Too many ties? An empirical analysis of the Venezuelan recall referendum counts

Jonathan Taylor*

November 7, 2005

Abstract

In this work, we study the question of detecting electronic voting fraud based on count data alone, specifically focusing on the 2004 Venezuelan recall election. The opposition (YES, in favour of recalling sitting president Hugo Chavez), which lost the referendum, alleged that the government had committed electronic voting fraud. In this work, we use the count data to investigate two of the opposition's claims, ultimately concluding that the claims cannot be upheld by the count data alone.

The first allegation is based on the total number of ties (defined below) for the YES/NO in each particular station. In this paper we compare four or five models for the count data. We present an FDR analysis of 6507 roughly independent test statistics that shows no systematic fraud in the form of "vote-capping" and no significant differences between the dispersion of the YES votes and the NO votes.

The second allegation involves Benford's Law, which is a model for the "significant digits" of an observation. If we look at the significant digit distribution of elections generated by a "fair election model" (the null model used by the YES side as evidence of fraud), we see that Benford's Law does not hold.

Finally, we conclude with a simple scenario of fraud that is statistically undetectable based on solely the count data. Hence, our

*Department of Statistics, Stanford University, Stanford, CA 94305; jtaylor@stat.stanford.edu

ultimate conclusion is that allegations of fraud based on count data are difficult to verify. This underscores the importance of a "paper trail" when using electronic voting machines.

1 Introduction

In the past several years, there have been several reports of the possible pitfalls of using electronic voting machines in the scientific literature as well as the news media: security failures, software bugs, lack of paper trail, etc. This technology was widely used in the 2004 presidential election in the United States. In the 2004 election, the technology used varied from state to state and county to county, which makes studying claims of electronic fraud difficult as the reliability of the count data may depend on the technology used.

Electronic voting machines were also used in 2004 Venezuelan recall referendum, decribed below. In this referendum, however, the setup was much more uniform across polling stations, providing a more uniform dataset in which to investigate the possibility of electronic voting fraud.

On August 15, 2004, Venezuela held a referendum to decide whether or not to recall the sitting president, Hugo Chavez. Since then, the losing side (YES: those favouring Chavez's recall) has alleged that the election was a fraud. The author was asked to investigate one such allegation on behalf of the Carter Center, the international observers of the Venezuelan election. The allegation is based on the observed number of ties in the YES side's votes, defined in the next paragraph.

The YES side claimed that the electronic voting machines used in the referendum "capped" the YES side's vote when they reached a certain number, the number depending on the mesa. They claim that if the number of voters at each machine in a given mesa were independent Poisson random variables with equal means then there would be approximately a 1 in 1000 chance of observing so many ties: 402 mesas of 7913 mesas had a tie for YES, which was defined as two or more machines within the mesa having the same number of YES votes[1]. For the NO side, there were 311 mesas of

[1]Actually, the model used by the YES side is multinomial, which is the model obtained from the Poisson by conditioning on the total number of voted within a mesa. This conditioning has almost no affect on the expected number of ties in the model. Further, in the final model the YES side settled on, the means were not equal but were assigned to

the 7913 that had a tie in the number of NO votes. As the NO side won, it is not unexpected that there would be more ties for YES than NO as there is more variability in a Poisson (or approximate Poisson) count with higher variance.

The YES side's claim can be summarized as follows:

> *The preponderence of ties for the YES side, compared to what would be expected under a Poisson model, indicates that "vote-capping" took place (i.e. machines were programmed to stop counting YES votes after a certain figure, depending on the mesa).*

The first part of the paper will investigate the expected number of ties under different models, which depends on the joint distribution of the counts within a mesa. For example, if all of the votes are independent and uniformly drawn from 0 to 240 (which is approximately the average number of NO votes in a station), the probability of having a tie if there are two machines is about 0.004, while if they are all degenerate and equal to 240, say, then the probability of having a tie is 1. That is, in the first scenario, if we generate integers X_1 and X_2 independently and uniformly from $[0, 240]$, then

$$\mathbb{P}\left(X_1 = X_2\right) \simeq 0.004.$$

While in the second scenario we generate X_1 and X_2 from the degenerate distribution that puts a point mass on $(240, 240)$, in which case the probability is 1. In the independent, uniform model it is easy to check that if there are three machines at the station, then this probability rises to

$$\begin{aligned} &\mathbb{P}(X_1 = X_2) + \mathbb{P}(X_2 = X_3) + \mathbb{P}(X_1 = X_3) - 2 \cdot \mathbb{P}(X_1 = X_2 = X_3) \\ &\simeq 0.0125. \end{aligned} \tag{1}$$

This means we would expect to see

$$0.0125 \, \frac{\text{ties}}{\text{station}} \times 7913 \text{ stations} \simeq 98.6 \text{ ties}$$

if the votes were drawn uniformly in this fashion and there were exactly three machines per station and exactly 7913 voting stations (some stations had 0

be proportional to the pre-assigned number of voters assigned to vote at a given machine within the mesa. This assignment is carried out based on the last two digits of the voter's Venezuelan ID. In this case the probability drops to approximately 1 in 10000.

total votes so were excluded from our model). Taken at face value, this figure seems to be evidence in favor of the YES side's claim as this number is much smaller than 311.

