Submitted to the Annals of Applied Statistics

BENFORD'S LAW ANOMALIES IN THE 2009 IRANIAN PRESIDENTIAL ELECTION

By Boudewijn F. Roukema

Toruń Centre for Astronomy, Nicolaus Copernicus University

The results of the 2009 Iranian presidential election presented by the Iranian Ministry of the Interior (MOI) are analysed based on Benford's Law and an empirical variant of Benford's Law. The null hypothesis that the vote count distributions satisfy these distributions is rejected at a significance of $p \leq 0.007$, based on the presence of 41 vote counts for candidate K that start with the digit 7, compared to an expected 21.2–22 occurrences expected for the null hypothesis. **Three of the six most populous voting areas have vote totals for K that start with 7. All three of these have greater proportions of votes for A than the other three voting areas. Interpreting this as an overestimate of the true vote assumed to be 50% to match other data, while retaining constant total vote numbers and increasing votes for the other three candidates in proportion to the average voting percentages, would imply that the difference between A's and M's vote totals would drop by about one million votes.**

1. Introduction. The results of the 12 June 2009 presidential election held in the Islamic Republic of Iran are of high political importance in Iran. International interest in these results is also considerable. On 14 June 2009, the Ministry of the Interior (MOI) published a table of the numbers of votes received by the four candidates for 366 voting areas (MOI Iran 2009a). In order to avoid focussing on personalities, the four candidates will be referred to here as A, R, K, and M, following the order given in the table. These letters correspond to the conventional Roman alphabet transliteration of the four candidates' names by which they are frequently referred to. The total votes for these four candidates from the MOI table give A as the winner with 24,515,209 votes, against R with 659,281 votes, K with 328,979 votes, and M with 13,225,330 votes.

The total numbers of votes in the 366 voting areas in the MOI's data vary from about 10^4 to 10^6, i.e. two orders of magnitude. This suggests that Benford's Law (Newcomb 1881; Benford 1938) may be applicable to test *the null hypothesis that the first digit in the candidates' absolute numbers of votes are consistent with random selection from a uniform, base 10 log-*

AMS 2000 subject classifications: Primary 62P25

arithmic distribution modulo 1. **Benford's Law for the distribution of the *second* digit has been independently applied to this same data set (Mebane 2009). Since use of the first digit distribution may be problematic (Mebane 2009), the standard form of Benford's Law is supplemented by an empirical form of Benford's Law generated directly from the data. The details of the method are** described in Section 2. A plain text form of the MOI data (Roukema 2009a) and a plain text OCTAVE script (Roukema 2009b) for reproducing these results are provided along with this article. Results are presented in Section 3. Discussion, **including an anomaly that follows from the basic result and some statistical characteristics of the vote distributions, are** given in Section 4. Conclusions are presented in Section 5.

2. Method. Benford's Law (Newcomb 1881; Benford 1938) for the relative frequency of the occurrence of the first digit i in decimal representations of real numbers

$$f(i) = \log_{10}\left(1 + \frac{1}{i}\right) \tag{1}$$

should be valid for real world samples that can be expected to be logarithmically uniform over several orders of magnitude.[1] The degree to which this assumption is accurate depends on the degree to which

$$\{\log_{10} v_j - \lfloor \log_{10} v_j \rfloor\}, \tag{2}$$

i.e. the folding of a sample $\{v_j\}$ to a single decade, is uniform, where $\lfloor x \rfloor$ is the greatest integer $\leq x$. This illustrates why data sets do not necessarily need to span many orders of magnitude in order to approximately satisfy Benford's Law. The most striking characteristic follows from Eq. (1): the first digit is 1 with a frequency of $\log_{10} 2 \approx 30\%$, i.e. much more frequently than any other digit.

Mebane (Mebane 2009) has pointed out that use of Benford's Law for distributions of the first digit is problematic. Hence, in order to apply an equivalent probability distribution to Benford's Law, an empirical version of the law can be constructed as follows. Let the total numbers of votes in voting area j be v_j and the global fraction of votes received by candidate $X =$A, R, K, and M be $\alpha_X = \left(\sum v_{Xj}\right) / \left(\sum v_j\right)$, where v_{Xj} is the vote count for candidate X in the j-th voting area. If voters in different areas vote fractionally in exactly identical ways independently of geography, then the distributions of first digits should follow the total vote

[1] Powers of 10.

