The beginning of the electronic vote in Venezuela and the debate on its performance

Guillermo Salas

Abstract

The Referendum that took place in Venezuela in 2004 coincided with the beginning of the electronic voting in the country. Ever since, there has been a great debate regarding whether there was an electoral fraud through the manipulation of the electronic vote, and whether it was proven or not. There is a null hypothesis that states there was no electronic fraud, and an alternative hypothesis that states that in computerized voting centers the official number of votes was forced through malicious manipulation of the electronic vote, which correlate linearly with the number of signatories that solicited the Referendum, and multiple statistical evidence that is not compatible with the null hypothesis. One of these statistical evidences, that No Votes reported by the voting machines ignored the Benford's Law, has already been recognized by the academic world as a anomalous behavior. In previous articles the probability that these anomalies were present in the official results had been estimated that in the order of 10^{-36} (virtually zero). In the present article it is stated that a family mathematical models, compatible with the alternative hypothesis, produces results with the same anomaly. Also, the arguments that support the null hypothesis and the group of arguments that have been published in Statistical Science by Raúl Jimenez against the statistical evidence that corroborate the hypothesis of generalized electronic fraud are refuted. Finally, this evidences, those ones that corroborate the hypothesis of generalized electronic fraud, are presented in a coherent and structured way.

1. Introduction

The Presidential Revocation Referendum that took place in Venezuela in 2004 (RRP-2004) is an extraordinary event to analyze the inherent risks of electronic voting.

The Venezuelan government at the time, presided by Lt. Col. Hugo Chávez, controlled the directive of the National Electoral Council (CNE in Spanish) in charge of organizing RRP-2004. In this electoral process there were voting centers where the voting procedure was manual and other voting centers (where the majority of the population voted) where the voting procedure was computerized, which allowed us to compare behaviors. That was the beginning of electronic voting in Venezuela, and the debut of voting machines of the company Smartmatic.

The audit of the results reported by the voting machine scheduled to be done immediately after the closure of the voting centers was abruptly prevented by the staff of the CNE. Three days after RRP-2004 (time in which the electoral urns were under the

supervision of the Armed Forces), the audit took place. In this occasion, the CNE staff did not impede it.

In both types of voting centers there were many stations to cast the votes. In the centers in which the computerized automatic voting system was used, there was a voting machine for each station. Table 1 shows how voting stations were distributed at the computerized voting centers.

Table 1

Number of Voting Stations	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Number of Electoral Centers	333	1060	829	682	505	392	309	213	112	57	42	19	12	4	5	2	2	1

The electoral centers located in urban municipalities, where the audits were to take place, the machines had a fingerprint readers connected to laptops which, simultaneously, were connected to a parabolic antenna that was permanently transmitting information (See Fig. 1).

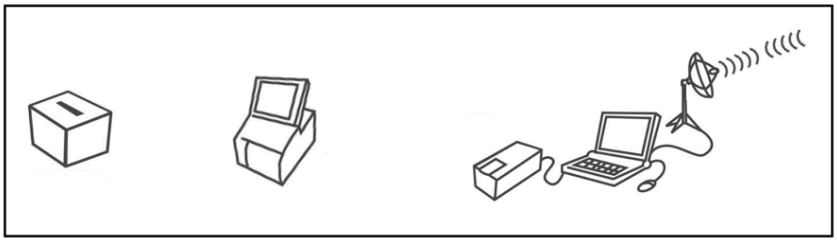

Fig. 1: The structure of computerized electoral centers: voting machine, electoral urn, fingerprint readers, laptop and parabolic antenna for satellite transmission

The electorate had only two options: “YES” to revoke the president’s time in office and “NO” for him to remain in office (in manual voting centers there were also the possibility of “null vote”).

In computerized voting centers the voter pressed a touch screen in the voting machine. Afterwards, the machine printed a paper ballot called “voting proof” (“comprobante de votación”) with the voter’s choice. The votes counted were not the paper ballots but the electronic votes generated automatically by the voting machines, which were connected to a central server and after the transmission was over, the acts with the results were printed.

Due to the fact that the CNE did not allow for the electoral urns to be opened and count the paper ballots, and that no audit could endorse the veracity of the results printed by

the machines (See Ref. 1. Section 9 and 10), the only mechanism left for Venezuelans to evaluate the results was to analyze them through statistical methods.

For the analysis, there was a considerable amount of data that the CNE had published and therefore could not hide. Amongst others:

- The list of petitioners of the Referendum. This list, which included the ID numbers of the signatories, was published by CNE in two papers of national circulation.
- The list of petitioners of the Referendum which's signatures were rejected because the CNE considered them "firmas planas" (similar calligraphy). This list, with the respective ID numbers, was published by the CNE in two papers of national circulation.
- The list of petitioners of the Referendum sent to "*Reparo*" (voters had to ratify their signatures). This list, with the respective ID numbers, was published by the CNE in two papers of national circulation.
- The Electoral Registry, which includes each voter's respective voting center.
- The location by State, Municipality and Township of every voting center.
- The official electoral results: number of YES votes and NO votes by voting station.
- The list of the 199 voting centers chosen for the audit that was scheduled to be done immediately after the voting centers' closure.
- Results of other elections.

And data from other sources:

- The list of the computerized voting centers, which was published by Aviel Rubing, member of the Panel of Specialists the Carter Center assembled to evaluate the allegations of fraud based on statistics <http://www.venezuela-referendum.com/>.
- The list of the 27 voting centers where the CNE did allow opposition witnesses to be present in the audit of official results. This list was done by the Coordinadora Democrática, the coalition that represented the YES option (also the NGO Ciudadanía Activa), and handed to the OEA and Carter Center with the endorsement of the Vice-President of the CNE at the time, Dr. Ezequiel Zamora.
- The *log* of the transmissions that were done by the voting machines before printing the results, data that was administered by the telephone company that facilitated such transmissions.
- The results of various exit polls.

Also, there was a Panel of Specialists that the Carter Center assembled *a posteriori* to evaluate the allegations of fraud based on statistics.

2. The null hypothesis of no electronic fraud and the arguments that support it.

The official version of what happened during RRP-2004 originated in the endorsement that the Carter Center gave official results, which are stated in the section *Conclusions*

and Looking Toward the Future of its final Carter Center report *Observing The Venezuela Presidential Recall Referendum Compressive Report* (Ref. 2):

> "The Venezuelan people voted not to recall President Chávez from office, with 59 percent of the population voting for Chávez and 41 percent voting against him. It is the opinion of The Carter Center that the Aug. 15 vote clearly expressed the will of the Venezuelan electorate. The Center did not observe, and has not received, credible evidence of fraud that would have changed the outcome of the vote"

In the end, the official version that this article will consider the null hypothesis was sustained by four arguments:

- The opposition's *quick count* done the night of RRP-2004 by adding the results of the official acts (mainly done by voting machines), coincided with the official result.
- The audit that was done three days after RRP-2004 proved the veracity of the official results.
- The official results coincided with the results of some polls.
- There is no credible evidence of fraud.

The first argument does not uphold the veracity of the official results, because the fraud would have taken place in the scrutiny of the electronic vote, which was the least transparent part of the process. This *quick count* would not have been able to detect it.

The second argument does not sustain the veracity of the official results, amongst other things because the urns with the paper ballots that were counted in the audit three days after RRP-2004 did not have a certified origin, they were provided by the Armed Forces. What happened in that audit and the mathematical arguments with which the Carter Center supports their validity are widely analyzed and rebut in the Section 10, Cold Audit, of the Ref. 1.

Regarding the third argument, it must be taken into account that the results of an election cannot be validated exclusively because they coincide with the results of some polls, especially when other polls had entirely opposite results (See Ref. 1, 3, 4 y 5).

To debate on the asseveration done to build the fourth argument, that there is no credible evidence of fraud, is precisely the purpose of this article.

3. The statistical evidence inconsistent with the null hypothesis

There is a group of statistical evidence inconsistent with the null hypothesis. The majority of them are based on unexpected behaviors (very unlikely behaviors under the assumption of no electronic fraud) that are present in the official results (See Ref. 6, 7, 8 , 9 and 10) or in the way they correlate to the data associated RRP-2004: list of petitioners of the Referendum, log of voting machines transmissions, results of some exit polls, etc. (Ref. 1, 3, 5 y 11[im]).

One of this unexpected behaviors has already been recognized by the academic world as an anomalous behavior (See Ref. 8 and 12). Also, there is an alternative hypothesis that proposes that the results reported by the voting machines were invented according to the number of petitioners for the Referendum (See Ref. 1 and 5).).

These evidences are coherent and complementary. This consistency and complementarity are summarized in *Section 6. Conclusions* of the present article.

