Statistical Considerations Over 70 Years Elections Data

Matteo Cirillo
Dipartimento di Fisica and MINAS Lab
Università di Roma “Tor Vergata”
00133 Roma, Italy

Abstract

The time evolution of the response of electors to political life is analyzed for the Italian Republic over a 70 years span in which 18 political elections have taken place. The basis for the performed analyses are the official data available from the Italian Ministry of Interiors exposing the results of the political elections from 1948 until 2018. The attention is concentrated on parameters providing information on the responsivity of the electors to country’s political life. These parameters, expressed in adequate percentages, are the effective number of voters and the percentage of these expressing blank or spoilt ballots. The time dependence of these parameters, over the analyzed period, shows regularities, correlations and interesting peculiarities. The analysis concerns the results for both Chamber of Deputies and Chamber of Senators available, for each election, all along the 70 years time span that are those relative to people voting on the national territory.
Analysis of votes expressed in national elections is an issue involving statistics, sociology, economics, and politics in a way that it is often not straightforward to achieve convergence and objectivity for the opinions. Beside the specific analyses an overall concern of all the political parties is the percentage of citizens expressing their vote. It has been often conjectured that the participation of citizens to the electoral process could decide the result of a specific election. The purpose of the present paper is to analyze the participation of voters to Italian elections and to analyze the percentage of those that expressed blank or spoilt ballots. We intend to show that few parameters related to these specific phenomena, when available for a substantial time span, can provide interesting information over the political sensitivity of a country ruled by democracy and free elections. The information that we analyze have been extracted from the public web site of the Italian Ministry of Interiors [1]. The data are organized for a technical analysis and we conclude commenting their impact and a possible sociological interpretation.

As anticipated in the abstract we perform our analysis for the two Italian Parliament chambers, namely Camera dei Deputati and Senato: from now on we will refer to these two chambers as Camera and Senato. From the available data the following parameters are extracted: the percentage of citizens effectively voting with respect the total number of electors having right to vote and this parameter is evaluated for both Camera and Senato and will be indicated respectively as \(RV(\text{Camera})\) and \(RV(\text{Senato})\). Then we introduce the parameters \(R_B\) and \(R_S\) which are respectively the ratio between the number of Blank and Spoilt ballots and the total number of effective voters: a parenthesis, as before, will indicate whether these parameters are relative to Camera or Senato. Finally we introduce the parameter \(R_T = R_B + R_S\) as the sum of the percentages of Blank and Spoilt ballots.

In Fig. 1a we show the time dependence, from 1948 until 2018, of the of the parameter \(RV\) for the Senato, which we there indicate as \(RV(\text{Senato})\). We see in the figure that the dependence of \(RV\) on time has two slopes. The first part of the dependence, which covers the time span up to 1979, is characterized essentially by a constant value with an intercept around 93% indicating a relevant, and
basically constant, participation/response of the voters and a noticeable will to express their opinion through the vote. After 1979, however, we see that the parameter has a clear tendency to decrease: a linear fit of the data now reveals that a linear dependence (the linear correlation coefficient of the fitting is equal to 0.964) can reasonably be envisaged through the data. The slope of the line indicates that the number of citizens expressing their political will went down, after 1979 of a rough 2% every 5 years (the constitutional duration of an elected parliament in Italy). Beside the adequacy of the linear fitting, the overall tendency of $R_V$ (Senato) to decrease is evident. The reasons for this phenomenon has been investigated in other publications [2], but the interesting thing that we wish to note in Fig. 1a is that the time response is essentially linear with just one change in derivative. We will go back to this aspect later after commenting more data.

