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. A less significant anomaly suggested by Benford's Law could be interpreted as an overestimate of candidate A's total vote count by several million votes. Possible signs of further anomalies are that the logarithmic vote count distributions of A, R, and K are positively skewed by 4.6, 5.8, and 2.5 standard errors in the skewness respectively, i.e. they are inconsistent with a log-normal distribution with $p \sim 4 \times 10^{-6}, 7 \times 10^{-9}$, and 1.2×10^{-2} respectively. M's distribution is not significantly skewed.

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 logarithmic distribution modulo 1.* The method of applying this principle is described in

AMS 2000 subject classifications: Primary 62P25

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 alternative hypothesis motivated by one of the results, is 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.

In order to apply Benford's Law more accurately, 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 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).

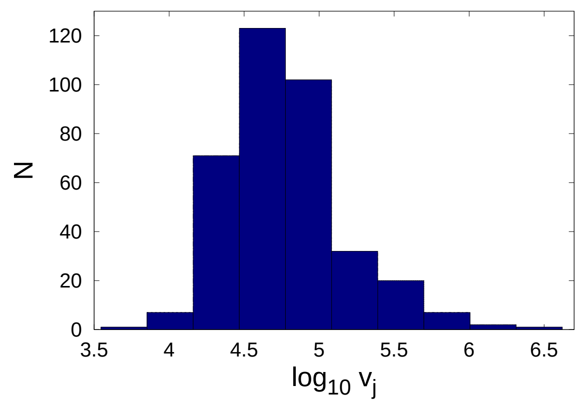

FIG 1. *Histogram showing distribution N of the total vote counts in equal bins of* $\log_{10} v_j$.

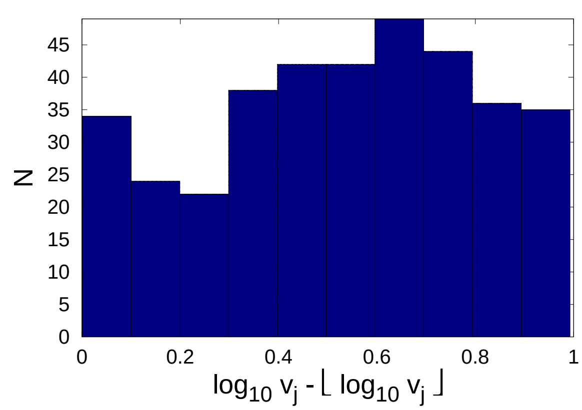

FIG 2. *Histogram showing distribution of the total vote counts folded into a single decade [Eq. (2)].*

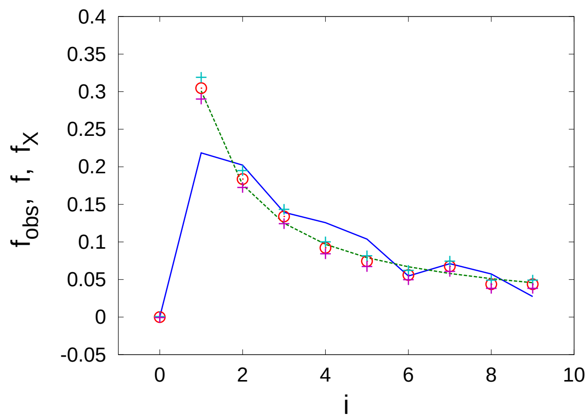

FIG 3. *Sample frequency distribution $f_{\rm obs}$ of the first digit of all individual vote counts v_{Xj}, $X \in \{{\rm A}, {\rm R}, {\rm K}, {\rm M}\}$, shown as (red) circles, with Poisson errors indicated by plus signs. The smooth, dotted (green) line shows the expected frequencies from Benford's 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.*

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 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,

[1]Powers of 10.