However, a little thought shows that this figure is evidence that the *(global) uniform model for all of the mesas* was wrong but does not provide an alternative. It does not provide any idea of how many of the individual mesas deviate from the mesa-level model, or how "far" they are from the model.

Of course, in the above example, the uniform distribution is, in a sense, the worst case scenario for counting the number of ties because it is so flat. If the counts had a more peaked distribution, say, Poisson, then the expected number of ties would increase. This model as well as others are addressed below in Section 3.

In Section 4 we investigate what conclusions can be drawn from the fact that 402 ties were observed, and most of the models predict that 350 ties is a reasonable number. The number of ties can be thought of as an *omnibus* test statistic, which we can formalize as follows. Given a model

$$\mathcal{M} = (\mathcal{M}_i, 1 \leq i \leq 7913)$$

for the counts in all 7913 mesas and the "true" models

$$\tilde{\mathcal{M}} = (\tilde{\mathcal{M}}_i, 1 \leq i \leq 7913)$$

for the counts, a test that compares the observed number of ties to the number of ties expected under $\mathcal{M}$ is a test of the *global null hypothesis*

$$H_{0,global} : \widetilde{\mathcal{M}_i} = \mathcal{M}_i, \quad \forall i. \tag{2}$$

Therefore, based on the observed number of ties, we can reject the hypothesis that *all* of the mesas followed the Poisson model, but we cannot conclude anything about which mesas did not follow the Poisson model. If we try to determine which mesas did not follow the Poisson model, then we would trying to determine which of the models $\mathcal{M}_i$ were wrong. In statistical terms, this can be thought of as a multiple comparisons problem[2].

In the spirit of multiple comparisons, and in order to investigate the YES side's claims further, in Section 4.1 we carry out an FDR analysis on the

[2]We emphasize also that even if a particular mesa does not follow the Poisson model does not necessarily mean that fraud was committed at that mesa.

deviance statistics of each individual mesa under the multinomial model, in order to see if widespread fraud is present in the elections. We are specifically looking for the type of fraud alleged by the YES side: "vote-capping" and other forms of fraud that would be expected to show the YES side's votes are underdispersed. Unlike the hypothesis tests based on the expected number of ties, this approach presupposes that a non-trivial number of mesas in the election did not follow the Poisson model. Given this hypothesis, the FDR analysis attempts to discriminate between machines that follow the Poisson model and "vote-capped" machines.

The results of this section show that there is no systematic way to discriminate between these two classes: alternatively, there is no evidence that there really is a non-trivial proportion of mesas in which either the YES or the NO counts do not follow the Poisson model. Further, there is no evidence that there is a non-trivial proportion of mesas in which the YES counts are underdispersed relative to the NO counts. This counters both the claims of "vote-capping" as well as claims that the YES votes are too narrowly dispersed.

We then investigate the YES side's second allegation of fraud, based on Benford's Law [5, 2, 6], which is a model for the distribution of significant digits $(D_1, D_2, D_3, \ldots)$ of observations in a dataset. In Section 5 we use the multinomial model of a "fair election" and find that its significant digit distribution is virtually identical to the observed distribution, which is different than Benford's Law. Hence, the "fair election" model is indistinguishable from the observations in terms of Benford's Law, but does have fewer ties than the observed data. In any case, if this model is accepted as a truly fair model of an election, then this shows that the fact that Benford's Law does not hold is not evidence of electoral fraud.

We conclude in Section 6 where we describe a very simple mechanism of fraud that is undetectable based solely on the count data, followed by a discussion.

2 Modelling the referendum

In this section, we describe a few models of "fair elections", and investigate the number of ties observed in these models.