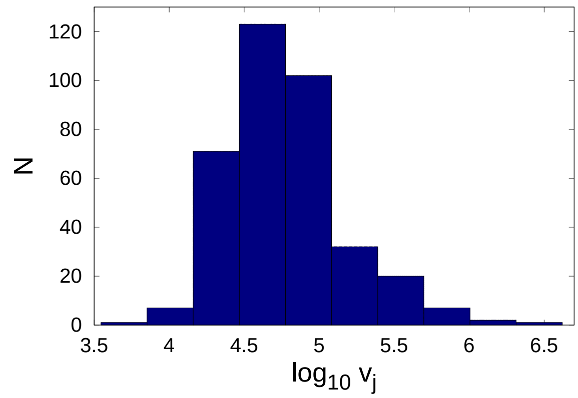

FIG 1. *Histogram showing distribution N of the total vote counts in equal bins of* $\log_{10} v_j$.

counts, apart from a constant logarithmic shift. That is, let us define $f_X(i)$ as the relative frequency of the digit i in the set of digits

$$\{\lfloor 10^{\log_{10}(\alpha_X v_j) - \lfloor \log_{10}(\alpha_X v_j) \rfloor} \rfloor\}. \tag{3}$$

In reality, geographic variation in voting preferences, and small town versus large town demographic variations in preferences make it unlikely that exact proportionality is valid, i.e. the actual vote counts v_{Xj} for candidate X are only approximated by the $\{\alpha_X v_{Xj}\}$. Nevertheless, $f - f_X$ should give an approximation to the inaccuracy introduced by the logarithmic uniformity assumption required in Eq. (1).

Another caveat is that if the vote counts v_{Xj} for a candidate who dominates the total vote count ($\alpha_X \approx 1$) themselves are anomalous, then $v_{Xj} \approx v_j$, so that the null hypothesis will contain nearly the same (anomalous) information as the sample. In this case, the empirical Benford's Law will be a weak test for detecting anomalies.

3. Results. Figure 1 shows that as expected, the distribution of total vote counts mostly covers only about two orders of magnitude, while Fig. 2 shows that the folded distribution [Eq. (2)] is more uniform.

Figure 3 shows that the concatenation of all four candidates' vote counts is much better fit by Benford's Law for a uniform logarithmic distribution rather than the empirical Benford's Law. This is reasonable since the mean voting rates for the different candidates' vary widely, so that the concatenated data v_{Xj} cover more orders of magnitude than the total vote data set v_j. For the same reason, the fact that the concatenated list of all the

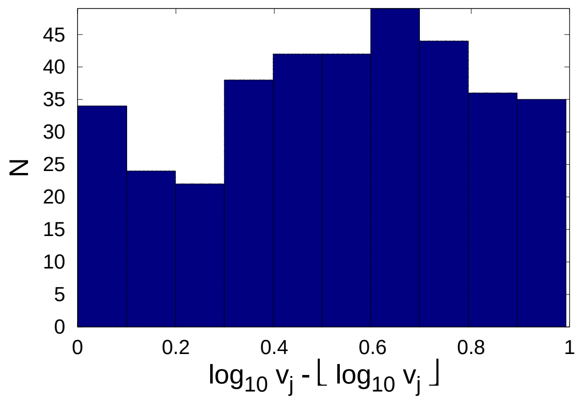

FIG 2. *Histogram showing distribution of the total vote counts folded into a single decade [Eq. (2)].*

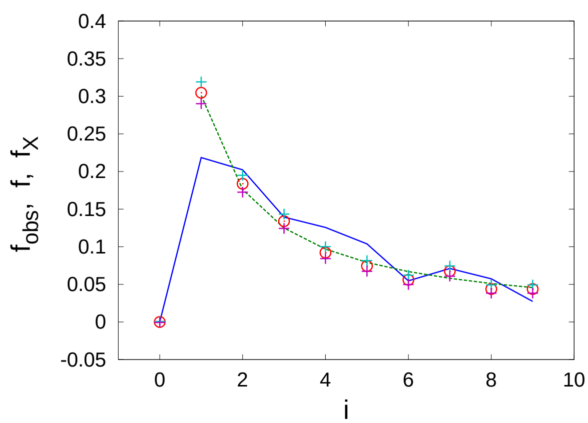

FIG 3. *Sample frequency distribution* $f_{\rm obs}$ *of the first digit of all individual vote counts* v_{Xj}*,* $X \in \{\mathrm{A, R, K, M}\}$*, shown as (red) circles, with Poisson errors indicated by plus signs. The smooth, dotted (green) line shows the expected frequencies from Benford's Law [*f *in Eq. (1)]; the jagged continuous (blue) line shows the expected frequencies using the empirical Benford's Law [Eq. (3)] with* $\alpha_X \equiv 1$*. The frequency of the digit 0 is zero since this plot concerns the first digit.*

