid even when sampling $n \leq N$ without replacement from a population of size $N \rightarrow \infty$. In practice, the unbiased estimator of σ is usually defined according to:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \Rightarrow E[S^2] = \mu_{S^2} = \frac{n-1}{n} \sigma^2 \quad (24)$$

$$\hat{S}^2 = \frac{n-1}{n} S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \Rightarrow E[\hat{S}^2] = \mu_{\hat{S}^2} = \sigma^2 \quad (25)$$

Hence, \hat{S} is typically used instead of S as an unbiased estimator for small-sized samples. Based on these fundamentals from estimation theory, the confidence intervals of the (sample) mean value for any given confidence level $z_c > 0$ is given by:

$$\bar{X} - z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{X} + z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad (26)$$

for known σ and N, n as described above, i.e., the rightmost fractions are omitted for sampling with replacement or as $N \rightarrow \infty$. Similarly, if σ is unknown it is replaced by \hat{S} or S and the same definition is given by:

$$\bar{X} - t_c \frac{\hat{S}}{\sqrt{n}} \leq \mu \leq \bar{X} + t_c \frac{\hat{S}}{\sqrt{n}} \quad (27)$$

The statistic z_c is defined according to the *Standard Normal* distribution given a specific significance level (two-tailed cumulative distribution function), e.g., for $\alpha = 0.05 \Rightarrow z_c = 1.960$. Similarly, t_c is defined according to the *Student's t* distribution given a specific significance level (two-tailed cumulative distribution function) and *degrees of freedom* (sample size), e.g., for $n = 20, \alpha = 0.05 \Rightarrow t_c = 2.093$. The t_c statistic provides a somewhat wider confidence interval due to the uncertainty of having a small sample size and/or unknown σ . Evidently, for $n > 30$ the confidence intervals provided by the two distributions are practically equal.

Using the previous formulation, the confidence intervals defined by Eq.26 and Eq.27 can be specifically rewritten for the Bernoulli probabilities, i.e., the normalized estimations or *proportions* \hat{p} and $\hat{q} = 1 - \hat{p}$ in Eq.21. Specifically, setting $\hat{p} \propto \bar{X} = k/n$ with k 'safe' individuals detected in n tests, i.e., \bar{X} is the relative frequency in the current sample, and since for Binomial distribution $\mu = np$ and $\sigma = \sqrt{npq} \approx \sqrt{\hat{S}^2} \rightarrow \sigma$, then:

$$\bar{X} - z_c \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}} \leq \hat{p} \leq \bar{X} + z_c \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}} \quad (28)$$

for known σ and N, n as described above, i.e., with the rightmost fractions omitted for sampling with replacement or as $N \rightarrow \infty$. Here, p is used for replacing $\sigma = \sqrt{p(1-p)}$, but in practice σ is also considered unknown (related to \hat{p}) and, thus, estimated via $\sigma \approx \hat{S}$ according to Eq.25, while additionally using t_c instead of z_c statistic if the size of the sample is small ($n < 30$). A more accurate definition of Eq.28 is [40]:

$$\hat{p} = \frac{\bar{X} + \frac{z_c^2}{2n} \pm z_c \sqrt{\frac{\bar{X}(1-\bar{X})}{n} + \frac{z_c^2}{4n^2}}}{1 + \frac{z_c^2}{n}} \quad (29)$$

where for large samples ($n > 30$) both Eq.28 and Eq.29 are reduced to:

$$\hat{p} = \bar{X} \pm z_c \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \quad (30)$$

using the z_c statistic with the Standard Normal distribution or, more properly for unknown (estimated) $\sigma \approx \hat{S}$:

$$\hat{p} = \bar{X} \pm t_c \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \quad (31)$$

using the t_c statistic with the Student's t distribution.

