

# Effectively managing diagnostic tests to monitor the COVID-19 outbreak in Italy

Lorenzo Lampariello<sup>1</sup> and Simone Sagratella<sup>2</sup>

<sup>1</sup>Department of Business Studies, Roma Tre University, Rome, Italy,  
lorenzo.lampariello@uniroma3.it

<sup>2</sup>Department of Computer, Control and Management Engineering Antonio  
Ruberti, Sapienza University of Rome, Rome, Italy,  
sagratella@diag.uniroma1.it

March 13, 2020

## Abstract

Urged by the outbreak of the COVID-19 in Italy, this study aims at helping to tackle the spread of the disease by resorting to operations research techniques. In particular, we propose a mathematical program to model the problem of establishing how many diagnostic tests the Italian regions must perform in order to maximize the overall disease detection capability. An important feature of our approach is its simplicity: data we resort to are easy to obtain and one can employ standard optimization tools to address the problem. The results we obtain when applying our method to the Italian case seem promising.

## 1 Aims and scope

We focus on strategies to effectively managing diagnostic tests to monitor the COVID-19 outbreak in Italy. In fact, there is a widespread belief that diagnostic testing is crucial to detect the largest possible number of people infected by the disease. At the beginning of the outbreak, finding even just an additional positive case is important to adopt suitable early containment policies in order to slow the outbreak. Moreover, having as much as possible a precise picture of the total number of people actually infected is key to develop mathematical models (see [4, 5] for some early results) to understand the disease evolution and when the peak of infection will happen. For technical details on diagnostic test procedures, see [3].

In Italy, tests are performed independently by laboratories managed by regional health care systems. However, results of the diagnostic tests need to be confirmed by the central

national authority, namely Istituto Superiore di Sanità. From this respect, we remark that the Italian health care system is federal-like, even if the national authority still plays an important control role. We focus on the decision problem of the central authority that, aiming at effectively monitoring the evolution of the disease outbreak, wishes the regions to perform a suitable number of diagnostic tests. Clearly, diagnostic tests are a limited resource because of their cost in terms of money and time, but also given the finite capacity of the central national authority to check the tests results. Hence, the need for optimization techniques to address this critical decision problem emerges.

We propose a mathematical program to model the problem of deciding how many COVID-19 diagnostic tests the regions must perform in order to maximize the overall disease detection capability, hence, roughly speaking, minimizing the number of negative tests. The main aim of our approach is to improve on the results obtained by the strategy actually adopted by the Italian health care system. Moreover, our method is designed to be as simple as possible so that it can be understood and implemented even by non-experts in operations research. In fact, we rely only on public data that are easy to obtain. Furthermore, our method needs only simple and readily available software tools to be called for. In Section 2, we describe our mathematical model, while in Section 3, we present some results for the Italian COVID-19 outbreak case.

## 2 A mathematical model

In a federal Health Care System (HCS), assume each regional unit  $i \in \{1, \dots, n\}$  to be given (by a central decision unit) a number  $x_i \in \mathbb{R}_+$  of diagnostic tests to perform during a disease outbreak. For the sake of simplicity, we assume the outcome of a diagnostic test to be positive if the patient is infected, while negative otherwise. In this paper we consider the case of a central unit aiming at assigning to the regional units an optimal number of diagnostic tests such that the number of infected people that are detected (i.e. the positivity rate) is maximized. We assume that the following data are publicly available for each region  $i$  at some time during the outbreak:

$d_i \in \mathbb{R}_+$  is the total number of deaths from the disease (cumulated from the beginning of the outbreak to the time considered);

$c_i \in \mathbb{R}_+$  is the total number of cases, that is the total number of people diagnosed with the disease (cumulated from the beginning of the outbreak to the time considered);

$p_i \in \mathbb{R}_+$  is the size of the population;

$\sigma_i \in [0, 1]$  is the positivity rate for the diagnostic test (observed at the time considered):  
if, for example, in a day, the region  $i$  performs  $t_i$  tests and the observed number of the diagnosed cases is  $r_i$ , then  $\sigma_i = r_i/t_i$ .

We also assume an estimate  $\varphi \in [0, 1]$  is available for the Case to Fatality Rate (CFR), that is the ratio of deaths from the disease to the total number of people diagnosed with the disease during its outbreak.