In all our analyses, we will keep the number of stations to be 7913, as well as the number of machines per station to be fixed. Of these 7913 voting stations, 17% had only one voting station, so there could be no ties there, 23% had 2 voting stations and 59% had three voting stations. In other words, if we think of the N_i's as IID then the empirical p.m.f. of N_i is

$$p_N(1) = 0.17, \quad p_N(2) = 0.23, \quad p_N(3) = 0.59. \tag{3}$$

Let the YES/NO vote tallies by labeled $(X_{i,1}^{Y/N}, X_{i,2}^{Y/N}, X_{i,3}^{Y/N}), 1 \leq i \leq 7913$, with $X_3^{Y/N}$ and/or $X_2^{Y/N}$ possibly zero, depending on the total number of machines N_i at the i-th station. For instance, if the i-th mesa has three machines, we can summarize this information in the following table

	Machine # 1	Machine # 2	Machine # 3	Total
YES	$X_{i,1}^Y$	$X_{i,2}^Y$	$X_{i,3}^Y$	$\sum_{j=1}^3 X_{i,j}^Y = T_i^Y$
NO	$X_{i,1}^N$	$X_{i,2}^N$	$X_{i,3}^N$	$\sum_{j=1}^3 X_{i,j}^N = T_i^N$
Total	$X_{i,1}^{Y+N}$	$X_{i,2}^{Y+N}$	$X_{i,3}^{Y+N}$	$\sum_{j=1}^3 X_{i,j}^{Y+N} = T_i^{Y+N}$

Table 1: Summary of votes in a mesa with three machines

The models we will consider are the following:

- $X_{i,j}^{Y/N}, 1 \leq i \leq 7913, 1 \leq j \leq N_i$ are IID Poisson$(\lambda^{Y/N})$ random variables;
- $X_{i,j}^{Y/N}, 1 \leq i \leq 7913, 1 \leq j \leq N_i$ are IID random variables, not necessarily Poisson;
- for each i, $X_{i,j}^{Y/N}, 1 \leq j \leq N_i$ are IID Poisson$(\lambda_i^{Y/N})$ random variables;
- for each i, $X_{i,j}^{Y/N}, 1 \leq j \leq N_i$ are Multinom$(T_i^{Y/N}; N_i^{-1}, \ldots, N_i^{-1})$
- a multiple hypergeometric model obtained by conditioning on the marginal row/column sums;
- a hierarchical parametric bootstrap model described below.

We find the final four models the most realistic, for reasons described below, and their results agree quite well.

3 The expected number of ties

3.1 Independent, Identically Distributed Poisson

The Poisson random variable, arguably the statisticians' favorite model for count data, is a natural first approximation. The MLEs for $\lambda^{Y/N}$ are

$$\widehat{\lambda}^{Y/N} = \frac{\sum_{i=1}^{7913} T_i^{Y/N}}{\sum_{i=1}^{7913} N_i}$$

or

$$\widehat{\lambda}^{Y} = 188.1, \qquad \widehat{\lambda}^{N} = 258.1.$$

If the model is correct, if we looked at the histogram, or the empirical p.m.f. of the YES votes, it should look (at least qualitatively) like the p.m.f. of a Poisson(188) random variable. Likewise, the empirical p.m.f. of the NO votes should look like the p.m.f. of a Poisson(258) random variable. Figure 1 shows that this is clearly not the case.

The expected number of ties under this model are 320 for YES and 274 for NO.

3.2 Independent, Identically Distributed

The next model we consider is one where the observed votes are independent and identically distributed, but not necessarily Poisson. The non-parametric MLE $\widehat{f}^{Y/N}$ of the p.m.f. $f^{Y/N}$ for the YES and NO votes is just the empirical p.m.f. which yields an expected number of ties of 58 for YES and 55 for NO.

This model does indeed have a very low expected number of ties, however it assumes that every mesa in the country had on average the same number of voters, and behaved in the same way. That is, this model ignores variability across different regions of the country. A model that allowed some flexibility across stations may capture this variability. The next model allows the distribution of the number of people to vary across the different mesa.

3.3 Poisson counts, varying by mesa

The next model we consider is one where the observed votes are independent, and Poisson but the parameter varies by mesa. The MLE $\widehat{\lambda}_i^{Y/N}$ is just