Similar approaches can be employed for hypotheses testing regarding the prevalence level, i.e., accepting or rejecting a specific estimation $\hat{p} \approx \bar{X}$ based on z_c or t_c at a specific significance level α [39]. Specifically, the z_c statistic can be reformulated in a way that enables the significance testing of hypothesis H_0

that the sample-estimated \bar{X} is ‘close enough’ to the assumed mean value μ (H_0 not rejected) at a significance level α , or reject it otherwise:

$$H_0 : -z_c \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_c \quad (32)$$

where z_c is defined for a specific significance level, e.g., for $\alpha = 0.05 \Rightarrow z_c = 1.96$. Taking into account the properties of the Binomial distribution in Eq.21 and, again, substituting for proportions $\bar{X} = \hat{p}$, $\mu = p$, $\sigma = \sqrt{p(1-p)}$, Eq.32 can be rewritten as:

$$H_0 : -z_c \leq \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \leq z_c \quad (33)$$

which has the meaning of testing whether an estimated \hat{p} is ‘close enough’ to the assumed prevalence p for the global population or if the specific sample (testing group) is α -significantly different. As previously described for the confidence intervals, proper adjustments can also be employed here for small-sized samples, using \hat{S} instead of σ and t_c instead of z_c statistics.

5.2 Evidence-based posterior updates

In Eq.11 the simplest and most common form of the Bayes rule [40] is defined for m mutually exclusive outcomes O_j , using some evidence E to update their corresponding *posterior* probabilities. This is the most fundamental and generic way to express the fact that some *prior* probability is updated when new evidence becomes available, e.g., a supplementary or more recent testing sample, data related to another property of the associated subject, etc. Bayes approaches have been applied for many years as the basis for statistical modelling and empirical estimation of incidence rates [9,20].

A more generic definition of Eq.11 is when the underlying probability distribution is continuous, as it is in most cases. Let $\mathbf{x} = \{x_1, \dots, x_n\}$ be a sample of some random variable X , $f(\mathbf{x})$ the corresponding probability density function that depends on some unknown parameter θ and Θ the random variable associated with that parameter θ having a probability distribution function $\pi(\theta)$. Then $f(x)$ may be described as the conditional density function of variable X given some $\Theta = \theta$, i.e., denoting it as $f(x|\theta)$ or ‘ x given θ ’. Similarly, the joint probability of Θ given some $X = x$ is defined as $\pi(\theta|x)$. Then the joint probability of X and Θ can be defined as $f(x;\theta) = f(x|\theta)\pi(\theta)$. Moreover, given a sample \mathbf{x} of X , the joint density function or *likelihood* can be written as $f(\mathbf{x}|\theta) = f(x_1|\theta) \cdot \dots \cdot f(x_n|\theta)$ and the density function of θ given \mathbf{x} as $\pi(\theta|\mathbf{x})$. With these definitions at hand, Eq.11 can be rewritten in its more generic form as:

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x};\theta)}{f(\mathbf{x})} = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{\Theta} f(\mathbf{x}|\theta)\pi(\theta)d\theta} \quad (34)$$

where the integral is over the range of values for θ . In practice, calculating the integral over the entire range of θ is not necessary, since the denominator is

independent of θ and the individual (exclusive) outcomes can be treated comparatively. This translates to redefining Eq.34 in a simpler way as:

$$\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta) \Leftrightarrow \pi(\theta|\mathbf{x}) = C \cdot f(\mathbf{x}|\theta)\pi(\theta) \quad (35)$$

where C is a proportionality constant independent of θ .

Based on this generic formulation of the Bayes rule, likelihood $f(\mathbf{x}|\theta)$ can be interpreted as the updated or posterior probability of observing samples \mathbf{x} from random variable X given its conditioning parametrization by θ . Similarly, $\pi(\theta|\mathbf{x})$ can be interpreted as the probability density of parameter θ for X after observing samples \mathbf{x} . This later case is very interesting when θ is to be ‘discovered’ optimally from sample data in the sense of *maximum likelihood*. Although this task is similar to the approach presented earlier with the confidence intervals, Bayesian approaches are entirely different, since they always treat the corresponding target parameters as continuous probability distributions rather than spot statistics within some α -significance range of values [40,41].

The difference between these two approaches, i.e., the classic confidence intervals versus the Bayesian methods, can be described more clearly for the case of estimating the proportion parameter p in the Binomial distribution of Eq.21. In the more generic formulation, n is the number of trials and θ is the unknown parameter of the Binomial distribution of random variable X . Then θ has a Beta-related probability density function [40]:

$$\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < \theta < 1, \quad \{\alpha, \beta\} > 0 \quad (36)$$

where $B(\alpha, \beta)$ is the Beta function:

$$B(\alpha, \beta) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du, \quad \{\alpha, \beta\} > 0 \quad (37)$$