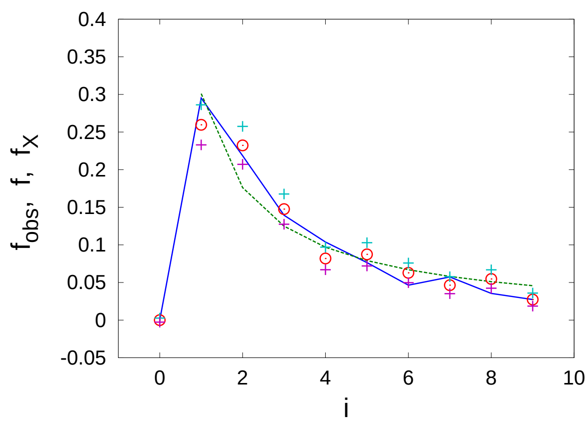

FIG 4. *As for Fig. 3, for candidate A vote counts only.*

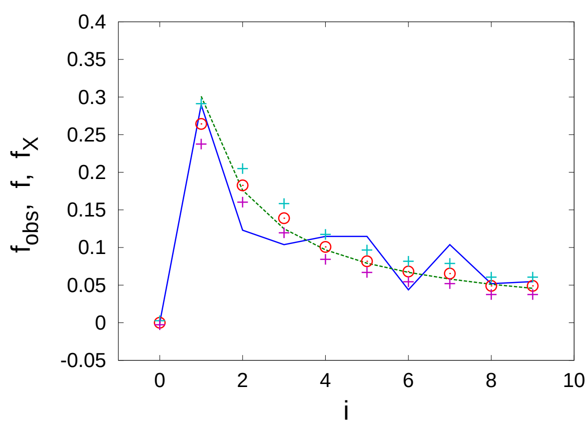

FIG 5. *As for Fig. 4, for candidate R.*

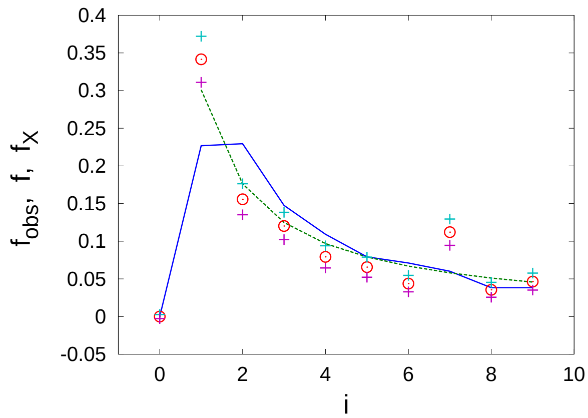

FIG 6. *As for Fig. 4, for candidate K. The excess number of 7's is about 3 standard deviations in excess of the expected values for both the idealised and empirical Benford's Laws.*

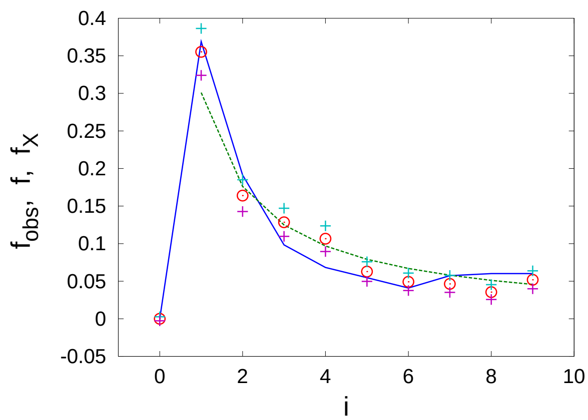

FIG 7. *As for Fig. 4, for candidate M.*

candidates' votes fits Benford's Law well does not imply that the votes for a single candidate should provide an equally satisfactory fit.

Figure 4 shows that first digits of the vote counts for candidate A have an excess of 2's and a lack of 1's relative to Benford's Law f (smooth dotted line) by roughly 2 to 1.5 standard deviations respectively. On the other hand, the frequencies of 1's and 2's match the empirical Benford's Law f_{A} much better. However, since A has about 60% of the total vote, he dominates it. Hence, as mentioned above, the similarity between A's first digit frequencies and the empirical distribution is mainly due to a built-in constraint and does not provide much useful information.

For minority candidates, the empirical Benford's Law f_X appears to provide a good complement to the uniform Benford's Law f. Figures 5 and 7 show candidate vote counts for R and M that are approximately bounded by the two probability density functions.