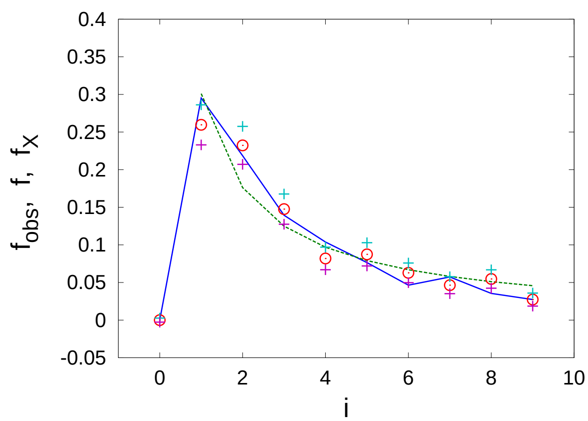

FIG 4. *As for Fig. 3, for candidate A vote counts only.*

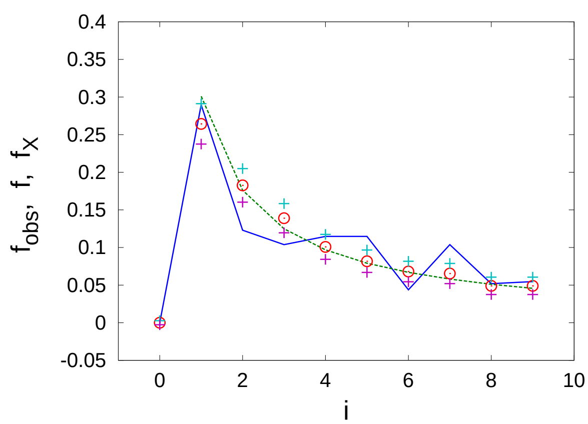

FIG 5. *As for Fig. 4, for candidate R.*

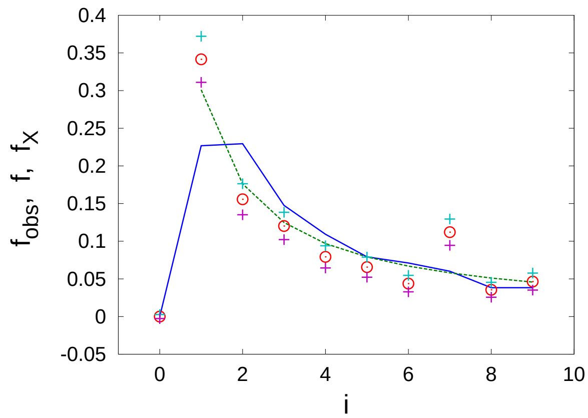

Fig 6. *As for Fig. 4, for candidate K. The excess number of 7's is about 3 standard deviations in excess of the expected values for both the idealised and empirical Benford's Laws.*