If $\alpha = \beta = 1$ then $\pi(\theta)$ becomes the uniform density function on $[0, 1]$, meaning that no assumption can be made for θ 's distribution. Using Eq.36 and the simplification of Eq.35, the posterior density $\pi(\theta|x)$ given any observation x becomes:

$$\pi(\theta|x) \propto \frac{\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}}{B(x+\alpha, n-x+\beta)}, \quad 0 < \theta < 1, \quad \{\alpha, \beta\} > 0 \quad (38)$$

which is actually similar to Eq.36 but with parameters $x + \alpha$ and $n - x + \beta$. In words, if a random variable X is Binomial with parameters n and θ , with the prior density of θ being beta with parameters α and β , then the *posterior* density of θ after observing some $X = x$ is also beta with (updated) parameters α and β [40].

If more strict assumptions can be made for the prior density function of θ , two other common options are *Haldane's prior* with $\alpha = x$, $\beta = n - x$ in Eq.37:

$$\pi(\theta) = \frac{1}{\theta(1-\theta)} \Rightarrow \pi(\theta|x) \propto \frac{\theta^{x-1}(1-\theta)^{n-x-1}}{B(x, n-x)}, \quad 0 < \theta < 1 \quad (39)$$

and Jeffrey's prior with $\alpha = x + 1/2$, $\beta = n - x + 1/2$ in Eq.37:

$$\pi(\theta) = \frac{1}{\sqrt{\theta(1-\theta)}} \Rightarrow \pi(\theta|x) \propto \frac{\theta^{x-1/2}(1-\theta)^{n-x-1/2}}{B(x+1/2, n-x+1/2)}, \quad 0 < \theta < 1 \quad (40)$$

The definition in Eq.38 is particularly useful in the context of the 'checking' game described in the previous sections. Specifically, it describes how the assumed prevalence p of the underlying Binomial distribution, associated to the probability p_k^s of selecting 'safe' individuals, is to be updated as new testing samples x become available. With confidence intervals this translated to increasing the sample size n and, thus, narrowing the limits of the α -significant range. Here, the Bayesian approach begins with little or no assumption ($\alpha = \beta = 1$) regarding the prior distribution of parameter $\theta = \hat{p}$ and after observing x the density function gets updated, with the posterior distribution being 'reshaped' more narrowly around the maximum likelihood 'best guess' of θ , while still remaining a continuous density function.

In order to see how the Bayesian approach exploits the evidence in \mathbf{x} and improves the estimation of θ , the Binomial distribution of can be treated via Normal distribution approximation as described earlier, in order to make the analysis of X posteriors more tractable. In particular, for a random sample of size n for X drawn from a Normal distribution with unknown mean value θ (\bar{X} in the sample) and known variance σ^2 , as well as a prior distribution of θ that is Normal with mean value μ and variance v^2 , it can be proven [40] that the posterior distribution for θ is also Normal with mean value μ_{post} and variance v_{post}^2 given by:

$$E[\theta]_{post} \approx \mu_{post} = \frac{\sigma^2\mu + nv^2\bar{X}}{\sigma^2 + nv^2} \quad (41)$$

$$Var[\theta]_{post} \approx v_{post}^2 = \frac{\sigma^2v^2}{\sigma^2 + nv^2} \quad (42)$$

In words, Eq.41 defines how μ prior for the mean value of θ is updated after observing \bar{X} in the current sample of size n , provided that the corresponding variances σ^2 and v^2 are both known. Similarly, Eq.42 defines how v^2 prior for the variance of θ is updated based on that same sample. For better understanding, the comparison of prior versus posterior for the variance can be made in terms of the reciprocal of it, thus defining the *precision*:

$$\xi_{prior} = \frac{1}{v^2} \Rightarrow \xi_{post} = \frac{1}{v_{post}^2} = \frac{1}{v^2} + \frac{n}{\sigma^2} = \xi_{prior} + \xi_{data} \quad (43)$$

where the second term in ξ_{post} can be considered as the precision of the observed data, denoted by ξ_{data} . It is clear that according to Eq.43, the smaller the variance of a distribution, the larger is its precision value. Then Eq.41 and Eq.42 can be rewritten as:

$$E[\theta]_{post} \approx \mu_{post} = \frac{\xi_{prior}\mu + \xi_{data}\bar{X}}{\xi_{prior} + \xi_{data}} \quad (44)$$

$$\text{Var}[\theta]_{post} \approx \frac{1}{\xi_{post}} = \frac{1}{\xi_{prior} + \xi_{data}} \quad (45)$$