However, it is clear from Fig. 6 that the number of 7's in K's first digit distribution is about 3 standard deviations too high for *both* versions of the null hypothesis, f and f_{K}. The significance of this excess can be calculated more precisely using the cumulative Poisson distribution $P_{\mathrm{Poiss}}(x, \lambda)$ of mean λ. K has 41 vote counts that start with the digit 7. For a sample size of 366, the uniform and empirical versions of Benford's Law predict 21.2 and 22.0 values starting with 7 respectively. This gives $p = 1 - P_{\mathrm{Poiss}}(41, 21.2) = 4 \times 10^{-5}$ and $p = 1 - P_{\mathrm{Poiss}}(41, 22.0) = 9.6 \times 10^{-5}$ respectively. Converting these to two-sided probabilities, since we have not hypothesised any particular form of anomalies, gives $p = 1 - P \leq 1.9 \times 10^{-4}$.

This is a strong rejection of the null hypothesis in either form. However, let us suppose that this is the only anomalous frequency for all the first digits of all four candidates, and to be conservative, let us suppose that these constitute 36 independent samples of a statistical test. In that case, we have $p = 1 - P \leq 0.0069$ for the full set of tests, conservatively using just one clearly divergent point.

4. Discussion.

4.1. **Excess 7's for K**.

The rejection of the null hypothesis at $1 - P \leq 0.0069$ is estimated using just 41 vote counts starting with the digit 7 for candidate K, in excess to an expected 20–21.2 vote counts starting with 7. Could this just be a copying error by employees under pressure in a stressful situation? Various sources of unintentional errors are possible. The present analysis only concerns the data as published by the MOI.

This is unlikely to be a transliteration error: the different files appear to contain the same substantial content. The number of entries that start

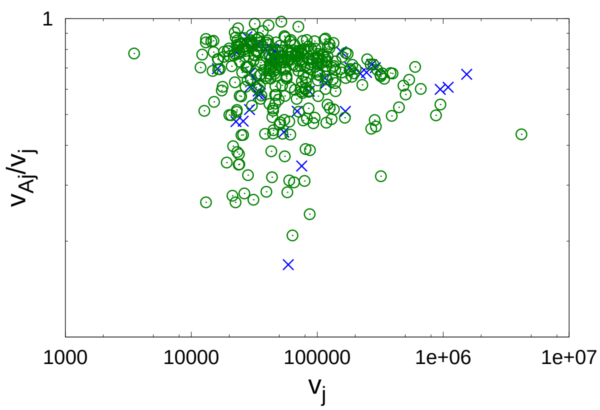

FIG 8. *A's proportions of vote counts v_{Aj}/v_j againt total votes v_j. Voting areas where K's vote count has the first digit 7 are shown by (blue) "x" signs, other areas are shown as (green) circles.*

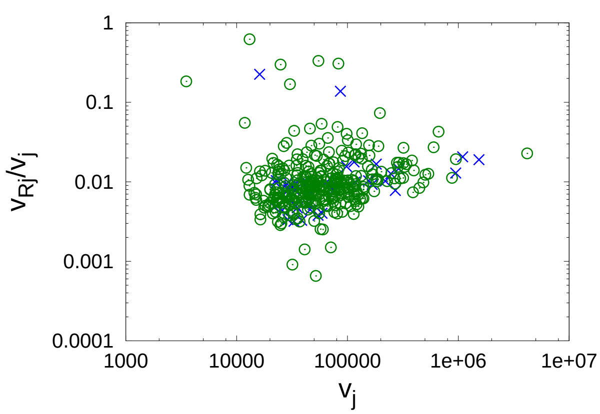

FIG 9. *R's proportions of vote counts, as for Fig. 8.*

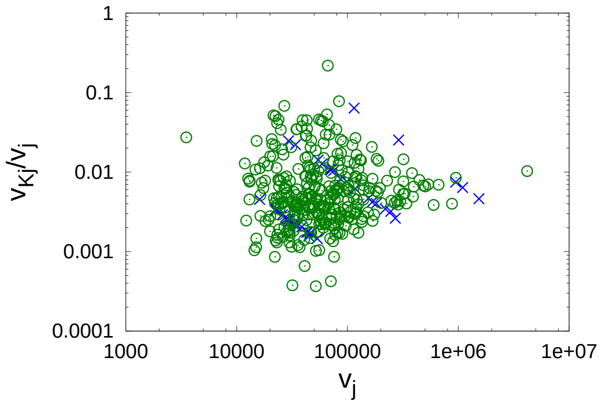

FIG 10. *K's proportions of vote counts, as for Fig. 8. The selection by the first digit 7 is clear.*