For each region  $i$ , indicating with  $\text{undetected}_i$  an estimate of the number of not diagnosed people that are affected by the disease, we define the following utility function (see also Figure 1):

$$\mu_i(x_i) \triangleq \frac{\text{undetected}_i}{p_i} x_i - \frac{\sigma_i}{2p_i} x_i^2.$$

The function  $\mu_i : \mathbb{R} \rightarrow \mathbb{R}$  provides a measure of the positive outcome for the diagnostic tests  $x_i$  assigned to region  $i$ . In fact, the linear term coefficient  $\text{undetected}_i/p_i$ , that is the undetected people ratio with respect to the size of the population in region  $i$ , gives the marginal utility of performing the first test: more specifically, it can be viewed as an empirical estimate of the positive outcome for the first diagnostic test performed randomly on the population of region  $i$ . The (negative) correction quadratic term is introduced based on the idea that, if a number of diagnostic tests  $x_i = \text{undetected}_i/\sigma_i$  is performed, then ideally all the undetected people have been checked, and the marginal utility of an additional test should be zero (see Figure 1). As a result, we get a concave quadratic utility function; this is clearly advantageous from a numerical point of view, see [1].

The quantity  $\text{undetected}_i$ , which is crucial in defining the utility  $\mu_i$ , is difficult to estimate. Here, we propose a possible simple way to compute it, based on the publicly available data introduced above:

$$\text{undetected}_i = \max \left\{ 0, \frac{d_i}{\varphi} - c_i \right\}.$$

In fact, the ratio  $d_i/\varphi$  gives an estimate on the total number of infected people during the disease outbreak, based on the actual number of deaths from the disease. Subtracting from it the number of actually diagnosed cases provides an estimate on the number of undetected cases.

Clearly, the choices of the central unit are subject to some constraints. In particular, it is natural to assume the total number of diagnostic tests to be upper-bounded by some available budget  $b \in \mathbb{R}_+$ , and the the number of diagnostic tests assigned to each region  $i$  to be lower and upper-bounded by some quantities  $l_i, u_i \in \mathbb{R}_+$ , respectively. Overall, the optimization problem that the central unit faces reads as follows:

$$\begin{aligned} & \underset{x}{\text{maximize}} && \sum_{i=1}^n \mu_i(x_i) \\ & \text{s.t.} && \sum_{i=1}^n x_i \leq b \\ & && l_i \leq x_i \leq u_i, \quad i = 1, \dots, n. \end{aligned} \tag{1}$$

We remark that problem (1) is a convex quadratic program that can be solved by means of many efficient optimization methods.

### 3 The COVID-19 outbreak: the Italian case

On a day-by-day basis, we analyze the beginning of the COVID-19 outbreak in Italy: more specifically, we consider the week from March 4th to March 10th 2020. We remark that

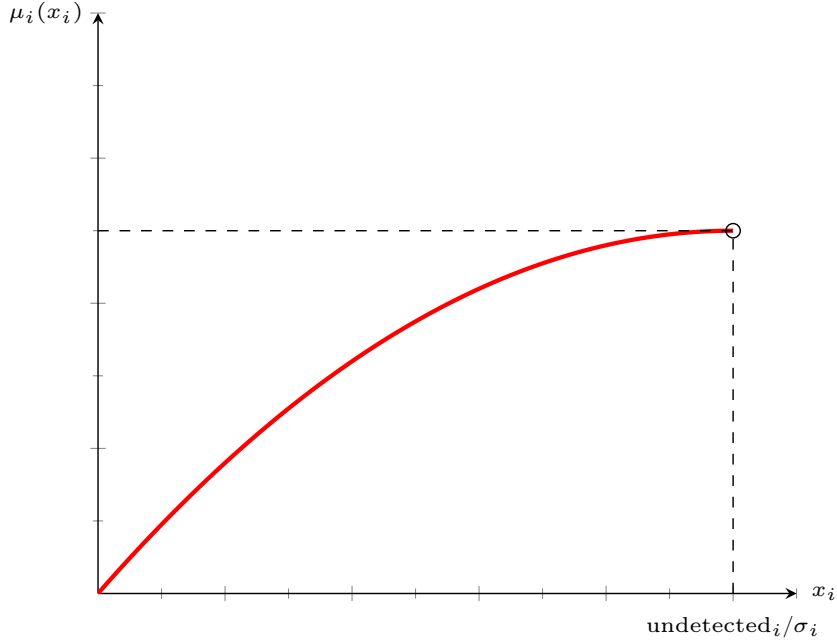


Figure 1: behavior of the utility function: at  $x_i = \text{undetected}_i/\sigma_i$  the curve is flat.

the region-oriented Italian HCS fits with the federal framework we assume in Section 2. In particular, we focus on the five regions that, during the week we focus on, are the most affected by the COVID-19 outbreak: Lombardia, Emilia Romagna, Veneto, Marche and Piemonte. All data can be found at the Italian Ministry of Health webpage <http://www.salute.gov.it/nuovocoronavirus>. For every  $i$ , one can download from the latter webpage parameters  $d_i$  and  $c_i$ ; the daily value for the positivity rate  $\sigma_i$  can be obtained by dividing the daily number of people diagnosed with COVID-19 by the total number of diagnostic tests that are performed during the day in region  $i$  (which are available on the same webpage). For our numerical experience, according to [2], we take  $\varphi = 0.034$ .