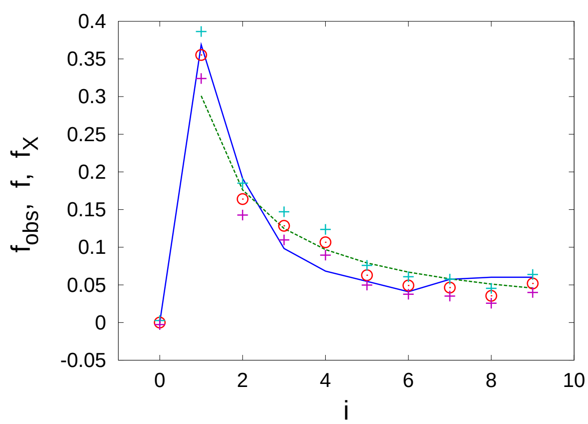

Fig 7. *As for Fig. 4, for candidate M.*

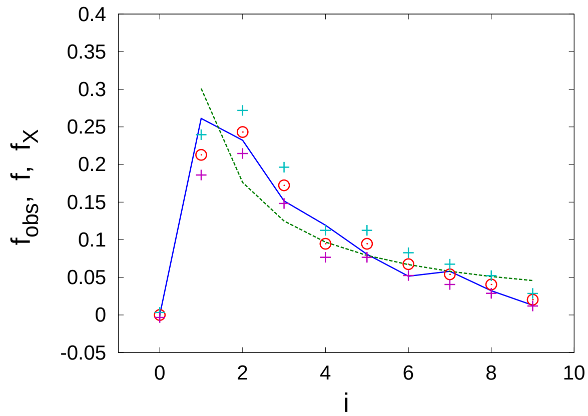

FIG 8. *As for Fig. 4, for candidate A in the range 10,000 to 99,999 inclusive, and using total votes in the corresponding range 10,000/0.6247 to 99,999/0.6247 inclusive for the estimate of* f_A.

the frequencies of 1's and 2's match the empirical Benford's Law $f_{\rm A}$ 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 $f_{\rm K}$. The significance of this excess can be calculated more precisely using the cumulative Poisson distribution $P_{\rm Poiss}(x, \lambda)$ 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 $p = 1 - P_{\rm Poiss}(41, 21.2) = 4 \times 10^{-5}$ and $p = 1 - P_{\rm Poiss}(41, 22.0) = 9.6 \times 10^{-5}$ 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,

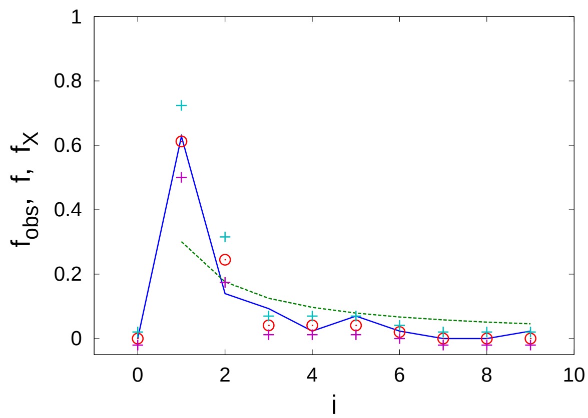

Fig 9. *As for Fig. 8, for candidate A vote counts in the range 100,000 to 999,999 inclusive, and using total votes in the range 100,000/0.6247 to 999,999/0.6247 inclusive for estimating f_A. The range of the vertical scale in this plot differs from that of the previous plots.*

we have $p = 1 - P \leq 0.0069$ for the full set of tests, conservatively using just one clearly divergent point.

4. Discussion. 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 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, let us return to the apparent excess of 2's and lack of 1's relative to Benford's Law f for candidate A, as shown in Fig. 4. If we were to consider *the alternative hypothesis that someone interfered with the data in order to* *increase A's votes,* replacement of 1's by 2's in a few dozen voting areas would be one method of achieving this without leading to numbers that are "obviously" artificial. Such a modification would be most useful for the voting areas with higher numbers of voters, i.e. 10,000-99,999 or 100,000-999,999 votes for A.

Can Benford's Law be used to test individual decades for candidate A? Again, there is the problem that A's vote dominates the total vote count, so using the empirical Benford's Law — based on the logarithmic distribution of the same limited ranges shifted higher by a factor of $1/\alpha_{\mathrm{A}} = 1/0.6247$ — again has limited value. Figures 8 and 9 show the two main decades of A's votes. If f_A is considered, then there is little significant difference between A's votes and the expected numbers of votes.

On the other hand, compared to the uniform Benford's Law f, the 10,000 and above range has about 3 standard deviations too few 1's, about 3 standard deviations too many 2's, and about 2 standard deviations too many 3's. Even more dramatically, the 100,000 and above range shows about 2.5 standard deviations too many 1's, and 2 to 3 standard deviations lack of all digits from 3 to 9. The 100,000 to 999,999 vote range for A is very significantly rejected in comparison to f.

Nevertheless, unless we know the distribution of the numbers of people who voted from an independent source, the use of the uniform Benford's Law as an approximation for the intrinsic distribution remains speculative.