4. The evidence of fraud based on the anomalous behavior

In 2004, a few weeks after the Referendum, two independent reports were presented to the Carter Center: *Evidencia de la Manipulación Artificial de los Resultados al Aplicar la Ley de Benford al Referéndum Venezolano de Agosto de 2004* (*Evidence of the Artificial Manipulation of the Results by Applying Benford's Law to the Venezuelan Referendum of August 2010)*, by I. Mikoss (Ref. 6), and *La Ley de Newcomb-Benford y sus aplicaciones al Referéndum Revocatorio en Venezuela. Reporte Técnico no-definitivo 2a. versión: Octubre 01, 2004* (*Newcomb-Benford's Law and its application to the Revocation Referendum in Venezuela. Non-definitive technical Report. 2nd version: October 01, 2004)*, by L. Pericchi and D. Torres (Ref. 7). Both show the same indication of fraud: NO votes reported by the voting machines ignored Benford's Law for the second dígit [1].

In one of the reports, L. Pericchi and D. Torres (2004) state:

> "The conclusions are clear. The NO votes, in the computerized mesas violate Newcomb-Benford's Law with Probability virtually 1 (the probability of the complementary hypothesis, which comply the Law, is de 1.34×10^{-36} , virtually 0)"

The authors add:

> "This makes it relevant to inquire on the mechanisms of the intervention that altered the NO computerized votes in such a way that they violate so clearly Newcomb-Benford's Law".

In Figure [2] , YES votes and NO votes in manual voting stations comply with Benford's Law, YES votes of computerized voting stations comply and NO votes of computerized stations ignore it.

---

[1] The Benford Law is used in the second digit, because the frequency in the distribution of the first digits is very affected by the imposition of the máximum of total voters in each voting machine. This imposition barely affects the behavior of the second digits (See. Ref. 8).

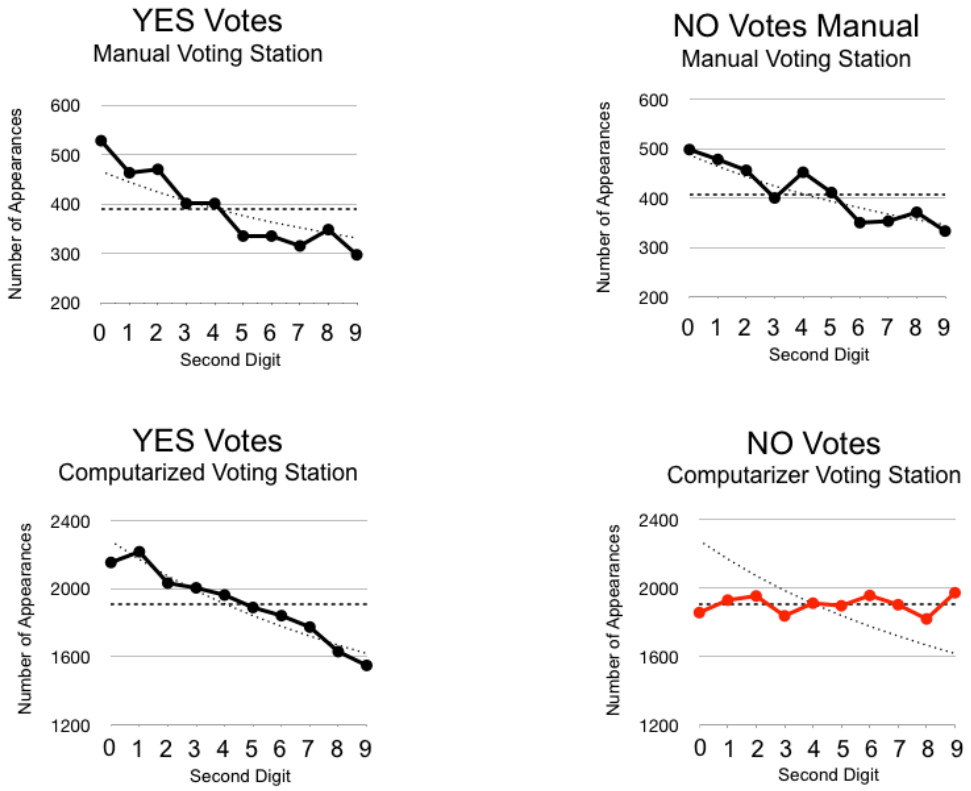

Fig. 2 These graphs show the number of times that every second digit appeared in the official results, YES and NO votes of the manual voting stations (top), and YES and NO votes of voting machines (bottom). Each computerized voting station corresponds to a voting machine. Like references spotted line show the Benford Distribution and the dashed line the Uniform Distribution. serve as references. Observe that NO votes reported by computerized voting machines ignore Benford's Law.

In Figure [3] show that the goodness of fit between the distribution of the second digits of NO votes reported by the voting machines and the Uniform Distribution ($\chi 2$ = 12,9) is the same that would be expected of the second digits of a group of numbers originated by random numbers generator (RGN). This is especially suspicious.

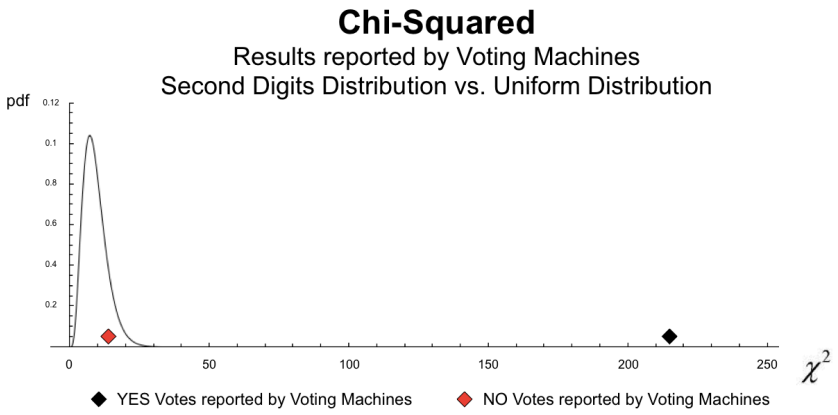

Fig. 3. The graph on the horizontal axis shows the values of Chi-Square between the frequency distribution of the observed second digit and the Uniform Distribution. On the vertical axis, as a reference, the probability density function (pdf) for the chi-square distribution (with 9 degrees of freedom) is shown. Note that the chi-square distribution of the second digit of the No votes is in the area where the probability density function is relatively high, while that for the Yes votes is where this function is virtually zero.

This situation led to investigate how commonly the frequency of the second digits of the results reported for each voting station has goodness of fit to this distribution as well the results were.

To this end, we used 48 results (by voting station) of elections that took place in Venezuela between 1998 and 2008. We obtained the distribution of frequencies of their second digits and then calculated for each of them the χ2 and the *p-value* with respect to the Uniform Distribution. The only votes that fit well with the Uniform Distribution were the NO votes reported by the voting machines in RRP-2004 (See Ref. 9). The results (*p*-values) are in Fig. 4.

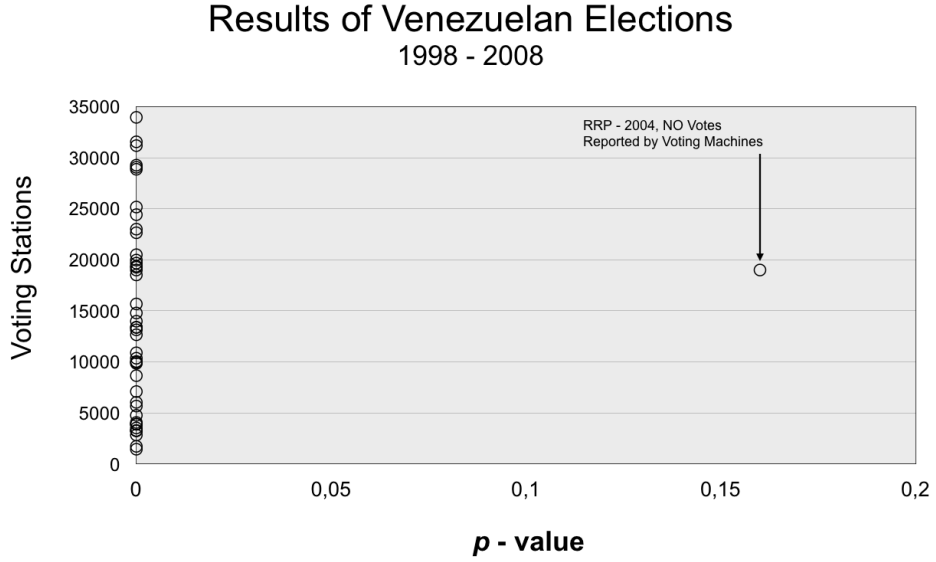

Fig. 4 In the gráfic the horizontal axis of the graph are the *p-value* of the chi-square that exist between the distributions of the second digits of the 48 results observed (by voting station) and the Uniform Distribution. In the vertical axis the number of voting stations of the elections observed (with 10 or more votes). Observe that the only result that fit to the Uniform Distribution (whit a *p-value* non close to 0), is the NO votes reported by voting machines in 2004 (pointed with an arrow).