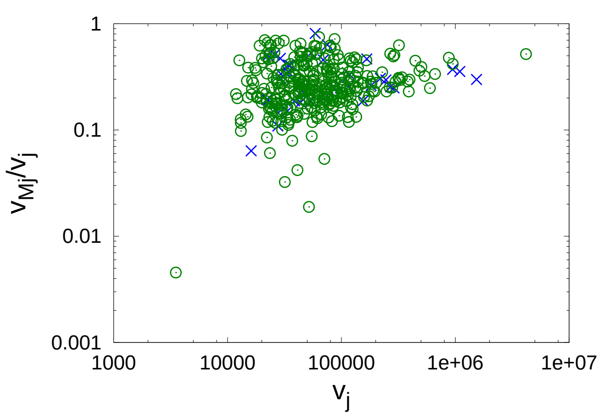

FIG 11. *M's proportions of vote counts, as for Fig. 8.*

TABLE 1
Votes for K and proportion of votes for A for the six voting areas with the greatest numbers of total votes.

voting area	v_j	v_{Kj}	v_{Aj}/v_j
Tabriz	876919	3513	0.497
Shiraz	947168	7078	0.600
Karaj	950243	8057	0.537
Isfahan	1095399	7002	0.609
Mashhad	1536106	7098	0.669
Tehran	4179188	43073	0.433

with 7 under candidate K in the Persian-Arabic numerals PDF file is 41 (MOI Iran 2009c).

The fact that "just a few dozen 7's" may intuitively seem insignificant could itself be a reason for the anomaly. In the case of artificial modification of the data, "just a few dozen 7's" may have seemed sufficiently "random" not to be detectable.

One possible method to test whether this is just an odd fluke would be to check the validity of the vote counts for candidate K in the voting areas where the official number of votes for K starts with the digit 7.

However, if the excess 7's for K indicate interference in the data, then other signs of interference could be expected. A reasonable hypothesis would be that vote counts for one of the major candidates were decreased or increased. Figures 8, 9, 10 and 11 show proportions of votes that each candidate received as a function of total votes, where those voting areas selected by having 7 as the first digit in K's vote count.

It is clear in Fig. 8 that for candidate A, among the six voting areas with the greatest numbers of total votes, the three of these that voted for A in the highest proportions are all selected by the K first digit 7. **Data for these six voting areas are listed in Table 1. Figure 11 shows correspondingly that the three of the six most populous voting areas who voted least for M are also those selected by the K first digit 7. Since the vote fractions for K and R are only about 1% each, the high proportions of votes for A necessarily imply low proportions of votes for M.**

If we suppose that (i) these three voting areas (Shiraz, Isfahan and Mashhad for 7078, 7002 and 7098 votes for K respectively) should have proportions of about 50% for A in agreement with

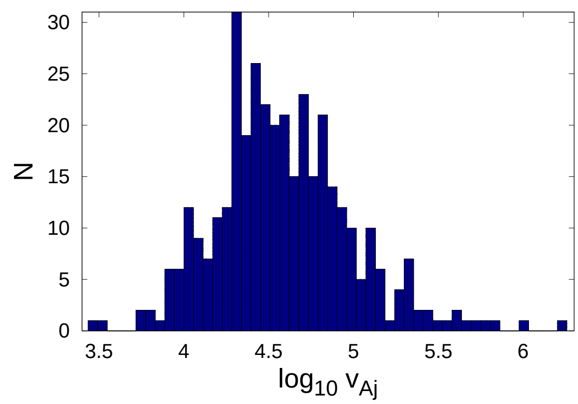

FIG 12. *Distribution of the logarithmic vote counts for A, shown as numbers per logarithmic bin in* $\log_{10} v_{Aj}$.

Tabriz and Karaj, which follow an approximately linear upper boundary to A's proportions of votes in the log-log plot in Fig. 8, and (ii) the total number of votes should remain constant, then from Table 1 this would imply that the correct number of votes for A would be about 473,000 less than in the MOI table. To keep the total number of votes constant, M's, K's and R's votes would also have to be corrected. If these are corrected in proportion to the three candidates' overall vote percentages, then the difference between A's and M's total vote counts would be reduced by about one million votes.