On the other hand, the alternative hypothesis that a person or persons interfered with the data in order to increase A's votes could bring the vote counts in Figs 8 and 9 much closer to f. Given the lack of 1's in Fig. 8 and the excess of 1's in Fig. 9 and assuming that Benford's Law should be normalised to match the frequencies of digits 3 to 9, we could estimate that the approximately 20 "extra" 1's in Fig. 9 represent artificial increases of about 15,000 to 150,000, i.e. roughly 3 million votes would have been artificially added to A's vote counts. Corresponding explanations for the anomalous (according to this alternative hypothesis) excess 2's and 3's in Fig. 8 and excess 2's in Fig. 9 could account for another few million votes.

Are there other statistics that could help to distinguish the null hypothesis from the alternative hypothesis? To motivate further analyses, Figs 10, 11, 12, and 13 show the logarithmic distributions of the four candidates' votes. Fig. 10 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. However, to some degree we have already detected these spikes in Figs 8 and 9, so applying an additional test would risk the problem of *a posteriori* biases.

On the other hand, Fig. 11 appears to be very far from a lognormal

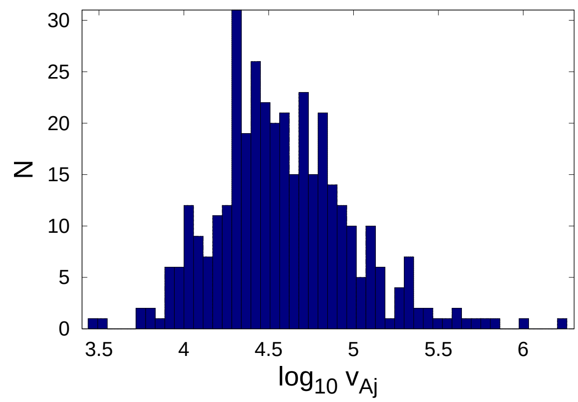

FIG 10. *Distribution of the logarithmic vote counts for A, shown as numbers per logarithmic bin in* $\log_{10} v_{Aj}$.

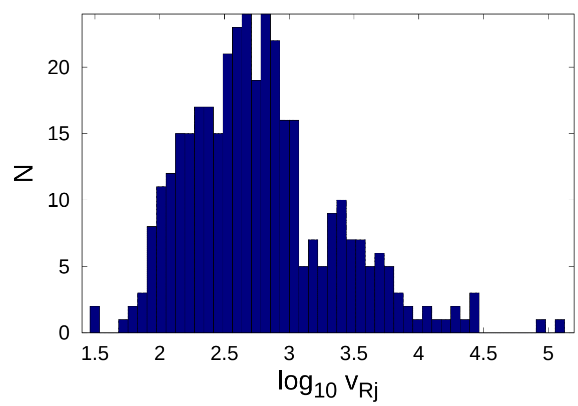

FIG 11. *Distribution of the logarithmic vote counts for R, shown as numbers per logarithmic bin in* $\log_{10} v_{Rj}$.

TABLE 1
Skewness γ_1, *standard error in the skewness* $\sigma_{\langle\gamma_1\rangle} \equiv \sqrt{\mathrm{Var}(\gamma_1)}$, *and width* $\sigma(\log_{10} v_{Xj})$ *of the four candidates' vote counts.*

candidate	A	R	K	M
γ_1	0.59	0.74	0.32	-0.09
$\sigma_{\langle\gamma_1\rangle}$	4.6	5.8	2.5	-0.7
$\sigma(\log_{10} v_{Xj})$	0.40	0.56	0.59	0.51