Despite this strong indication of fraud, ever since its publication there's been an effort to undermine it. In 2004, when it was presented to the Carter Center, they analyzed and answered:

> "More recently, claims have been made by Mikoss and by Pericchi and Torres that a comparison of the recall referendum results with "Benford's Law" shows that the results are fraudulent"

To evaluate the report, the Carter Center limited itself to compare three groups of electoral results produced by computers: the results of the voting machines and results produced by two simulations. The results of both simulations, as did the results of the voting machines, violated Benford's Law. Based in that observation, Carter Center Ref. 2) said:

> "The panel concludes that there is insufficient evidence that Benford's Law applies to election results in general. Furthermore, a simple but plausible model of the election does not produce results that conform to Benford's Law".

With time, after observing the way real electoral results behave and thanks to the formulation of more realistic mathematical models than those used by the Carter Center in 2004, the fact that Benford's Law does apply to electoral results is now a reality, at least in the academic world (See. Ref. 8 and 12).

After this acknowledgement, and in spite of the fact that the NO Votes of voting machines were considered anomalous, in 2011 two arguments were published in *Statistical Science* to continue to diminish the importance of the evidence:

1. The fact that the number of votes reported by the voting machines violate Benford's Law does not imply that the voting machines were biased for one of the options. In other words, a series of non-biased errors in the counting of votes (which would have to annul each other) could produce this anomaly (biased; YES votes comply with Benford's Law and NO votes ignore it) in the official results.

2. Given the fact that the result of the sum of YES votes and NO votes by voting center do comply with Benford's Law, this indication of fraud does not add anything to the controversy about the veracity of the official results.

In the article *Forensic analysis of the Venezuelan recall referendum* (Ref. 13), R. Jimenez states:

> "We should remark that violations of Benford's Law may be due to unbiased errors".

And based on the fact detected by Jimenez, that the result of the sum of YES votes and NO votes by voting center do comply with Benford's Law, he adds:

> "All voting options confirm the law. According to this analysis, there is no reason to doubt the official results by center, despite that the test suggests the contrary when we use the results by notebook. Is the former a false negative or the latter a false positive? Could unbiased errors in the vote counting by notebooks reproduce such a scenario? Or, conversely, could the results by centers be masking a fraud in notebooks? Benford's test does not address this controversy"

When R. Jimenez published this article no one had yet left written evidence of the link between types of frauds that bias the results and the way results that reported the voting machines violated Benford's Law.

4. The alternative hypothesis

Previous Notes

According to the Venezuelan Constitution, for the CNE to call for the Referendum it was necessary that it received a petition with at least 20% of the electorate (entered in the Electoral Registry). In a three day event organized by the CNE the names, ID numbers and signatures of the petitioners were collected. This list was published by the CNE in two papers of national circulation and also handed to the OEA and the Carter Center.

There were 2.674 centers the electorate could choose from to go sign the petition for the RRP-2004; there were 8.394 voting centers, and every voter was assigned to a specific one. When crossing the list of signatories with the Electoral Registry data (which specifies each voter's voting center), it is possible to determine the number of signatories assigned to each voting center. We will call this variable "Signatories".

The presumed intention of the signatories was to vote YES in the RRP-2004. The relations between YES votes (variable "YES votes") and "Signatories" in the computerized voting centers is the base for the alternative hypothesis.

The Carter Center's Inform (Ref. 2) includes an analysis of the relation between YES votes and Signatories for each computerized voting center. The report states the following:

> "A very high correlation between the number of signers and the number of YES votes per center in the universe of automated voting machines has been found a correlation coefficient of 0.988. This means that in voting centers where a high signer turnout was obtained, a high YES vote also was obtained."

It further concludes:

> "There is a high correlation between the number of YES votes per voting center and the number of signers of the presidential recall request per voting center; the places where more signatures were collected also are the places where more YES votes were cast. There is no anomaly in the characteristics of the YES votes when compared to the presumed intention of the signers to recall the president."

The article *of the 2004 Venezuela Referendum: The Official Results Versus the Petition Signatures* by G. Delfino and G. Salas (Ref. 1) proposes the following alternative hypothesis:

> "In computerized centers, official results were forced to follow a linear relationship with respect to the number of signatures".

This would be a deceitful and subtle way to subtract votes from the YES option (See Ref.[ds] y [rm]).

In references [1] and [5] shows a group of unexpected facts inconsistent with the null hypothesis and which's occurrence can be explained with the alternative hypothesis.

In Ref. [5], R. Medina proposes an algorithm that reproduces the type of relation between YES votes and Signatories in computerized voting centers.

Medina's Algorithm:

$$\text{YES votes} = 1{,}16 \cdot \text{Signatories} + (t \cdot 3{,}6 \cdot (1{,}16 \cdot \text{Signatories})^{-1/2}).$$

The variable *t* represents a number with normal distribution of media zero with standard deviation of 1, which is obtained with a generator of pseudorandom numbers.

Is the anomaly present in the official results consistent with this alternative hypothesis? Is it consistent with R. Medina's algorithm?

5. Consistency between the observed anomaly and the alternative hypothesis

5.1 Number of signatories per voting machine

According to this alternative hypothesis the number of YES votes reported by voting machines are a series of multiple changes of scale of the numbers of Signatories assigned to each machine. The number of signatories per voting machine [2], as showed in Fig. [5], comply with Benford's Law.

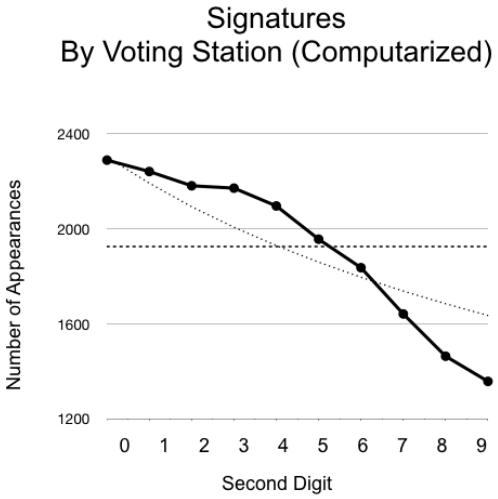

Fig. 5 This graph shows how the distribution of second digits of the numbers of signatories per computerized voting station comply with Benford's Law. The Benford's Distribution and Uniform Distribution (spotted and dashed lines) are included as references.

5.2 Simulation-0

To do the simulation-0 of the number of YES votes per voting machine according to the alternative hypothesis we will use:

[2] The number of signatories per voting machine was calculated as follows. Based on the list of signatories and voter registration (valid for the referendum) determined that voters were assigned to each voting center. Then, taking into account the last two numbers of the identification document of the signatories, was distributed (following the same methodology used by the CNE) between the different voting machines.

YES votes = 1,16 x Signatories

In this simulation the same factor or change of scale (1.16) is applied to all the numbers of signatories assigned to each voting machine to obtain the YES votes.

The alternative hypothesis omits everything about NO Votes, which are precisely the anomalous ones. For this Simulation-0, they are calculated with a simple generator of random numbers (number of NO Votes = random number between 0 and 600 [3]). Regarding how much this choice limits the result of the simulation is discussed in Section 5.4.

Fig. [6] shows the way results of voting machines and results of Simulation-0 behaved according to Second Digit Benford's Law. In both cases YES votes per voting machine complied with Benford's Law, NO votes and Total votes ignored it, and in the case of the sum of Yes votes and NO votes per voting center, the results do comply with Benford's Law.