What Eq.44 and Eq.44 describe under the Bayesian scope is fundamental: As the sample size increases, so does the precision of the (posterior) variance, while the estimated mean value gets weighted more and more towards the sample (data) mean instead of the its prior. In words, *the larger the sample size, the better estimates the posteriors provide for the mean and variance over their corresponding priors*. Not surprisingly, this is similar to what happens with the range of their confidence intervals for spot value estimations, as noted earlier. However, the Bayesian context provides a more intuitive way of viewing this effect on parameter θ as transitioning from a state of little information (wide prior distribution) to more specific information (narrower posterior distribution).

Besides confidence interval estimation, the Bayesian framework also enables the analytical calculation of conditional distributions for future observations based on a currently available observed sample. Although the approach is similar, the difference is that instead of the overall posterior density function, the probability of a specific outcome is now estimated, hence the name *Bayesian predictive distributions* for this family of methods. As an example, based on Eq.38 for a beta distribution for a random variable X , the joint probability of observing $Y = y$ after obtaining a posterior for θ can be defined as [40]:

$$f(y, \theta|x) = f(y|\theta)\pi(\theta|x) = \binom{m}{y} \frac{\theta^{x+y+\alpha-1}(1-\theta)^{m+n-x-y+\beta-1}}{B(x+\alpha, n-x+\beta)} \quad (46)$$

where $0 < \theta < 1$, $\{\alpha, \beta\} > 0$, $y = \{0, \dots, m\}$ and $\{n, m\}$ the sizes of the first (observed) and the second (future) sample sizes. Then, the predictive probability function for Y is the marginal density obtained by integrating Eq.46 over θ :

$$f^*(y) = \int_0^1 \binom{m}{y} \frac{\theta^{x+y+\alpha-1}(1-\theta)^{m+n-x-y+\beta-1}}{B(x+\alpha, n-x+\beta)} d\theta \quad (47)$$

$$= \binom{m}{y} \frac{B(x+y+\alpha, m+n-x-y+\beta)}{B(x+\alpha, n-x+\beta)} \quad (48)$$

What Eq.48 provides is the point probability of observing y in a future sample of size m , based on the posterior update as previously defined for θ . In a way, it is much like hypothesis testing of whether the future sample's distribution parameter θ is consistent with its estimation on the currently available sample, but with the Bayesian approach it is, again, a (predictive) distribution function rather than an accept/reject outcome.

6 Further complications in real-world testing

The description thus far in the previous sections was focused on three main aspects:

- Game-theoretic optimal strategies for ensuring risk mitigation.

- Optimal estimation of the critical probabilistic parameters.
- Optimal posterior updates for safety margin assurances.

For the game-theoretic strategies, only the general setup of the testing process is necessary to define the optimal way to plan the risk-mitigation testing for infectious diseases, given that the goal is to maximize the pool of ‘safe’ (or, equivalently, minimize the pool of ‘non-safe’) individuals and provided that the capability of the (one or more) decision-makers in selecting those from the general pool can be accurately estimated. For this estimation, samples can be used for confidence intervals or Bayesian methods, while subsequent observations can also be exploited via posterior updates.

The issue that is prevalent in real-world testing and not covered by the aforementioned framework is related to the evidence used, i.e., the observed samples. Normally in statistics the observations are considered with absolute certainty, counting different outcomes or properties in a pool of samples without any possibility of measurement error or lack of information. However, in reality these measurements are also subject of probabilistic functions that dictate their reliability and the certainty of the outcome. If this certainty is adequately close to 100% it is usually ignored as a factor, but the typical estimation models can not address situations where this is not a valid assumption. Instead, the evidence itself must be estimated via confidence intervals or Bayesian methods, before it can be used as observation to subsequent estimation procedures.

In testing for infectious diseases this situation is a very common case-specific factor that needs to be taken into account. Neither of the two main categories of tests, molecular or antigen ‘rapid’, exhibit perfect sensitivity or specificity and, hence, their diagnostic accuracy is close but certainly lower than 100%. There are several ways to assess the accuracy of these medical testing procedures and, hence, the certainty of the evidence which they provide in statistical terms. More commonly, *sensitivity* and *specificity* represent the two major factors in any empirical (data-driven) assessment related to the confidence on the evidence regarding positive and negative hypotheses, respectively. In addition, medical testing can also be characterized by the positive and negative *predictive value*, which are related to the confidence on the evidence regarding positive and negative test outcomes, respectively.