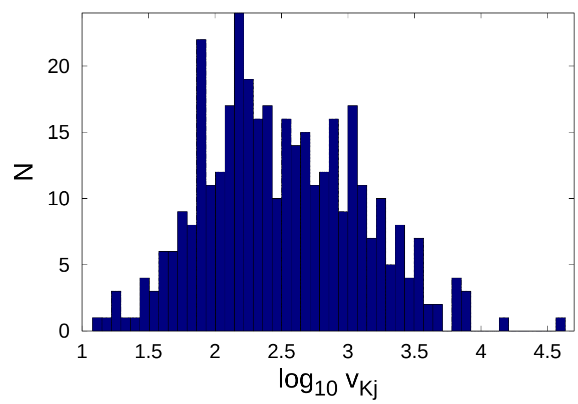

FIG 12. *Distribution of the logarithmic vote counts for K, shown as numbers per logarithmic bin in* $\log_{10} v_{Kj}$.

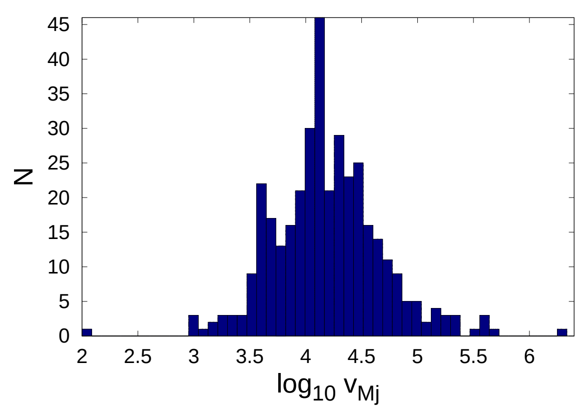

FIG 13. *Distribution of the logarithmic vote counts for M, shown as numbers per logarithmic bin in* $\log_{10} v_{Mj}$.

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 $\sigma_{\langle\gamma_1\rangle} \approx \sqrt{6/366} \approx 0.128$. Table 1 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.

It is possible that these reflect the genuine characteristics of the Iranian voting population, e.g. bimodal or trimodal demographic combinations. More demographic information would be needed if these distribution shapes were to be used to test the null hypothesis that the data have not been artificiallly interfered with.

However, the widths of the four candidates' four distributions $\sigma(\log_{10} v_{Xj})$ listed in Table 1 show another oddity. While the candidate with the highest popularity has numerically the widest freedom to have individual vote counts that range from 1 to a majority of the total number of votes, minority candidates have a more restricted range. Given various factors that spread out the different distributions, it would seem reasonable that the majority candidate has the largest log standard deviation $\sigma(\log_{10} v_{Xj})$. Table 1 shows that the reverse is the case. Candidate A has $\sigma(\log_{10} v_{Aj}) = 0.40$, while the other three candidates have $0.51 \leq \sigma(\log_{10} v_{Xj}) \leq 0.59$. This suggests another characteristic to be explained in models of this particular election and the corresponding voting population.

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. $1 - p \geq 99.3\%$.

Less significant anomalies suggested by Benford's Law can be interpreted using an alternative hypothesis in which a few dozen vote counts for candidate A in the range 10,000–19,999 had an extra digit added, shifting them to the 100,000–199,999 range, and/or a few dozen vote counts in the 20,000–29,999 and 200,000–299,999 ranges were artificially added to these ranges. Corrections for these would-be anomalies would amount to several million votes.

The highly significant excess of 7's for K and the speculative alternative hypothesis could be checked by examining the credibility of the total vote numbers (and likely voting patterns) for those particular voting areas with these numerical characteristics. 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 is that all the candidates' logarithmic vote count distributions are highly skewed, especially R's vote counts, which are positively skewed by about 5.8 standard errors, except for M, whose logarithmic vote counts are skewed (negatively) by *less* than one standard error. Any demographic models of Iranian voting patterns will need to either reproduce these statistical characteristics, or else make hypotheses regarding systematic anomalies in the data.

Acknowledgements. Thank you to the pseudo-anonymous Wikipedia editor "128.100.5.143" who alerted me to the MOI publication of the data set. 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.

[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.[2]

[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

[2]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.