$$\widehat{\lambda}_i^{Y/N} = \frac{1}{N_i} T_i^{Y/N}.$$

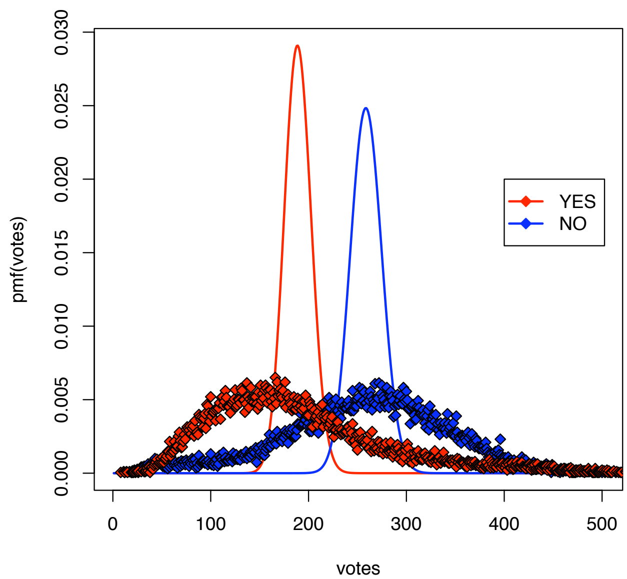

Figure 1: Empirical and Poisson fit probability mass functions for the YES and NO votes

Keeping number of machines per station is independent of the counts, the MLE of the expected number of ties is therefore

$$\begin{aligned}\mathbb{E}\,(\ \#\ \text{ties for YES}) &= \sum_{i=1}^{7913} \mathbb{P}_{\widehat{\lambda}_i^Y, N_i}\,(\text{there is a tie between } N_i \text{ machines}) \\ &= 344 \pm 19 \\ \mathbb{E}\,(\ \#\ \text{ties for YES}) &= \sum_{i=1}^{7913} \mathbb{P}_{\widehat{\lambda}_i^Y, N_i}\,(\text{there is a tie between } N_i \text{ machines}) \\ &= 290 \pm 17.\end{aligned}$$

The error cited above represents the error from the same assumption as in the IID Poisson model.

3.3.1 Conditional model

Another way to carry out this analysis, as proposed by some on the YES side is to condition on the total number of YES votes in a mesa, and look at the expected number of ties conditioned on having observed this many yes votes. Under the Poisson assumption, a simple exercise shows that (assuming the votes are split among machines in a given mesa according to the proportions $(\pi_{i,1}, \ldots, \pi_{i,N_i})$

$$(X_{i,1}^{Y/N}, \ldots, X_{i,N_i}^{Y/N})\big|T_i^{Y/N} \sim \text{Multinom}(T_i^{Y/N}, \pi_1, \ldots, \pi_{N_i}). \tag{4}$$

Assuming that the votes are split up equally among the machines, i.e. that $\pi_{i,j} = \frac{1}{N_i}$ yield an expected number of ties of 344 for YES and 290 for NO.

3.4 Multiple hypergeometric model

Yet another model for the data is to condition on the total number of votes cast in each machine within a mesa, as well as the total number of YES and NO votes within the mesa. That is, we condition on

$$X_{i,j}^{Y+N}, T_i^{Y/N}.$$

If we assume that the observed table is drawn uniformly at random from all tables with these fixed marginal totals, we obtain the multiple hypergeometric model. Simulations for this model were carried out [4] yielding an expected number of ties of 360 for YES and 317 for NO. It is also possible, with the help of a computer, to explicitly compute the expected number of ties in this model. However, the routines used to do this computation are prone to some roundoff error. Nevertheless, these calculations yield an expected number of ties of 367 ties for YES and 322 ties for NO.

3.5 Parametric bootstrap model

The final model we considered was a parametric bootstrap model where the counts were generated according to the integer part of a multivariate Normal distribution. Specifically, we used the empirical covariance matrices of the normalized total vote counts

$$r_{i,j}^{Y+N} = \frac{X_{i,j}^{Y+N} - \widehat{\lambda}_i^{Y+N}}{\sqrt{\widehat{\lambda_i^{Y+N}}}}$$

to generate new "residuals" $\varepsilon_{i,j}^{Y+N}$, the new counts were the integer part of

$$\widetilde{X}_{i,j}^{Y+N} = \widehat{\lambda}_i^{Y+N} + \varepsilon_{i,j}^{Y+N} \cdot \sqrt{\widehat{\lambda}_i^{Y+N}}.$$

Then, conditional on $\widetilde{X}_{i,j}^{Y+N}$, the YES votes were assumed Binomial with proportions determined by the observed number of YES votes to the total number of votes in the particular mesa, i.e

$$\widetilde{X}_{i,j}^{Y} \big| \widetilde{X}_{i,j}^{Y+N} \sim \text{Binom}\left(\widetilde{X}_{i,j}^{Y+N}, \frac{\widehat{\lambda}_i^{Y}}{\widehat{\lambda}_i^{Y+N}}\right).$$

Under this model, the expected number of for YES is 345 and 292 for NO.