Regarding SARS-CoV-2, Table 2 presents the corresponding numbers for five commercially available antigen ‘rapid’ tests (2020), for different levels of sampled-estimated prevalence [29]. It is clear that testing outcomes can not be assumed to exhibit perfect certainty, hence the statistical evidence on ‘safe’ or ‘non-safe’ individuals is strong but not absolute. This means that every priors estimation and posterior updates should take this into account, leading to much more complex probabilistic treatment than what was presented in the previous sections.

What Machine Learning provides is a data-driven view of these estimation tasks and an abstraction level ‘above’ the underlying statistical complications that are unknown or too complex to express analytically [42,23]. At the same time, the theoretical foundations from Artificial Intelligence and, more precisely,

the structural risk minimization of models that are trained with empirical evidence (datasets) ensure that the final estimations are optimal w.r.t. some core criterion, which is normally the maximum likelihood estimation for the ‘true’ parameters of these ‘unknown’ probabilistic models [41]. Two such typical examples are Support Vector Machines (SVM) [42], which can be designed specifically to maximize the discrimination margin between predicted classes or, equivalently, to minimize the structural risk of the empirical error of the trained model, i.e., generalized from a limited training dataset to the global problem; and Bayesian Networks [41], which naturally incorporate the notion of imperfect information (uncertainty) and cascaded propagation of evidence-based probabilistic estimation of outcomes from node to node as a continuous flow.

One specific factor that caught the attention of researchers during after the initial surge of the SARS-CoV-2 pandemic and the emergence of readily available ‘rapid’ test kits was the option of having lower accuracy but massive tests in the general population [30]. Low accuracy in testing translates to low sensitivity and/or specificity, which in turn produces low positive and/or negative predictive value. After observing the evolution of the national epidemic in several countries, especially after lifting the international travel bans during the summer of 2020, researchers have argued that in practice these policies may work much better than expected. Although this seems counter-intuitive in statistical terms, ‘rough’ but massive screening in the population can be a decisive pre-emptive contingency measure against the spreading of the virus. Having many false positives means putting into quarantine more individuals than necessary, but this can be considered as acceptable cost during such a global emergency. Having many false negatives means missing some individuals, but the massive scale of tests overcomes this drawback by detecting much more ‘non-safe’ individuals in total. In words, both cases of low-performance testing may lead to better overall mitigation strategies and contingency policies against a pandemic such as SARS-CoV-2.

7 Conclusions

In this study, the general problem of developing optimal testing strategies for infectious diseases like SARS-CoV-2 was explored under the scope of Game Theory, sampling and estimation methods from classic Statistics, as well as Bayesian methods for the proper treatment of posterior updates. Overall, it is a very challenging research topic that requires deep understanding and somewhat new point of view, combining multiple aspects of crowd dynamics, risk management and Machine Learning, which are usually employed individually by researchers depending on their main field of expertise.

Six main guidelines have been established by this work:

1. The core task of any such screening process via testing in transit hubs and gateways is minimizing the risk of introducing infectious individuals in the general population; there is no point in maximizing the ‘hit rate’ of the tests, especially when the testing pool is very limited.

Table 2. Indicative performance of antigen ‘rapid’ SARS-CoV-2 tests from five different companies (2020), for different levels of sampled-estimated prevalence \bar{X} , based on $n = 1000$ sample patients. [29].

$\bar{X} = 0.02$	<i>STY</i>	<i>SPY</i>	TP	FP	FN	TN	PPV	NPV
comp.1	100.00%	94.29%	20	56	0	924	26.32%	100.00%
comp.2	95.00%	98.47%	19	15	1	965	55.88%	99.90%
comp.3	100.00%	93.06%	20	68	0	912	22.73%	100.00%
comp.4	90.00%	97.24%	18	27	2	953	40.00%	99.79%
comp.5	85.00%	97.24%	17	27	3	953	38.64%	99.69%