4.2. **Other statistics.** Are there other statistics that could help to distinguish the null hypothesis from the alternative hypothesis? To motivate further analyses, Figs 12, 13, 14, and 15 show the logarithmic distributions of the four candidates' votes. Fig. 12 does appear to have high, locally significant spikes at $\log_{10} v_{Aj} \approx 4.3, 5.3$, i.e. $v_{Aj} \approx 20,000$ and $v_{Aj} \approx 200,000$ respectively.

On the other hand, Fig. 13 appears to be very far from a lognormal distribution: there appears to be a significant dip at $3 < \log_{10} v_{Rj} < 3.3$, i.e. in the range 1000 to 2000. This was not detectable in the Benford's Law tests since the above-100 and above-1000 decades were combined.

Can we quantify the characteristics of these four plots in a way that is independent of the relative popularity (assuming that the data are correct) of the four candidates? For 366 values, the standard error of skewness $\sigma_{\langle\gamma_1\rangle} \approx$

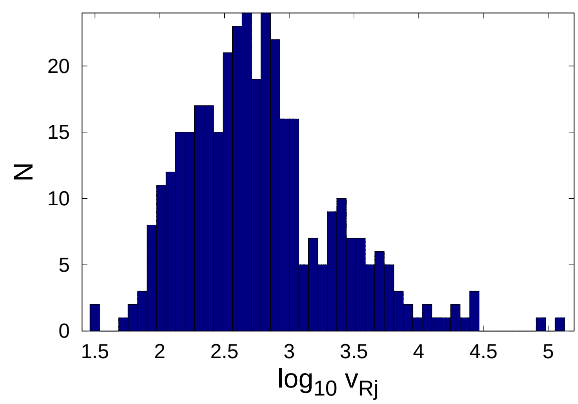

FIG 13. *Distribution of the logarithmic vote counts for R, shown as numbers per logarithmic bin in* $\log_{10} v_{Rj}$.

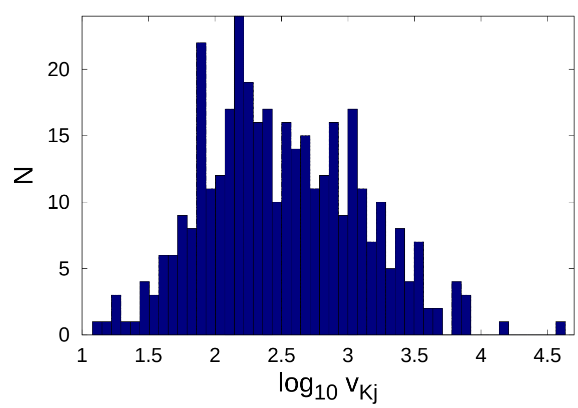

FIG 14. *Distribution of the logarithmic vote counts for K, shown as numbers per logarithmic bin in* $\log_{10} v_{Kj}$.

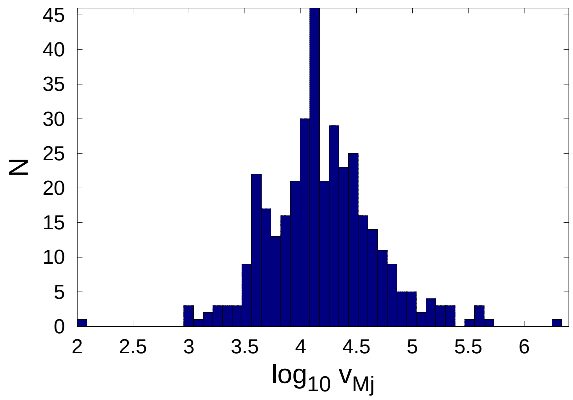

FIG 15. *Distribution of the logarithmic vote counts for M, shown as numbers per logarithmic bin in* $\log_{10} v_{Mj}$.

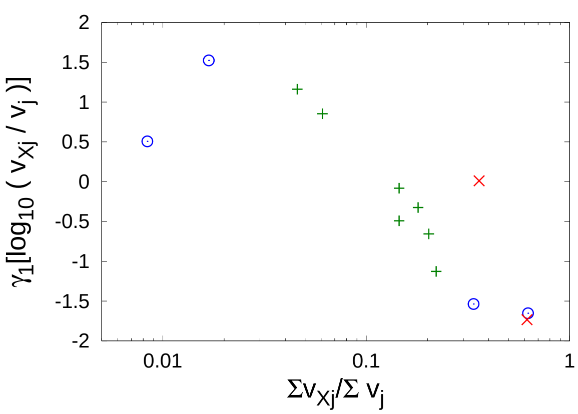

FIG 16. *Logarithmic skewness of proportional counts, as in Table 2, comparing the 2005 first ("+") and second ("x") round and 2009 first ("o") round candidates' results, shown against the overall proportions of votes.*