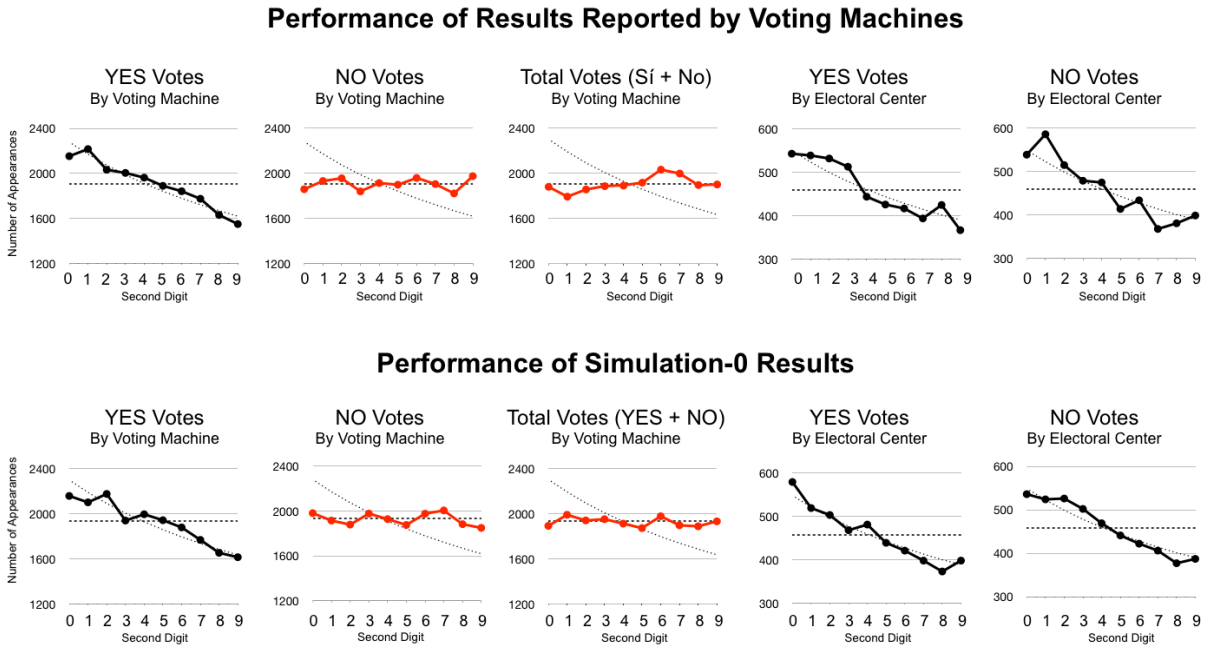

Fig. 6. Behavior of computerized results and behavior of simulation-0 results. These graphs show the number of times that every second digit appeared in YES vote per voting station, NO votes per voting station, Total votes (YES + NO) per voting station, YES votes added per voting center and NO votes added per voting center (Top refers to the results reported by voting machines and bottom Simulation-0). The Uniform Distribution and Benford Distribution (spotted line) are added as references. Observe that simulation-0 reproduces the behavior and anomalies that are present in the official results.

[3] 600 is the maximum numbers of electors by voting machine.

It is clear that YES votes reported by voting machines do not follow the extremely simple behavior of Simulation-0. Can more complex mathematical models in which to calculate YES votes per voting machine used no the same change of scale is applied for all the voting machines produce results that comply with Benford's Law?

5.3 Simulation with Medina's algorithm

YES votes will be simulated following Medina's proposed algorithm (See Ref. [rm]). According to this algorithm, that refers only to the number of Votes Yes for Voting Centers, there are many possibles way to calculate, from the Signatures, the YES votes per voting machine. The chosen way to simulate YES votes per voting station (voting machine) will be the following:

For each of computerized voting center the numbers of YES Votes were estimated using Medina's Algorithm. In base a this estimation, to each voting center a value k_i was assigned.

k_i = YES Votes / Signatories

Then for all voting stations (voting machine) of the same electoral center Yes Votes were simulated (using the k_i value assigned):

YES Votes = k_i x Signatories

Whit this procedure, that use differences factor or change of scale (k_i) for the voting station of differences voting center, the YES Votes per voting station were simulated.

Whit this procedure, that uses the same factor or change of scale (k_i) for the voting stations of same voting center, but uses differences factor for the voting station of differences voting center, the YES Votes per voting station were simulated.

This procedure was done 1,000 times. The resulting behavior is in Fig. [7].

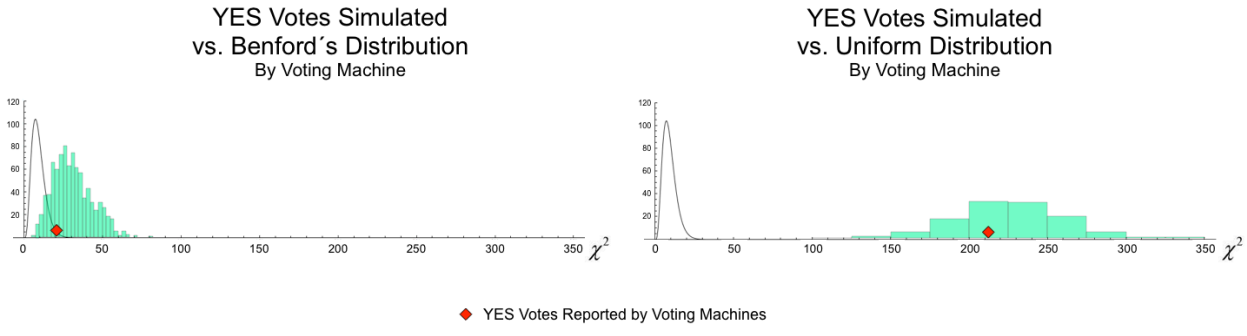

Fig. 7. The left graph on the horizontal axis shows the values of Chi-Square that there are between the frequency distribution of the second digit YES votes simulated and the Benford's Distribution. The right graph on the horizontal axis shows the values of Chi-Square that there are between the frequency distribution of the second digit YES votes simulated and the Uniform Distribution. In both graphics the vertical axis is used for show the chi-square histogram that resulted from the YES votes coming from the simulations. Note that chi-squared of YES votes reported by voting machine (red diamond), is consistent whit simulates ones. As a reference, the probability density function (pdf) for the chi-square distribution with 9 degrees of freedom (multiplied by 1,000) is shown.

In these results, YES votes are the product of different changes of scale, note that as the official result, comply with Benfor's Law [4].

The calculation of NO votes, using the same methodology as simulation-0, was done 1000 times. In figure [8], the behavior of these 1000 simulations are compared to the official result.

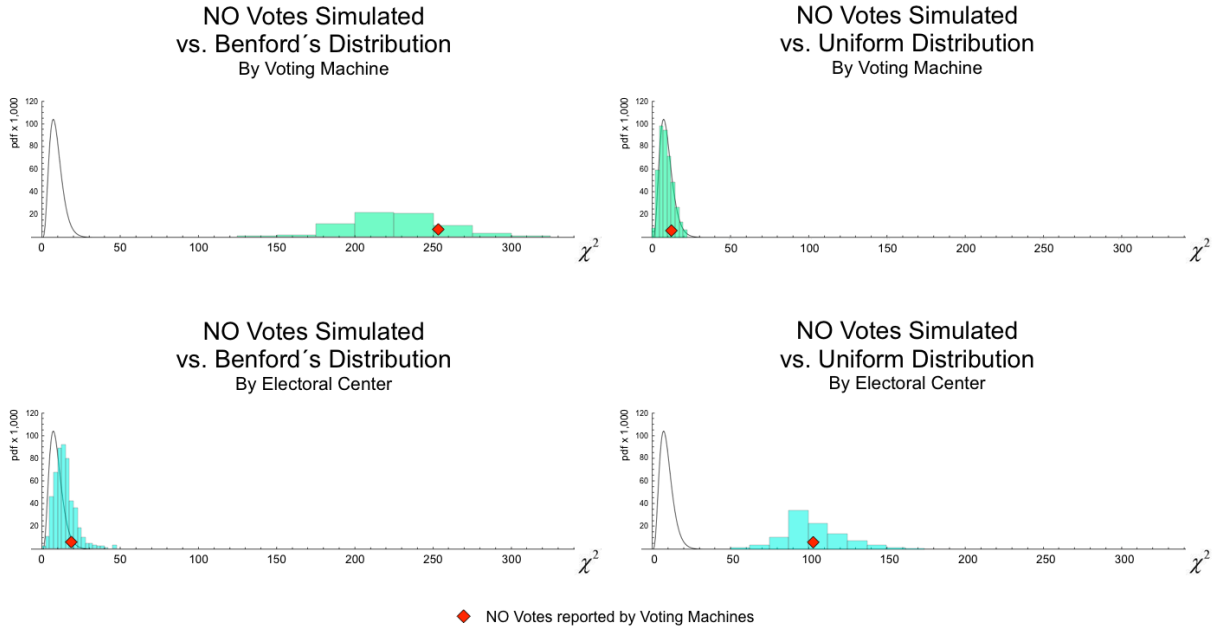

Fig. 8 The left graphics on the horizontal axis shows the values of Chi-Square that there are between the frequency distribution of the second digit NO votes simulated and the Benford's Distribution. The right graphics on the horizontal axis shows the values of Chi-Square that there are between the frequency distribution of the second digit NO votes simulated and the Benford's Uniform. In all graphics the vertical axis is used for show the chi-square histogram that resulted from the NO votes coming from the simulations. The graphics in the top refers to NO Votes by voting station (voting machines). The graphics in the bottom refers to NO Votes by voting Center. Note that in all case chi-squared of NO votes reported by voting machine (red diamond), is consistent whit simulates ones. As a reference, the probability

4

density function (pdf) for the chi-square distribution with 9 degrees of freedom (multiplied by 1,000) is shown.