A plot analogous to the one shown in Fig. 1a can be obtained for the parameter $R_V$ (Camera), namely the percentage of people voting with respect to total number having right for the Camera dei Deputati. Indeed results similar (dated up to 2008) to those shown in Fig. 1a, were reported for the Camera in ref. 2 (see Fig. 1 of that publication). Plotting the results for $R_V$(Senato) and $R_V$(Camera) on two Cartesian axes with linear scales for the whole analyzed time scan, as shown in Fig. 1b, we realize that there is no difference between the two. The data lie on the bisector line of the quadrant meaning that the two set of data express identical information; the linear correlation coefficient for the straight line is in this case differs from 1 for less than 0.1% meaning that we can reasonably assume $R_V$(Senato) $\cong$ $R_V$(Camera).

In Italy the Constitution allows to vote for the Chamber of Senators only people 25 years old and above; we can see that, as far as the will to express the vote is concerned, this age constriction does not make any appreciable difference with respect to the Chamber of Deputies, where all people above the age of 18 years can vote. The fact that $R_V$ is the same for both Camera and Senato is likely related to the limited percentage of electors who can make the difference between the two Chambers:
these are just those with the age in the interval (18-25) with respect to the total numbers of voters at the time of a given election.

In Fig. 2a we show, for the Camera, a plot where we put $R_B$ on the horizontal axis and $R_S$ on the vertical axis; we can see here that the data, except for the circled point [3], fall mostly on two parallel lines whose slope is equal respectively to $2.04 \pm 0.12$ and $2.15 \pm 0.05$ and intercepts $1.04 \pm 0.20$ and $-0.55 \pm 0.10$. The linear correlation coefficient of the two lines are respectively 0.986 (upper line) and 0.998 for the lower line. Thus, the two slopes are consistent within the errors and their weighted average is $2.11 \pm 0.05$ meaning that, the percentage of spoilt ballots can be obtained, with a rough 4% discrepancy, from percentage of blank ballot by multiplying a factor two. We note that all the 7 points fitted by the lower line are results of elections up to 1994 (included) while all the results after 1994 lie in the upper curve. Apart for the specific numerical proportionality constant, this plot shows that the two parameters are linearly correlated: the overall amount of spoilt ballots does not correspond to random mistakes made by the electors, but it is related to the same intention expressed by the blank ballots to not vote for any party. It is known in fact that most of the spoilt ballots contain just scribblings and graphic fantasies indicating, more than a mistake, the will to not vote. A plot very similar to the one of Fig. 2b can be obtained for the Senato.

From what we said in the previous paragraph we argue that it is reasonable to consider the sum $R_T = R_S + R_B$ as an indicator of the intention of the voters to not express a vote in favor of any political party in a given election. In Fig. 2b we show the time dependence of the parameter $R_T$ (Camera) over the whole time span investigated where an overall rise of the parameter in time is evident. Although the behavior does not seem regular at a first sight, we see in Fig. 2b that the data of four elections are very well intercepted by a straight line and the other points are so well distributed below and above this line that the linear correlation coefficient of the fit is essentially the unity. We can see, however, that the points circled by the oval in the plot, in the low-right part, are clearly far from the tendency to increase in time of $R_T$. These points correspond to the data relative to elections
of 2006, 2008, 2013, 2018. The dashed arrowed line in the figure is a guide to the eye evidencing the abrupt factor 2.8 decrease of the $R_T$ occurred between the election of 2001 and those of 2006 whose interpretation is not easy. One could expect $R_T$ tending in time to a sort of plateau value, assuming that in a sane democracy this parameter cannot grow forever, but this expectation would be somewhat validated by a more smooth temporal variation and not by a sharp, and substantial, decrease like that shown in Fig. 2b. Even admitting that the election of 2001 $R_T$ could be affected by an unfortunate casualty [3], it is unavoidable to admit from 2006 onward a substantially different “tendency”.