$\bar{X} = 0.30$	<i>STY</i>	<i>SPY</i>	TP	FP	FN	TN	PPV	NPV
comp.1	100.00%	94.29%	300	40	0	660	88.24%	100.00%
comp.2	96.67%	98.57%	290	10	10	690	96.67%	98.57%
comp.3	98.67%	93.14%	296	48	4	652	86.05%	99.39%
comp.4	89.67%	93.14%	269	48	31	652	84.86%	95.46%
comp.5	85.33%	97.14%	256	20	44	680	92.75%	93.92%

2. Risk minimization is inherently associated with (estimated) variance minimization of incidence rates for each sampling group; hence, allocating testing resources according to this principle must be the core screening policy.
3. When planning an inflow-allocation strategy for a specific capacity from multiple groups, risk minimization dictates that maximizing quotas towards lower-ranked incidence groups is the optimal policy.
4. Multiple decision-making independent ‘experts’ (models) for detecting ‘safe’ versus ‘non-safe’ individuals can be combined optimally via analytical solutions from Coalitional Games, specifically Weighted Majority Voting.
5. Estimated confidence intervals must be used instead of point means for proper control of risk margins; even more, Bayesian methods provide a more intuitive way for continuous posterior updates exploiting testing results.
6. In order to cope with the increased complexity of uncertainty in evidence (test reliability), Machine Learning methods are appropriate for data-driven maximum likelihood estimation of parameters and structural risk minimization.

It is imperative to stress out the differences between common cost/benefit optimization methods like bandit algorithms from Operational Research and the risk-minimization target of the problem treated here: *maximizing the detection value of any single test does not minimize the posterior incidence risk for the population*. If tests themselves are put in the center of the optimization goal, there is no guarantee whatsoever that the risk of infection propagation is minimized. In other words, *minimum infection risk, not optimal resource allocation, is the real challenge and top priority*.

Under the scope of these core principles described above, any screening policy designed by the decision-making authorities is guaranteed to minimize the risk of introducing infectious individuals to the general population.

References

1. M Ahsen, R Vogel, and G Stolovitzky. Unsupervised evaluation and weighted aggregation of ranked classification predictions. *Journal of Machine Learning Research*, 20(166):1–40, November 2019.
2. N Augenblick, J Kolstad, Z Obermeyer, and A Wang. Group testing in a pandemic: The role of frequent testing, correlated risk, and machine learning. *National Bureau of Economic Research*, 2020.
3. A Bairagi, M Masud, D Kim, and et.al. A noncooperative game analysis for controlling covid-19 outbreak. *medRxiv*, pages 1739–1776, May 2020.
4. R. Ben-Yashar and S. Nitzan. The optimal decision rule for fixed-size committees in dichotomous choice situations: The general result. *Int. Econ. Rev.*, 38:175–186, 1997.
5. O Besbes, Y Gur, and A Zeevi. Stochastic multi-armed-bandit problem with non-stationary rewards. *Advances in neural information processing systems*, pages 199–207, 2014.
6. Boeing. Covid-19 screening strategy comparisons, May 2021.
7. A Caparros and M Finus. The corona-pandemic: A game-theoretic perspective on regional and global governance. *Environmental and Resource Economics*, pages 1–15, August 2020.
8. M. de (1785) Condorcet. *An essay on the application of probability theory to plurality decision making: An election between three candidates*. Univ. of Oxford, UK, 1989.
9. O Devine, T Louis, and E Halloran. Empirical bayes methods for stabilizing incidence rates before mapping. *Epidemiology*, pages 622–630, 1994.
10. E Dong, H Du, and L Gardner. An interactive web-based dashboard to track covid-19 in real time. *The Lancet Infectious Diseases*, Feb 2020.
11. Centers for Disease Control and Prevention (CDC). Overview of testing for sars-cov-2 (covid-19), August 2021.
12. Centers for Disease Control and Prevention (CDC). Testing strategies for sars-cov-2, August 2021.
13. European Centre for Disease Prevention and Control (ECDC). Covid-19 testing strategies and objectives, September 2020.
14. European Centre for Disease Prevention and Control (ECDC). Testing strategies for sars-cov-2, May 2021.
15. Y Freund and R Schapire. A short introduction to boosting (in japanese). *Journal of Japanese Society for Artificial Intelligence*, 14(5):771–780, September 1999.
16. H. Georgiou, , M. Mavroforakis, and S. Theodoridis. A game-theoretic approach to weighted majority voting for combining svm classifiers. In S. Kollias, A. Stafylopatis, W. Duch, and E. Oja, editors, *Artificial Neural Networks – ICANN 2006*, pages 284–292, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
17. H Georgiou. Collective decision efficiency and optimal voting mechanisms: A comprehensive overview for multi-classifier model. *arXiv:1502.02191*, pages 1–21, February 2015.
18. H Georgiou. *Games people play - Conflicts, mechanisms and collective decision-making in expert committees*. (e-book) viXra:1506.0168, 1st edition, 2015.
19. H. Georgiou. A game-theoretic and statistical approach to designing optimal testing strategies for infectious diseases (to appear), 2021.
20. S Greenland and J Robins. Empirical-bayes adjustments for multiple comparisons are sometimes useful. *Epidemiology*, pages 244–251, 1991.