The NO votes of voting machines behave the same way as the results of the simulation; by voting station they ignore Benford's Law (they fit well to the Uniform Distribution) and added by voting center they comply with Benford's Law [5].

Finally, to determine how Total Votes behaved in these simulations, we added the YES votes of the simulation-0 and NO votes of the 1,000 simulations. In Fig. [9], the behavior of Total Votes reported by the voting machines are compared with those of the simulations.

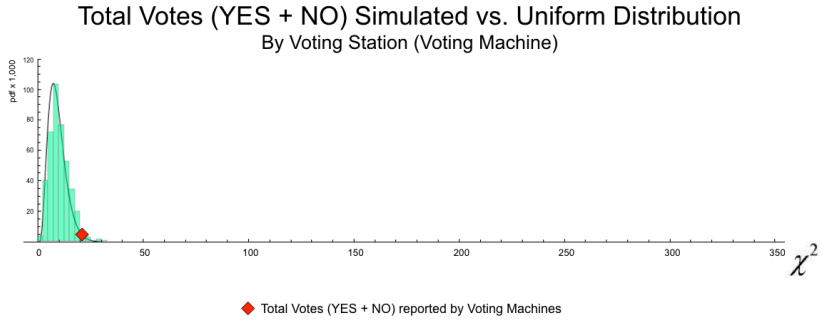

Fig. 8 The graph on the horizontal axis shows the values of Chi-Square that there are between the frequency distribution of the second digit Total Votes (YES Votes + NO Votes) simulated and the Uniform Distribution. Note that chi-squared of Total votes reported by voting machine (red diamond), is consistent whit simulates ones. As a reference, the probability density function (pdf) for the chi-square distribution with 9 degrees of freedom (multiplied by 1,000) is shown.

Note in Fig. [9], Total Votes product of the simulations and Total Votes of the voting machines have the same anomalous behavior; they ignore Benford's Law (they fit well to the Uniform Distribution).

5.4 About mathematical models that reproduce these types of behaviors

The mathematical models used to do the previous simulations in many aspects do not coincide with the official results. For example, the models of some machines produce more number of voters than the number of potential voters registered to vote in those machines. But there are mathematical reasons, asseverated by a group of mathematical models far more complex than those used in these simulations and

[5] The discussion in Apendix X, makes relevant that the number of simulated NO votes added by voting centers comply with Benford's Law.

consistent with the alternative hypothesis, that explain results with the same behavior of the official results:

1) When a change of scale is done to each number that belongs to a set of numbers that comply with Benford's Law (not necessarily the same change for each one of them), the resulting set of numbers will comply with Benford's Law.
2) When a random number from a distribution of equipotencial probabilities (such as those produced by RNG) is added to a group of numbers that does comply with Benford's Law, the resulting set of numbers could ignorates Benford's Law (the distribution of frequencies of appearance of the second digit of the numbers might tend to the Uniform Distribution)
3) When a random number from a distribution of equipotential probabilities (such as those produced by RGN) are distributed in a random way in different subset, the sum of numbers that belong to each resulting subset comply with Benford's Law.

The first of these mathematical properties makes the mathematical models of each voting station force the YES votes to be proportional to the number of signatories and therefore produce YES votes that comply with Benford's Law. The fact that mathematical models (compatible with the alternative hypothesis) that by voting station impose YES votes to be proportional to the number of signatories, produce a number of YES votes that comply with Benford's Law, is a direct consequence of such property.

The second mathematical property makes: a) Mathematical models of each voting station that alter NO votes adding numbers originated in a generator of random numbers, produce NO votes that ignore Benford's Law (the distribution of frequencies of appearance of the second digit of numbers tend to the Uniform Distribution); and b) Mathematical votes that impose, by voting station, that Total Votes (simulated YES votes plus simulated NO votes) be the sum of a set of numbers calculated in such a way that YES votes comply with Benford's Law. For example, simulated YES Votes (that comply Benford's Law) plus NO Votes simulated (that ignore this law) produce Total Votes (YES votes plus NO votes) that ignore Benford's Law (the distribution of frequencies of appearance of the second digit of the numbers tend to the Uniform Distribution).

The third mathematical property makes NO votes that ignore Benford's Law comply with it when added by voting center.

In the Appendix A there are a discussion about this three mathematical properties.

6. Conclusions

The evidence that corroborate the thesis of electronic fraud can be divided into two groups: statistical and non-statistical.

The group of irregularities that occurred during the preparation of the Referendum, the voting procedure, the scrutiny of the votes and the audits of the results are part of the non-statistical evidence and are exposed in the article *Fraude a la Democracia, Informe Final* (Ref. 4).

The statistical evidence that corroborates the same thesis is presented in the following series of articles (organized by date of publication):

1. En la búsqueda del Cisne Negro: Análisis de la evidencia estadística sobre el fraude electoral en Venezuela (In search of the Black Swan: Analysis of statistical evidence of the electoral fraud in Venezuela), by R. Haussman and R. Rigobon.
2. Evidencia de la Manipulación Artificial de los Resultados al Aplicar la Ley de Benford al Referéndum Venezolano de Agosto de 2004 (Evidence of the Artificial Manipulation of the Results when Applying Benford's Law to the Venezuela Referendum of August 2004), by I. Mikoss
3. La Ley de Newcomb-Benford y sus aplicaciones al Referéndum (The Newcomb-Benford Law and its application to the Referendum)
4. Revocatorio Venezolano (versión: Octubre 01, 2004) (Venezuelan Recall) (version: October 01, 2004), by L. Pericchi and D. Torres
5. Elections Forensics: Vote Counts and Benford Law, by W. Mebane
6. A statistical approach to asses referendum results: The Venezuelan recall referendum of 2004, by M. Febres Cordero and B. Marquez
7. Analysis of the 2004 Venezuelan Referendum: The Official Results Versus the Petition Signatures, by G. Delfino and G. Salas.
8. Quick Anomaly Detection by the Newcomb-Benford Law, with Applications to Electoral Processes Data form the USA, Puerto Rico and Venezuela by L. Pericchi and D. Torres
9. The 2004 Venezuelan Presidential Recall Referendum: Discrepancies Between Two Exit Polls and Official Results, by R. Prado and B. Sanso
10. 2004 Venezuelan Presidential Recall Referendum (2004 PRR): A Statistical Analysis from the Point of View of Electronic Voting Data Transmissions, by I. Martin
11. Comprobación de la hipótesis de Delfino y Salas (Verification of the hypothesis of Delfino and Salas), by R. Medina
12. The present article.

Even though these articles study the veracity of official results of computerized voting centers from diverse points of view, the evidences that corroborate the thesis of fraud resulted coherent and complementary. These are a few of such evidences:

1. The exchange of information of the voting machines and a central server just before the acts were printed, is that of an expected generalized electronic fraud. There were two patterns of exchange of information, both bi-directional, and in both of them the voting machines received more information than they sent (Art. 10)
2. None of the audits that were done could endorse the results reported by the voting machines (Articles 1 and 7).

3. The fact that the CNE personnel impeded the audit that was scheduled immediately after the voting procedure closed was a malicious act, and not fortuitous as stated in the Carter Center's report. The probabilities of chance of all 27 voting centers where CNE personnel did not impede the presence of opposition observers would have a behavior so different from the rest of the computerized voting centers is 2×10^{-8} (Articles 7 and 11).
4. The official results of computerized voting centers are not compatible with the results of two exit polls directed and supervised by Penn, Schoen & Berland (Articles 1, 7, 9 y 11).
5. The official results of computerized voting centers are less credible as the percentage of YES votes diminishes (Articles 1, 6, 7, 8 and 11).
6. The group of numbers of NO votes reported by voting machines is anomalous: it ignores Benford's Law. It is estimated that the probability of having such an abnormal result by chance is 1×10^{-36}.
7. The relations between YES Votes/Signatories in computerized voting centers were unexpected: they did not reflect neither the socio-political heterogeneity in Venezuela nor the fact that the conditions to sign the petition for the Referendum were considerably different from the conditions to vote, and ignore the effect of an essential singularity that are in this relation (Articles 7, 11 and 13).

And this is how they complement each other:

The alternative hypothesis explains why these facts, not consistent with the null hypothesis, occurred.