Table 2

Skewness of vote counts $\gamma_1(\log_{10} v_{X\,j})$ **and its normalisation by the standard error in the skewness** $\sigma_{\langle\gamma_1\rangle} \equiv \sqrt{\mathrm{Var}(\gamma_1)}$**, skewness of proportional vote counts,** *and width* $\sigma(\log_{10} v_{X\,j})$ *of the four candidates' vote counts.*

candidate	A	R	K	M
$\gamma_1(\log_{10} v_{X\,j})$	0.59	0.74	0.32	−0.09
$\gamma_1(\log_{10} v_{X\,j})/\sigma_{\langle\gamma_1\rangle}$	4.6	5.8	2.5	−0.7
$\gamma_1[\log_{10}(v_{X\,j}/v_j)]$	−1.65	1.52	0.51	−1.53
$\gamma_1[\log_{10}(v_{X\,j}/v_j)]/\sigma_{\langle\gamma_1\rangle}$	−12.9	11.9	4.0	−12.0
$\sigma(\log_{10} v_{X\,j})$	0.40	0.56	0.59	0.51

$\sqrt{6/366} \approx 0.128$. Table 2 shows that candidate R's distribution is skewed by about $5.8\sigma_{\langle\gamma_1\rangle}$. A's is skewed by $4.6\sigma_{\langle\gamma_1\rangle}$ and K's by $2.5\sigma_{\langle\gamma_1\rangle}$. For a log-normal null hypothesis, these would correspond to rejection probabilities of $p \sim 7 \times 10^{-9}, 4 \times 10^{-6}$, and 1.2×10^{-2} respectively. In contrast, M's distribution has no significant skew at all.

One possible explanation of why at least candidates with high vote counts should have skewed log-normal distributions is that their votes are constrained by the total numbers of votes. A tail to lower values is allowed, but not a tail to higher values. Hence, a strong negative skew is reasonable. However, $\gamma_1(\log_{10} v_{A\,j})$ **is positively skewed, not negatively skewed. The skewness of the proportional counts** $\gamma_1[\log_{10}(v_{X\,j}/v_j)]$ **shown in Table 2 is easier to understand. In this case A's distribution is negatively skewed, consistently with a long tail to lower values.**

Positive skewness for candidates with low vote numbers could be expected since candidates' votes cannot fall below 1; fractional votes are not possible in a standard voting system and vote counts of zero would have to be excluded for evaluating logarithmic skewness. R's and K's vote proportions do not go this low, but are nevertheless positively skewed. However, Fig. 16 shows a comparison with candidates' vote distributions from the first and second rounds of the 2005 Iranian presidential election.[2] Strong negative skews occur for candidates with high proportions of the vote as expected. In addition, the low popularity candidates of the 2005 first round election suggest a general monotonic relation. In the absence of a detailed model for how the skewness should behave for realistic voting populations, it is not obvious whether or not

[2]Kindly provided by Walter Mebane (Mebane 2009).

K's and R's distributions are exceptional compared to the low popularity candidates of the 2005 first round election.

In any case, more demographic information would be needed if these distribution **characteristics and** shapes were to be used to test the null hypothesis that the data have not been artificiallly interfered **with.**

5. Conclusion. The vote counts per voting area published on 2009-06-14 by the Ministry of the Interior of the Islamic Republic of Iran for the 2009 presidential election show a highly significant excess of the first digit 7 for candidate K, compared to the expectations either from a uniform Benford's Law or from an empirically derived equivalent of Benford's Law. Given that the test was applied for all four candidates, for all nine possible first digits, the null hypothesis that the first digit in the candidates' absolute numbers of votes are consistent with random selection from a uniform, base 10 logarithmic distribution modulo 1 is rejected at a significance of $p \leq 0.0069$, i.e. $1 - p \geq 99.3\%$.

Of the six voting areas with the greatest total numbers of voters, three of these (Shiraz, Isfahan, Mashhad) satisfy this criterion, i.e. they have vote totals for K that start with 7 (7078, 7002, 7098 votes respectively). All three of these have greater proportions of votes for A than the other three voting areas. If this is interpreted as a misestimate of the true vote and the true voting proportions for A are set to 50%, while retaining constant total vote numbers and increasing votes for the other three candidates in proportion to the average voting percentages, then the difference between A's and M's vote numbers would drop by about one million votes.