Another interesting information comes from the plot of Fig. 3 which further clarifies the “spurious” nature of four points of Fig. 2b. In this figure we report on the horizontal axis the parameter $R_T (Camera)$ while on the vertical axis we report the parameter $R_V (Camera)$; recall that $R_V$ is the number of people expressing their will in an election normalized to the total number of individuals having right to vote. The straight line we see between the data is a linear fit which returns a negative slope equal to $-0.015\pm0.002$ and an intercept equal to $99.7\pm1.4$. The linear correlation coefficient of the line returned by the statistical fit is 0.912. The red points inside the oval contour correspond to the same ones inside the oval in Fig. 2b; even we see here that these points are far from the somewhat regular dependence of the other election points. Fig. 3 is interesting because it shows that the “historical” tendency is that more people participate to the elections the less is the amount of invalid votes: in other words an increased percentage of voters indicates an increased tendency to express a valid vote. Thus, the conjecture that an increase of the number of voting electors increases the probability of getting valid votes makes sense, but this is an overall result which does not relate the increase to any specific politic direction, it is just an overall tendency.

The straight line in Fig. 3 can be interpreted as a politics “approval-rating line” and the interesting thing is that the intercept to zero of this line is $100\%$, within a $1.4\%$ error, meaning that if all the electors having right were to vote the percentage of “unexpressed” votes would tend to zero. It would be definitely interesting to analyze similar data for other countries. From what we just said,
however, it is clear that the points inside the oval in Fig. 3 (the same in the oval in Fig. 2b) represent and evident deviation from the linear tendency: these point, in Fig. 3 correspond to a low $R_V$ and a low $R_T$ which are far from the “approval-rating line” (representing a 50 year tendency).

Overall, the statistical analysis herein reported is at undergraduate level and the straight lines interpolating the data are just least square fittings. All the linear fittings, however, have correlation coefficients above 0.9 meaning that, considered that in the worst case the number of data generating the fit was 7, the probability of the considered data not being linearly correlated was less than 0.05% [4]. This result represents a reasonable degree of confidence for the presented analysis. It is remarkable that complex socio-political phenomena analyzed over a substantial time span display such simple dependencies. However, this it is not true in general and we can realize it in Fig. 1a, Fig. 2b and Fig. 3. In Fig. 1a we observe a significant discontinuity in the slope of the straight lines beginning after 1979, while in Fig. 2b and Fig. 3 we see that an abrupt transition taking place in 2006 takes the response far away from what one could expect from a 50 years “regular” tendency. The general behavior of natural phenomena is nonlinear and the reduction/approximation to linear responses often provides just a first approach in taming problems, and often for slight intervals of the independent variables. Here instead we see a scenario a bit different for the response of a percentage of voters: it is mostly linear with some discontinuities/irregularities.

It would be interesting to justify the objective anomalies from the perspective of the socio-political mass phenomena. The conclusion that we extract from the data is that overall, simply ruled, relations exist, but also “singular” points in which we see the response of the percentages of voters we are interested in changes abruptly. The only “technical” treatise we are aware on for of the onset of discontinuities in behavioral science, decision making, biology and other disciplines was put forward by Renée Thom [5]. Thom’s analysis was concerned with a general model for the singularities of nonlinear dynamical models, and related potentials. Excellent sources of information and references for this model can also be found in refs. 6 and 7. Within Thom’s analyses an abrupt
transitions in behavior is caused by the background existence of different stable states. He classified seven of these kind of abrupt discontinuities baptizing these as “catastrophes”. We can say that the stepping observed in Fig. 1a from a very high, almost “constant”, participation of citizens to the voting process, to an attitude expressing a constantly decreasing participation is reminiscent of the sharp transitions due to the folding of behavioral surfaces described in catastrophe theory [8]. In this specific case the two stable states could be “participation” and “indifference”, both in relation to political life. In these terms it is clear that from 1979 on “indifference” prevails generating a constant decrease of people participating to the elections. This is, however, all we can say since framing, technically, the results presented in terms of the 7 catastrophes models would be hardly achievable at this point. Commenting the socio-political reasons generating the discontinuity is not a purpose of the present work, we just present, and organize, the available data.