1. The alternative hypothesis could explains why all the voting machines received more information than the information they sent. The null hypothesis does not.
2. The alternative hypothesis explains why none of the audits of the official results reported by voting machines could be endorsed. The null hypothesis does not. (See Articles [7] and [11]).
3. The alternative hypothesis explains why CNE personnel impeded the audit scheduled for immediately after the voting centers closed, and why the 27 voting centers where the CNE did not impede the presence of opposition observers had a behavior so different from the rest of the computerized tests. The null hypothesis does not. (See Art. [7] and [11]).
4. The alternative hypothesis explains why official results of computerized centers are not compatible with the results in two exit polls (directed and supervised by Penn, Schoen & Berland). It even predicts the structure of the systematic discrepancies between both results. The null hypothesis does not (See Art. [7] and [11]).
5. The alternative hypothesis explains why official results of computerized voting centers are less credible as YES votes diminish. The null hypothesis does not. (See Art. [7], [11] and [13]).
6. The null hypothesis could explain why YES votes reported by voting machines complied with Benford's Law and why the number of NO votes reported by the same machines ignored it completely. The null hypothesis does not. (See Art. [13]).

7. The alternative hypothesis explains why the relations between YES votes and Signatories in computerized centers are unexpected: they do not reflect the socio-political heterogeneity in Venezuela because they did not reflect the fact the conditions to sign the petition for the Referendum were very different from those to vote, and because they ignored the effect of the essential singularity present in this relation that should have had effects in the real results. The null hypothesis does not. (See Art. [7], [11] and [13]).

The null hypothesis (no electronic fraud) is not consistent with the statistical facts shown in the mentioned articles. Neither are the multiple irregularities observed during the voting process.

Given that they are coherent, complementary and blunt, the statistical evidences presented in the referenced articles form a robust, well-founded and documented indication against the null hypothesis.

It is untrue that there are no credible evidences of fraud. What is true, is that there is no credible evidence that the results reported by the voting machines during RRP-2004 represent the will of the electorate.

In electoral matters, it is the electoral organism, not the citizens, the one responsible of demonstrating that the results emitted are true. It is not the citizens' responsibility to prove with statistical maneuvers or any other type of gizmo, whether the results are fraudulent or not. Given the absolute lack of transparency demonstrated by the electronic vote and its scrutiny, and given the fact that none of the audits of the results of the voting machines could validate them, there is no evidence that these results represent the will of the electorate.

Note: the group of arguments that have been published in Statistical Science by Raúl Jimenez against the statistical evidence that corroborate the hypothesis of generalized electronic fraud are refuted in Appendix B.

Appendix A: Transformations between the sets of numbers that satisfy Benford's Law and the sets of numbers that ignore

When random numbers generators (RNG) are created for computers, which ensures that when this is asked 36.000 integers numbers from 1 to 36, the set of numbers that occur tend to have the same amount of numbers 1, 2, 3, etc.. That if asked to generate 100,000 integers numbers between 1 and 1000, the group produced tend to have equal amount of numbers 1, 2, 3, etc.. In other words are created attempting to behave like a roulette.

Benford's Law revealed that the sets of random numbers that occur in nature (or which produce human beings), when its are measurable quantities, with regard to the frequency of occurrence of each digits, are very different to those one produced by the RNG who ignore this law. See Fig. 10. Here are several ways on how can transform and relate these two types numeric sets.

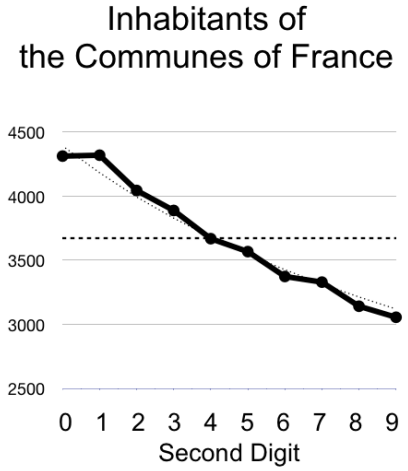

Fig. 10. This graph shows how the distribution of second digits of the numbers of inhabitants of the 36,681 *Communes* of France comply with Benford's Law. The Benford's Distribution and Uniform Distribution (spotted and dashed lines) are included as references.

A-1 How to transform a set of numbers that satisfies Benford's Law in another set that also satisfies this law

It is well known theorem Ted Hill (1995) says that Benford's Law is invariant under change of number base or scale. Which is invariant under scaling implies that if all the numbers of a set that satisfies Benford's Law are multiplied by the same factor, the resulting set will also comply with this law.

Now if we take a single random number that comes from a measurable quantity that satisfies Benford's Law, such as the number of inhabitants of any of 30 comunes which divided France, most likely is that this begins by one. And if we multiply this number by

17, 38 or 77, it is likely that the numbers resulting from these products, beginning with 1. This is a direct consequence of Theorem Hill.

This should cause, that when a set of numbers that satisfies Benford's Law be transformed multiplied their numbers each by a different number, not necessarily the same for each of them, the resulting set of numbers, comply with the Benford's Law.

To corroborate this conjecture, the numbers of inhabitants of the 36,681 Communes of France (that satisfies Benford's Law) were taken, and each was multiplied by a real random number between 1 and 100 out of a GNA, and distribution of their seconds digits was observed. This experiment was repeated 1,000 times. In all simulations observed met the second digit Benford's Law. The results of this experiment are shown in Figure 11.

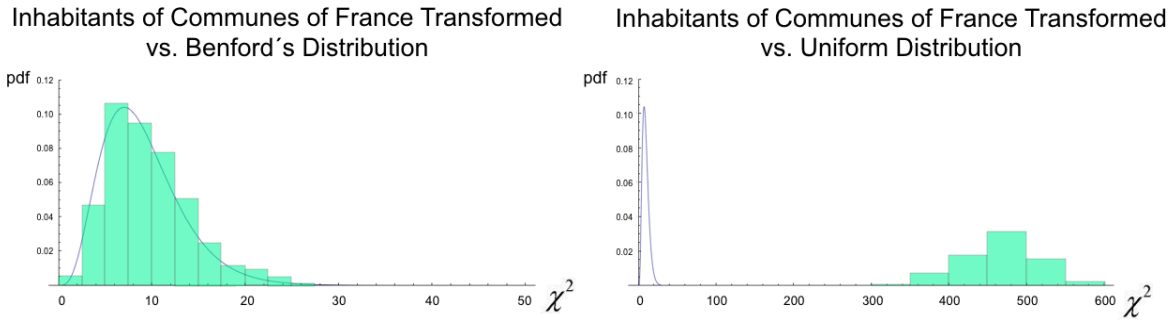

Fig. 11. The left graph on the horizontal axis shows the values of Chi-Square that there are between the frequency distribution of the seconds digits of the inhabitant of Communes of France transformer and the Benford's Distribution. The right graph on the horizontal axis shows the valuesof Chi-Square that there are between the frequency distribution of the second digit of the inhabitant of Communes of France transformer and the Uniform Distribution. In both graphics the vertical axis is used for show the chi-square histogram shape that resulted from this transformations. As a reference, the probability density function (pdf) for the chi-square distribution (with 9 degrees of freedom) is shown.

This experiment and its result is a reiteration of the simulations made in Section 5.3 to calculate the number of Yes votes by voting station from the number of Signatories enrolled in each of these stations.

A-2 How to transform a set of numbers that comply the Benford's Law in another that ignores this law

When the number 1 adds a number between 1 and 100 from a GNA, the resulting number will have almost the same chance of have second digit second 0 that have second digit 9. The same happens when this operation is performed to the number 25 or 63.

So when to each number within a set of numbers that satisfies Benford's Law is added a number from a GNA, not necessarily the same for each of them (and that it magnitude

affects the first digits of the sum), the set of numbers resulting, not tend to meet Benford's Law.

This explains why in Section 5.3 when the simulated YES Votes by voting station, (which satisfies Benford's Law), were added to simulated No Votes by voting station (which ignore Benford's Law), the result, Total Votes (Votes YES + Votes NO) by voting station, ignores Benford's Law.

A-3 How to transform a set of numbers that ignores Benford's Law in another that meets

It is a curious fact that when the numbers of 36,681 Communes of France are invented with a GNA (that produces integers between 10 and 100)[6], and these numbers are aggregated by 4,032 Cantons [7], the number of inhabitants per municipality invented in this way, satisfies the Benford's Law. This test was made 1,000 times, the results are shown in Fig 12.