The highly significant excess of 7's for K **could** be checked by examining the credibility of the total vote numbers (and likely voting patterns) for **Shiraz, Isfahan and Mashhad, as well as for the other voting areas selected this way.** The voting areas' names are listed in the table published by the MOI (MOI Iran 2009a; MOI Iran 2009b; MOI Iran 2009c). A possible clue for further investigation **could be the positive skewness of the low popularity candidates of the 2009 first round election.** Any demographic models of Iranian voting patterns will need to either reproduce these statistical characteristics **for both the 2005 and 2009 elections**, or else make hypotheses regarding systematic anomalies in the data.

While it does seem that checking both the standard form of Benford's Law and an empirical variant based on the same idea has led to the detection of anomalies in an electoral poll, the reverse would clearly not be true. Benford's Law may detect anomalies

in a data set but cannot guarantee the absence of anomalies, because many randomising effects can combine to hide artefacts. The anomalies detected in the analysis presented above, suggesting an error of about one million votes, may not constitute the full set of anomalies. An alternative interesting in-depth analysis is that of Mebane (Mebane 2009).

Acknowledgements. Thank you to the pseudo-anonymous Wikipedia editor "128.100.5.143" who alerted me to the MOI publication of the data set **and to Walter R. Mebane Jr for providing a copy of the 2005 data set. Thank you to numerous people who provided useful comments and will remain anonymous unless they request otherwise.** This work has used the GNU OCTAVE command-line, high-level numerical computation software
(`http://www.gnu.org/software/octave`).

SUPPLEMENTARY MATERIAL

Supplement A: Plain text files containing the MOI data
(http://arXiv.org/archive/stat). The data from (MOI Iran 2009a) used in this analysis are listed in the two plain text files TOTAL and CANDS, which will be part of the source version of this article at ArXiv.org.

Supplement B: Plain text octave script
(http://arXiv.org/archive/stat). This plain text file BENFORD.M is an OCTAVE script for carrying out the analysis in this paper, using the input files TOTAL and CANDS. This file will be part of the source version of this article at ArXiv.org.

References.

[Benford 1938] BENFORD, F. (1938). *The law of anomalous numbers*, in *Proceedings of the American Philosophical Society*, **78**, 551–572.

[Mebane 2009] MEBANE, W.R., JR. (2009). *Note on the presidential election in Iran, June 2009*, `http://www.umich.edu/~wmebane/note20jun2009.pdf`

[MOI Iran 2009a] MINISTRY OF THE INTERIOR, ISLAMIC REPUBLIC OF IRAN (2009a). XLS file `http://www.moi.ir/Portal/File/ShowFile.aspx?ID=0793459f-18c3-4077-81ef-b6ead48a5065`, 2009-06-14; archived at the University of Toronto `http://www.webcitation.org/5hXHfYNbN`.[3]

[3] The main url linking to MOI Iran 2009a, 2009b, 2009c is `http://www.moi.ir/Portal/Home/ShowPage.aspx?Object=News&ID=e3dffc8f-9d5a-4a54-bbcd-74ce90361c62&LayoutID=b05ef124-0db1-4d33-b0b6-90f50139044b&CategoryID=832a711b-95fe-4505-8aa3-38f5e17309c9`; archived at `http://www.webcitation.org/5hYWAcdhW`.

[MOI Iran 2009b] MINISTRY OF THE INTERIOR, ISLAMIC REPUBLIC OF IRAN (2009b). XLS file, http://www.moi.ir/ostan.xls, 2009-06-15, containing the same numerical data as MOI Iran 2009a, but in multiple tables; archived at http://www.webcitation.org/5hYWBcT7w.

[MOI Iran 2009c] MINISTRY OF THE INTERIOR, ISLAMIC REPUBLIC OF IRAN (2009b). PDF file, http://www.moi.ir/ostan.pdf, 2009-06-15, containing the same numerical data as MOI Iran 2009b, using Persian-Arabic numerals; archived at http://www.webcitation.org/5hZbt9YhD.

[Newcomb 1881] NEWCOMB, S. (1881). *Note on the frequency of use of the different digits in natural numbers*, *American Journal of Mathematics*, **4**, 39–40.

[Roukema 2009a] ROUKEMA, B.F. (2009a). Supplement to "Benford's Law anomalies in the 2009 Iranian presidential election".

[Roukema 2009b] ROUKEMA, B.F. (2009b). Supplement to "Benford's Law anomalies in the 2009 Iranian presidential election".

TORUŃ CENTRE FOR ASTRONOMY, UMK
UL. GAGARINA 11, 87-100 TORUŃ, POLAND
E-MAIL: boud@astro.umk.pl