We judge that an abrupt change in behavior is also what occurs in Fig. 2b. While Fig. 1 concerns the electors who decide to not vote, namely to not express their opinion at all, Fig. 2b concerns the percentage of the electors who, voting, decided to make clear the intention to not like any political party (by a blank or spoilt ballot). Even in this case we see an abrupt transition from a tendency to increase constantly in time of \( R_T \), whose duration is about 50 years, to substantially lower values observed in the last four elections. We have also seen that this recent tendency, generates an abrupt change in another parameter plane: in particular, the plane \((R_T, R_V)\) of Fig. 3 shows a noticeable regularity of the relation between these two parameters for almost 60 years but, while the transition that started in 1979 did not affect the relation between these two parameters, the transition of 2006 did. Thus, what we can say is that the four “anomalous” points of Fig. 2b and Fig. 3 is an unprecedented result of political expression and it would be surely interesting to further investigate what the causes for the phenomenon are.

It is also worth mentioning that, apart for the discontinuity occurring after 1979 in Fig. 1a, the percentage of electors voting \( R_V \) has proceeded along the negative slope until 2018. Thus, the
discontinuity and abrupt change of Fig. 2b cannot related to the overall percentage of electors deciding to vote since this parameter has kept the same tendency. The “spurious points in the ovals of Fig. 2b and Fig. 3 are evidence of an “internal” structure inside the parameter $R_T$ which is not related to the “regular” overall decrease $R_V$. 
REFERENCES

1) https://elezionistorico.interno.gov.it/index.php?tpel=C

2) M. Cerruto, *Electoral Participation in Italy*, in Quaderni di Sociologia, 60 (2012), https://doi.org/10.4000/qds.537; see also references therein.

3) The circled point in the figure refers to the elections of 2001 when a fault in Government organization caused people stay in queues for voting for several hours. Who is writing, for example, spent more than two hours in a queue to vote. In these conditions a reaction from the voters to the rather uncomfortable situation could be expected. This fault, however, has only taken place in 2001. In all the other Italian elections the average waiting time for voting has been roughly 10 minutes or less.

4) See, for example, J. R. Taylor, "*An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*", University Science Books, 55D Gate Five Road, Sausalito, CA 94965 (1997); ISBN 0-935702-42-3 (cloth)-ISBN 0-935702-75-X (pbk.).

5) R. Thom, *Structural Stability and Morphogenesis: An Outline of a General Theory of Models*, W. A. Benjamin, Inc, (1972).

6) P. T. Saunders, *An Introduction to Catastrophe Theory*, Cambridge Univ. Press (1980)

7) E. C. Zeeman, *Catastrophe Theory*, Scientific American 234, 65-83 (1976)
FIGURE CAPTIONS

Figure 1. (a) Time dependence of the ratio $R_V$ between the number of voters and the total of all the
citizens having right to vote. The negative slope of the straight line indicates a rough 2% decrease every 5 years, the time duration of a Parliament in charge in Italian Constitution.
(b) The same ratio as above for Camera and Senato on the Cartesian axes shows that there
is no distinction between the two percentages which lie on a straight line.

Figure 2. (a) The same kind of plot like in Fig. 1b where we report now on one axis the percentage
of blank ballots and on the vertical the spoilt ballots. The slopes of the two straight lines,
which give roughly 2 as an average, tell us that for every blank ballot there are 2 spoilt
ones. The linear correlation between $R_B$ and $R_s$ tells us that these correspond somehow
to the same intention to not express a vote. (b) The time dependence of the sum of the
percentages of blank and spoilt ballots $R_T$ as a function of time. Even here we can see a
grossly linear dependence (see text for fitting parameters of the shown straight line).
However, the red points in the oval are far from the linear growth tendency.

Figure 3. The correlation between the percentage of voters at the Camera and the percentage of
these responding with blank and spoilt ballots. The four points inside the oval perimeter
correspond to the same shown in Fig. 2b in the “time domain”. We define the straight line
as a politics “approval-rating” line.
M. Cirillo, Figure 1
M. Cirillo, Figure 3