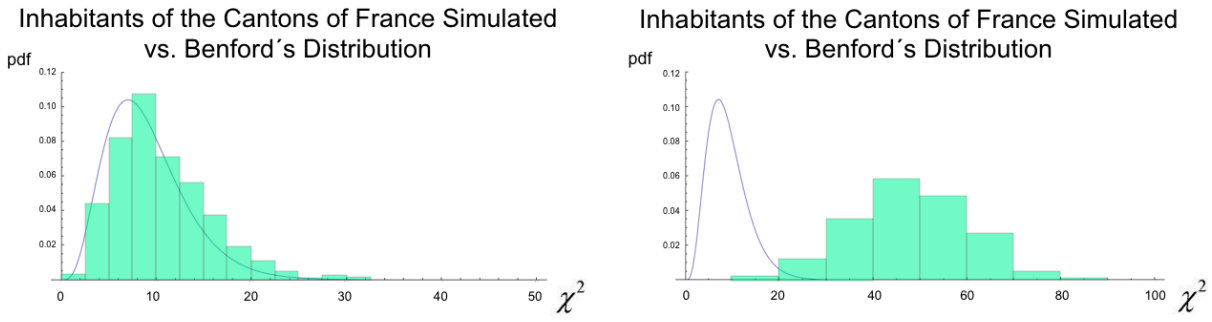

Fig. 12. The left graph on the horizontal axis shows the values of Chi-Square that there are between the frequency distribution of the seconds digits of the inhabitant of Cantons of France, invented (whit with RNG) by Communes and aggregated by Cantons, and the Benford's Distribution. The right graph on the horizontal axis shows the values of Chi-Square that there are between the frequency distribution of the second digit of the inhabitant of Cantons of France, invented (whit with RNG) by Communes and aggregated by Cantons, and the Uniform Distribution. In both graphics the vertical axis is used for show the chi-square histogram shape that resulted from this transformations. As a reference, the probability density function (pdf) for the chi-square distribution (with 9 degrees of freedom) is shown.

This experiment and its result is a reiteration of the mathematical curiosity, made with other data, which was observed in Section 5: Votes No by voting station ignored the Benford's Law, but when are aggregated by voting centers, meets this law.

Appendix B: Arguments that try to distort some of the evidences that corroborate the thesis of generalized electronic electoral fraud

[6] The test was made with integers between 10 and 100, to ensure that everyone had second digit.

[7] The territory of France is divided into 105 département, 342 arrondissements, 4,032 cantons and 36,681 Communes.

The article *Forensic Analysis of the Venezuelan recall referendum* by R. Jimenez, includes comments of the six previous articles about RRP-2004 that had already been approved by journals peer review. These articles were Febres Cordero and Marquez's *International Statistical Review* (2006) and five articles that are part of the Special Section *"Revisiting the Venezuelan referendum"* that *Statistical Science* published in November of 2011. The referred Special Section ends with R. Jimenez article as the last word.

In Jimenez's article the validity of the arguments, methodologies and origin of the data presented in favor of the generalized electronic electoral fraud are put to question. But far from diminishing such evidence, analyzing the counter arguments that Jimenez states in his article allows to understand how arbitrary the current Venezuelan electoral system really is.

Jimenez states, Ref [j], Section 2:

> "This article has two purposes: 1) To bring order to the ruckus caused by different statistical analyses, some of them carried out by non-experts, and 2) to examine, by a proper forensics analysis, the allegations of fraud"

Afterwards he mentions that the articles by Taylor (2006) and E. Felten, A. Rubin and A. Stubblefield (2004) conclude not having observed evidence of fraud. Then, referring to an alleged contradiction between both articles and those who asseverate having observed evidence of fraud, he asks who's right:

> "The paper by Taylor (2005) which concludes explicitly that there is no evidence of fraud. Taylor's paper is the best known reference on the subject, widely covered by media; partially because he was asked to investigate the allegations of fraud on behalf of The Carter Center. Another well-known reference is a paper by Felten et al (2004), which did not detect any statistical inconsistency that would indicate obvious fraud in the election. However, three papers in this issue of Statistical Science (Delfino and Salas, 2011; Prado and Sansó, 2011; Pericchi and Torres, 2011) support the claim of fraud. Who is right?"

In this statement, Jimenez omits mentioning the following information regarding J. Taylor's article (2005):

1. To investigate whether the results of the elections should comply with Benford's Law, Taylor limited his analysis to the comparison of three groups of electoral results generated by computers: the actual results of the voting machines and two simulations (done with mathematical models invented by Taylor). He concludes that given the fact that his simulations and the results of the voting machines did not comply with Benford's Law, there was no evidence of fraud.

2. That in Mebane's article (2006), the proof that both mathematical models used by Taylor to simulate the official results, were too simple to produce results that complied with Benford's Law.
3. That the results of the real elections do comply with Benford's Law.

Regarding the article of E. Felten, A. Rubin and A. Stubblefield (2004), Jimenez omits that in it the only allegation of fraud analyzed was that which received the name of "Los Topes" ("The Topes"), and that this allegation has nothing to do with the evidence of fraud in the articles by Febres-Cordero and Marquez (2006), Delfino and Salas (2011), Prado and Sanso (2011) and Pericchi and Torres (2011).

The arguments that R. Jimenez used to oppose the arguments and evidences that corroborate the thesis of electronic fraud are the following:

B.1. About the arguments that Jimenez uses to diminish importance to the fact that the results of the voting machines violate Benford's Law

These arguments were refuted in Section 5 of the present article.

C.2. About the arguments that Jimenez uses to try to discredit the two articles that promote the alternative hypothesis

Jimenez to discredit the data used in these articles (Ref. [1] and Ref. [m], affirms:

> "I am very skeptical about the use of data from other sources."
>
> "In this political atmosphere, we must assume that any unofficial information will be controversial"
>
> "I should add a comment related to the data. From the least squares regression lines shown in Figure 6 of Delfino and Salas (2011) one can estimate the total signatures in fully manual or computerized centers (excluding the mixed ones) on which the authors base their study. This total is 3,310,200, close to the 3,467,051 signatures submitted to the electoral umpire (Delfino and Salas, 2011). However, the total number of valid signers was 2,553,051 (The Carter Center, 2005). I leave the conclusion to the reader".

With these observations, Jimenez brings to discussion two issues of great importance: the severe contradiction that exposed the official results, and the use of non official data to analyze this case.

A contradictory official truth and the use of non official data

Delfino and Salas (2011) analyze the relations between YES votes and Signatories (of the petition for the Referendum) per voting center, which deep down is an answer to Appendix 4 of the Carter Center's Report (Ref. 2) - *Report on an Analysis of the*

Representativeness of the Second Audit Sample, and the Correlation between Petition Signers and the Yes Vote in the Aug. 15, 2004 Presidential Recall Referendum in Venezuela - (which Jimenez quotes in his article). Delfino-Salas (2011) analyze the relations YES Votes/Signatories using the same methodology used by the Carter Center in this appendix. It is important to highlight that the date used to calculate the number of signatories registered in every voting center is official data. Even the complete list of ID numbers of every signatory was published by the CNE in two papers of national circulation (See Ref. [cc]).

The origin of the doubt planted by R. Jimenez about the reliability of the data used in Delfino and Salas (2011) relies in the following argument:

According to the Venezuelan Constitution, to activate RRP-2004 the opposition had to hand the CNE a petition with the signatures of at least 2.340.305 voters (20% of registered voters). In an event organized and supervised by the CNE known was "*El Reafirmazo"* ("To Reaffirm"), the petitioners of the Referendum signed and put their finger prints in the numbered forms handed by the CNE. Such forms were of security paper and used in front of witnesses of the government, the opposition and the CNE. After "*El Reafirmazo"*, the opposition handed the CNE the petition for the Referendum with 3.476.051 signatures.

Disregarding the controls imposed in the recollection of the signatures, the CNE rejected more than a million and a half of them (alluding problems in the forms, similar calligraphy, similar fingerprints, and others) and denied the petition.

This attitude unchained a series of violent protests known as *"La Guarimba*", which were repressed and where many Venezuelans died. This combative attitude of the Venezuelan people forced the government to modify its decision, and so signatories were divided into three groups: signatories with accepted signatures, signatories with rejected signatures, and signatories with the possibility of repairing their signature. After a new process called *"Los Reparos"* ("The Repairs"), CNE recognized having received 2.553.051 valid signatures.

After this arbitrary process in which the CNE had rejected even signatures of entire municipalities, the official truth that Jimenez alludes to in his article was conformed: there were only 2.553.051 signatories.

In Appendix 4 of his report (Ref. 2), the Carter Center partially rejects this official truth. In the Section Methodology of Appendix 4 its stated:

> "This study is based on the assumption that all signatures were made in good faith by real voters; for this reason rejected signatures because of acta problems, similar handwriting, fingerprint, and other problems were included in the universe of signatures"

The Carter Center illustrates how, to compare the relations between YES votes and Signatories in computerized voting centers, it included valid signatures, rejected signatures and signatures sent to "*Reparo*" ("Repair"). See Fig. 13.