Each day of the time frame we deal with, we solve the decision problem (1) to establish the optimal number of diagnostic tests to be performed the day after. For example, consider the case in which we are solving the decision problem (1) on March 7th in order to compute the number of tests to be performed on March 8th. In Table I, we report the value for the parameters of March 7th that appear in the problem to be solved. On the other hand, as for the total budget, we set  $b$  equal to the total number of diagnostic tests that are made in the five regions on March 8th, i.e. 5784 (see the last row in Table III). Moreover, in accordance with the observed data, we set the upper-bound  $u_i = 3000$  for every  $i$ . Finally, regarding the the lower-bounds  $l_i$ , we preliminarily introduce the parameter  $\alpha \in [0, 1]$ . We choose the lower-bounds to be proportional to the total number of cases  $c_i$ , and so that  $\sum_{i=1}^n l_i = \alpha b$  and  $l_i \in [0, u_i]$ , according to the following scheme.

---

**Algorithm 1:** a procedure to compute lower bounds  $l_i$

---

$I \leftarrow$  vector of the indices of  $\left(\frac{c_1}{u_1} \dots \frac{c_n}{u_n}\right)$  sorted in descending order;  
 $\Delta \leftarrow \alpha b$ ;  
**for**  $i = 1, \dots, n$  **do**  
     $j \leftarrow I_i$ ;  
     $l_j \leftarrow \min \left\{ u_j, \frac{c_j}{\sum_{k=1}^n c_k} \Delta \right\}$ ;  
     $\Delta \leftarrow \Delta - l_j$ ;  
**end**

---

Based on the procedure described in Algorithm 1, the parameter  $\alpha$  determines the amount of the budget  $b$  to be assigned to the regions in a proportional way with respect to the total number of diagnosed cases  $c_i$ ; the remaining budget  $(1 - \alpha)b$  is allocated by solving the optimization problem (1). By varying the value of  $\alpha$  in  $[0, 1]$ , one passes from the case ( $\alpha = 0$  and, thus,  $l_i = 0$  for every  $i$ ) in which the whole budget is assigned by solving the optimization problem (1), to the scenario ( $\alpha = 1$ ) in which the entire budget is simply allocated in a proportional way. Finally, once the optimization problem (1) is solved, if some budget is still left over, then it is attributed again proportionally.

We recall again that the optimization problem (1), being a convex quadratic program (see [1]), can be easily solved by almost any available mathematical solver: e.g., Microsoft Excel, AMPL, Matlab, etc. In particular, we have simply resorted to the Matlab routine `quadprog`.

Concerning again the case of March 8th 2020, once we solve the optimization problem (1) using data from March 7th and setting  $\alpha = 0.5$ , we obtain the outcomes reported in Figure 2 and Table II. In Figure 2 we show the percentage of the 5784 diagnostic tests that have been actually performed by the HCS of the five regions versus the optimal assignment that is computed by our method. For each region, in Table II we report the positivity rate observed March 8th, as well as the number of daily cases actually diagnosed by the HCS of the regions compared to the estimated detections obtained by our approach. Note that the latter estimate is simply computed by multiplying the positivity rate  $\sigma_i$  observed March 8th times the number of the daily diagnostic tests to be performed according to our method. We remark that, by adopting our strategy, one could pass from a total number of actual daily detections equal to 1284 to 1421.

	$d_i$	$c_i$	$p_i$	$\sigma_i$	undetected $_i$
Lombardia	154	3420	10061k	0.364	1109
Emilia Romagna	48	1010	4459k	0.299	402
Veneto	13	543	4906k	0.039	0
Marche	6	207	1525k	0.208	0
Piemonte	5	207	4356k	0.253	0

Table I: March 7th 2020 data.

Finally, in Table III, we report the daily total number of people diagnosed with the