21. M Kasy and A Teytelboym. Adaptive targeted infectious disease testing. *Oxford Review of Economic Policy*, 36:s77–s93, 2020.
22. L. Kuncheva. Fuzzy versus nonfuzzy in combining classifiers designed by boosting. *IEEE Trans Suzzy Systems*, 11(6):729–741, December 2003.
23. L. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, USA, 2004.
24. S Lauer, K Grantz, Q Bi, F Jones, and et.al. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, March 2020.
25. S. Lipschutz and M. Lipson. *Discrete Mathematics (3rd/Ed.)*. McGraw-Hill, 2009.
26. N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212–261, 1994.
27. H Luo, C-Y Wei, A Agrwal, and J Langford. Efficient contextual bandits in non-stationary worlds. *Conf. Learning Theory*, pages 1739–1776, 2018.
28. G Lyng, N Sheils, C Kennedy, and et.al. Identifying optimal covid-19 testing strategies for schools and businesses: Balancing testing frequency, individual test technology, and cost. *PLoS ONE*, 16(3):e0248783, March 2021.
29. C McWilliams, K Kamran, and G Procop. Testing 1-2-3: Testing strategies for employers in the era of covid-19, May 2020.
30. M Mina, R Parker, and D Larremore. Rethinking covid-19 test sensitivity - a strategy for containment. *The New England Journal of Medicine*, 383(22):e120(1–3), Nov 2020.
31. C Montet and D Serra. *Game Theory and Economics*. Palgrave Macmillan, 2003.
32. S. Nitzan and J. Paroush. Optimal decision rules in uncertain dichotomous choice situations. *Int. Econ. Rev.*, 23:289–297, 1982.
33. M Norizuki, M Hachiya, A Motohashi, and et.al. Effective screening strategies for detection of asymptomatic covid-19 travelers at airport quarantine stations: Exploratory findings in japan. *Global Health and Medicine*, 3(2):107–111, 2021.
34. G. Owen. *Game Theory (3rd Ed.)*. Academic Press, USA, 1995.
35. M Ozkaya and B Izgi. Effects of the quarantine on the individuals’ risk of covid-19 infection: Game theoretical approach. *Alexandria Engineering Journal*, 60(4):4157–4165, August 2021.
36. D Paltiel, A Zheng, and R Walensky. Covid-19 screening strategies that permit the safe re-opening of college campuses. *medRxiv*, pages 1–36, July 2020.
37. W. Pierce. *Improving reliability of digital systems by redundancy and adaptation*. PhD dissertation, Stanford Univ., USA, 1961.
38. L. Shapley and B. Grofman. Optimizing group judgemental accuracy in the presence of independence. *Public Choice*, 43:329–343, 1984.
39. M.R. Spiegel, J. Liu, and S. Lipschutz. *Mathematical Handbook of Formulas and Tables (4th/Ed.)*. McGraw-Hill, 2012.
40. M.R. Spiegel, J. Schiller, and R.A. Srinivasan. *Probability and Statistics (3rd/Ed.)*. McGraw-Hill, 2009.
41. S. Theodoridis. *Machine Learning: A Bayesian and Optimization Perspective*. Academic Press, 2nd edition, July 2020.
42. S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 4th edition, November 2008.
43. World Health Organization (WHO). Recommendations for national sars-cov-2 testing strategies and diagnostic capacities - interim guidance, June 2021.
44. S Yuksel, J Wilson, and P Gader. Twenty years of mixture of experts. *IEEE Trans. on Neural Networks*, 23(8):1177–1192, August 2012.