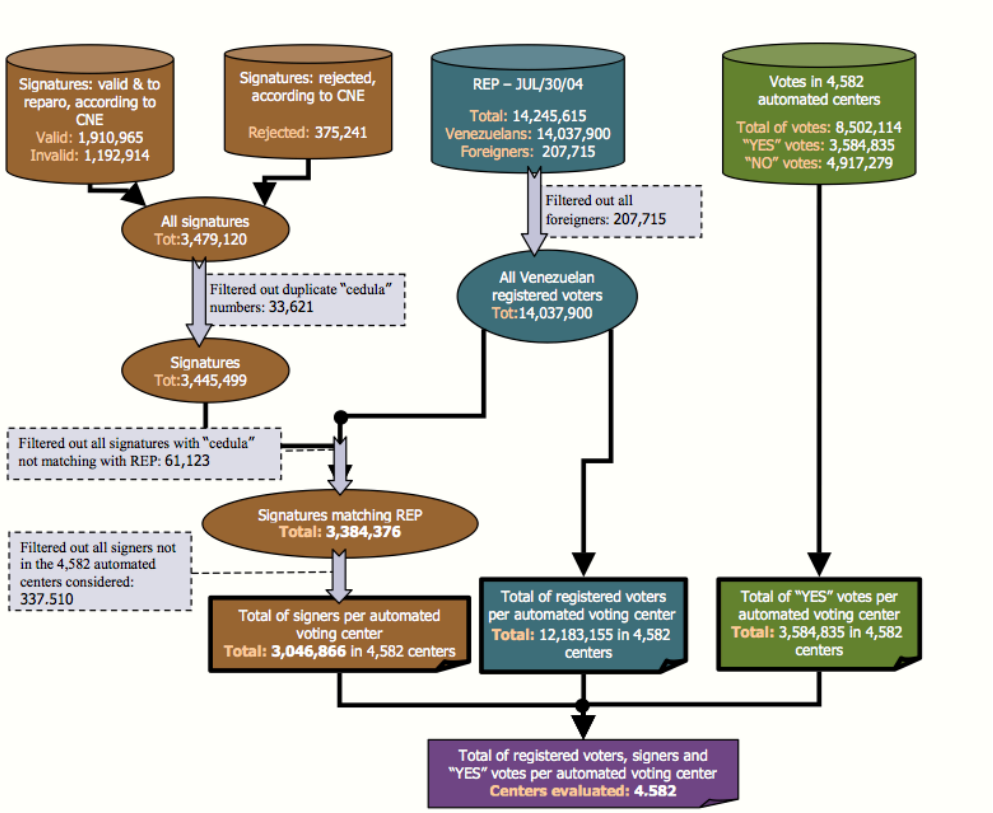

Diagram 1: Data processing flow chart

Fig. 13: Image from Carter Center Inform (Ref. 2. Appendix 4) of data processing flow chart used for Carter Center to analyze the correlation between petition signers and the YES Votes (by voting center). Observe that valid, to *reparo* and rejected signatures were included.

In Appendix 4, Section Results, the Carter Center refers exclusively to computerized voting centers:

> "There is a very high correlation between the number of signatories and the number of YES votes per voting center, of the universe of voting machines: a coefficient of 0,988"

The official results reported by the voting machines correlate (r = 0.988) with the non official truth; with the list of petitioners that the opposition handed the CNE (and that the CNE rejected). This high correlation diminishes when the signatures that the CNE rejected are not included, amongst other reasons because the CNE rejected 100% of the signatures collected in various municipalities. See Fig. 11.

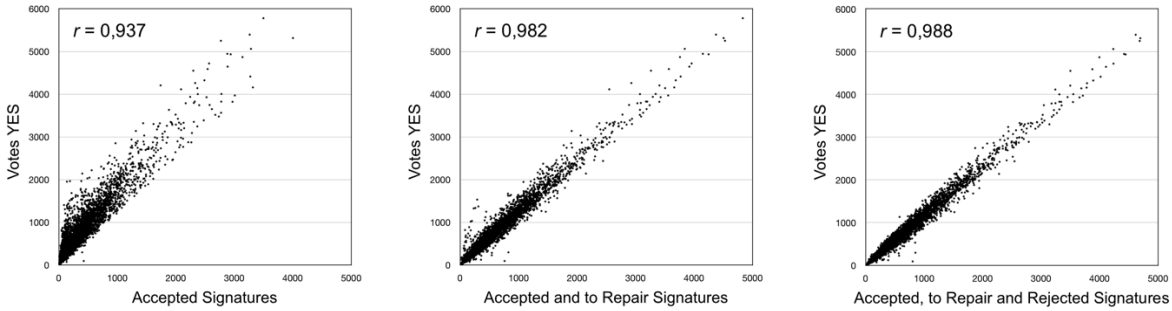

Fig. 14. This graphics shows (by voting center) how the YES Votes reported by voting machine (official truth) are related whit the signatures: a) left graphic, only with the accepted signatures (official truth), b) center graphic, only with the accepted and to repair signatures, and c) right graphic, with the accepted, to repair and rejected signatures (official non truth). Observes that the YES Votes reported by voting machine, official truth, are much better related with official non truth, accepted, to repair and rejected signatures, than with the official truth, accepted signatures.

After observing the contradictions between the official truth, the issue of use only official data and reject the use of non official data in the venezuelan case becomes relevant.

In Delfino and Salas (2011), Section 9, to calculate a probability that demonstrates that the suspension of the audit scheduled immediately after the closure of voting centers was a deliberate act and not an act of chance (like the Carter Center sustained), the list of the 27 voting centers where the CNE allowed the opposition to be present in the audits was used. This list was done by members of the Coordinadora Democratica (the civil group accepted by the OEA and the Carter Center to represent the Opposition in RRP-2004), by members of the NGO Ciudadania Activa, endorsed by the Vicepresident of the CNE at the time Dr. Ezequiel Zamora, and handed to the Carter Center and the OEA. In Appendix C, graphs were constructed with the results of a poll by NGO SUMATE, designed and supervised by Penn, Shoen & Berland Associates (this data was not used to calculate the probabilities).

In the article Prado and Sanso used data of two polls: the poll by SUMATE and another one by the political party Primero Justicia. Both polls were designed and supervised by Penn, Schoen & Berland Associates.

In his article, I. Martin also concurs with the thesis of generalized electronic fraud. He used the *logs* of transmission registered by the communications between the voting machines and central server instants before printing the acts of results. This data was handed by personnel of CANTV, the telephone company that provided the transmission channels between voting machines and this central server.

To analyze the results of an election organized by an electoral institution visibly subordinated to the government, should we limit ourselves to use official data even if it is so manifestly contradictory?

References:

[1] *Analysis of the 2004 Venezuela Referendum: The Official Results Versus the Petition Signatures* G. Delfino, G. Salas. *Statistical Science* 2011, Vol. 26, No. 4, 479–501.

[2] THE CARTER CENTER (February 25 of 2005). *Observing the Venezuela presidential recall referendum: Comprehensive report*.

[3] *The 2004 Venezuelan Presidential Recall Referendum: Discrepancies Between Two Exit Polls and Official Results*. R. Prado, B. Sansó. Statistical Science 2011, Vol. 26, No. 4, 517–527.

[4] *In Search of the Black Swan: Analysis of the Statistical Evidence of Electoral Fraud in Venezuela*. R. Hausmann, R. Rigobón. Statistical Science 2011, Vol. 26, No. 4, 543–563.

[5] Comprobación de la Hipótesis de Delfino y Salas. R. Medina.

[6] *La Ley de Newcomb-Benford y sus aplicaciones al Referendum Revocatorio en Venezuela. Reporte Tecnico no-denitivo 2a. version: Octubre 01, 2004*. L. Pericchi, D. Torres <http://esdata.info/pdf/pericchi-torres.pdf>

[7] *Evidencia de la Manipulación Artificial de los Resultados al Aplicar la Ley de Benford al Referéndum Venezolano de Agosto de 2004*. I. Mikoss <http://esdata.info/pdf/mikoss.pdf>

[8] *Quick Anomaly Detection by the Newcomb–Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela*. L. Pericchi, D. Torres. Statistical Science 2011, Vol. 26, No. 4, 502–516.

[9] *Las Distribuciones no Uniformes de los Segundos Dígitos Significativos de los Resultados Electorales*. G. Salas, G. Delfino, F. Vecchionacce. <http://esdata.info/pdf/s-d-v.pdf>

[10] *A Statistical Approach to Assess Referendum Results: the Venezuelan Recall Referendum 2004*. M. Febres Cordero, B. Márquez. International Statistic Review 2006, Volume 74, Number 3.

[11] 2004 Venezuelan Presidential Recall Referendum (2004 PRR): A Statistical Analysis from the Point of View of Electronic Voting Data Transmissions. I. Martín. Statistical Science 2011, Vol. 26, No. 4, 528–542.

[12] Election Forensics: Vote Counts and Benford's Law. W. Mebane. July, 2006. <http://www-personal.umich.edu/~wmebane/pm06.ps>

[13] *Forensic Analysis of the Venezuelan Recall Referendum*. R. Jimenez. Statistical Science 2011, Vol. 26, No. 4, 564–583.

[14] Fraude a la Democracia, Informe Final