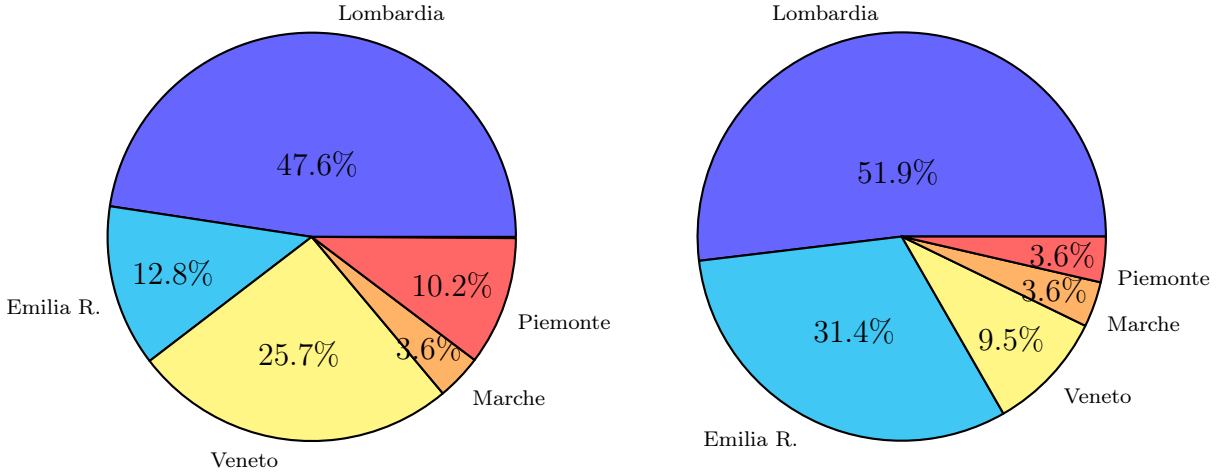


Figure 2: percentages of diagnostic tests that have been actually performed in the five regions on March 8th (on the left) vs. those obtained by our method for the same day (on the right).

	$\sigma_i$	HCS daily detections	our model daily estimated detections ( $\alpha = 0.5$ )
Lombardia	0.279	769	837
Emilia Romagna	0.230	170	418
Veneto	0.085	127	47
Marche	0.311	65	65
Piemonte	0.259	153	54

Table II: March 8th 2020 data and results: observed positivity rate and daily number of people diagnosed with the disease by means of the actual HCS method and our model, respectively.

disease in the five regions by means of the HCS actual method with respect to the estimated detections obtained by our model for different values of the parameter  $\alpha$ . Each day, for every target level of the parameter  $\alpha$ , our estimated number of detections is greater than the actual detections achieved by the regions HCS. Based on the data reported in the table, the overall performance (measured as the number of people diagnosed with the disease divided by the total number of diagnostic tests performed) of the method implemented by the regions HCS is 28.2%, while for our approach we obtain 40.1% ( $\alpha = 0$ ), 39.1% ( $\alpha = 0.5$ ) and 37.7% ( $\alpha = 1$ ).

## 4 Conclusions

We propose a mathematical method, that leveraging operations research techniques, aims at establishing how many COVID-19 diagnostic tests the Italian regions must perform in order to maximize the overall disease detection capability. An important feature of our approach is its simplicity: data we resort to are easy to obtain and the model enjoys nice mathematical properties so that one can rely on simple standard optimization softwares

	5th	6th	7th	8th	9th	10th	total
HCS daily detections	698	684	1115	1284	1601	755	6137
our model daily estimated detections ( $\alpha = 0.0$ )	1893	1132	1451	1436	1976	843	8732
( $\alpha = 0.5$ )	1790	1065	1423	1421	1984	832	8514
( $\alpha = 1.0$ )	1592	1006	1392	1398	1992	826	8206
number of daily diagnostic tests performed ( $= b$ )	2489	2950	4580	5784	2471	3499	21773

Table III: March 2020 results: total daily number of people diagnosed with the disease in the five regions by means of the HCS actual method and our model, respectively.

to treat numerically the problem. The results we obtain when applying our method to the Italian case seem promising. Also, we believe that our model is sufficiently flexible to be applied to other federal-like health care systems such as e.g. USA or Germany. We remark that in this work we provide a very preliminar analysis. Of course, the model we present here can be further generalized and enhanced by including additional features, such as, e.g., some degree of correlation among the regions' utility functions. We leave this investigation to future research. In fact, with this paper we mainly intend to provide a timely contribution to tackle the disease.

## References

- [1] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1992.
- [2] World Health Organization et al. Coronavirus disease 2019 situation report–46, 6 march 2020. [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf?sfvrsn=96b04adf\\_2,](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf?sfvrsn=96b04adf_2, 2020) 2020.
- [3] World Health Organization et al. Laboratory testing for coronavirus disease 2019 (COVID-19) in suspected human cases: interim guidance, 2 march 2020. Technical report, World Health Organization, 2020.
- [4] S.W. Park, B.M. Bolker, D. Champredon, D.J.D. Earn, M. Li, J.S. Weitz, B.T. Grenfell, and J. Dushoff. Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: framework and applications to the novel coronavirus (sars-cov-2) outbreak. *medRxiv*, 2020.
- [5] J. Zu, M. Li, Z. Li, M. Shen, Y. Xiao, and F. Ji. Epidemic trend and transmission risk of sars-cov-2 after government intervention in the mainland of china: A mathematical model study. *Available at SSRN 3539669*, 2020.