3.6 Summary

The following table is a summary of the models presented above. The approximate Z-score in the table is based on a Poisson approximation to the number of ties and the Normal approximation to the Poisson. That is, given we expected to observe $\lambda_{\mathcal{M}}^{Y/N}$ ties for YES/NO under a given model $\mathcal{M}$, the Z-score for this model is

$$Z_{\mathcal{M}}^{Y} = \frac{402 - \lambda_{\mathcal{M}}^{Y}}{\sqrt{\lambda_{\mathcal{M}}^{Y}}}, \qquad Z_{\mathcal{M}}^{N} = \frac{311 - \lambda_{\mathcal{M}}^{N}}{\sqrt{\lambda^{N}}_{\mathcal{M}}}.$$

Model	λ^Y	Z^Y	λ^N	Z^N
3.1	320	4.6	274	2.2
3.2	58	45.2	55	34.5
3.3	344	3.1	290	1.2
3.3.1	348	2.9	294	1.0
3.4	360	2.2	317	-0.3
3.5	345	3.1	292	1.1

Table 2: Summary of the various models considered

4 What can we conclude from the models?

In Table 2, we presented a summary of a number of models of the number of ties in an election similar to the Venezuelan recall election. The YES side (the losing side) claimed that the number of ties observed is much higher than would be expected under a reasonable election, therefore it is evidence of fraud.

While the Z-scores are fairly high, as was pointed out in the introduction this only means that we can reject the global null hypothesis (2), and not that there indeed was fraud. As the model the YES side has chosen to use as evidence of fraud is the multinomial model (4) , we will focus on this model as a model of a "fair election."

With this in mind, there are many other goodness-of-fit tests in the statistical literature. For instance, we could use the multinomial deviance as a test statistic[3]: if a mesa has N_i machines, then the deviance for the i-th mesa with probabilities $\pi_i = (\pi_{i,1}, \ldots, \pi_{i,N_i})$ is

$$D(X_{i,1}^{Y/N}, \ldots, X_{i,N_i}^{Y/N}; \pi_i) = -2 \cdot \sum_{j=1}^{N_i} X_{i,j}^{Y/N} \cdot \left(\log \pi_{i,j} - \log \left(\frac{X_{i,j}^{Y/N}}{T_i^{Y/N}} \right) \right). \quad (5)$$

If the Poisson model at the i-th mesa were correct then the two deviances are approximately independent $\chi^2_{N_i-1}$ random variables for all mesas with $N_i > 1$.

Summing all the χ^2 statistics yields a global goodness of fit test, or, equivalently a test of (2) with independent Poisson counts in each mesa. The results are: $\chi^{2,Y} = 10824, p = 0.016$, on 11142 degrees of freedom, and

[3] Instead of the deviance, we could use the multivariate Normal approximation to the multinomial, though the results presented below are virtually identical.

$\chi^{2,N} = 12466, p = 0$ on 11142 degrees of freedom. Further inspection shows that the χ^2 for the NO is driven by a handful of outliers in mesas with 2 machines. This can be seen by the peak at the right hand side of the histogram of p-values for the mesas with 2 machines (see Figure 2 below). If we remove these 21 mesas (0.3% of the mesas accounting for 24016 votes or 0.3% of the total votes cast), whose results are listed below, then the χ^2 test still rejects: $\chi^{2,Y} = 10180, p = 1.0 \cdot 10^{-10}$ on 11106 degrees of freedom and $\chi^{2,N} = 10672, p = 0.0016$ on 11106 degrees of freedom.

This indicates that, in a global sense, both the YES and the NO results seem to be slightly underdispersed relative to what we would expect under the multinomial model (4). However, these tests also do not indicate presence of widespread departures from the Poisson model: they indicate the global null $H_{0,global}$ is false. A more important question is: *How false is $H_{0,global}$?* We answer this question (on some level) in the following section.

4.1 FDR analysis of deviance statistics

Under the global model that *all* mesas follow a Poisson model, if we compute the multinomial deviance for YES and NO we would expect them to behave like two independent samples of size 6507 (the number of mesas with more than one electronic voting machine).

If there was a systematic departure from this Poisson model, we would expect a systematic departure from the respective theoretical χ^2 distributions for the 6507 deviance statistics. In particular, the YES side alleges that there has been "vote-capping." This would manifest itself in a large left tail near 0 for the χ^2 statistics. Figures 2 and 3 show the histograms of the deviance statistics for the YES and NO votes stratified by the number of machines in the mesa, the coloured curve show the theoretical density we would expect if *all* of the mesas followed the Poisson model. Also included are the transformed p-values

$$U_i^{Y/N} = \mathbb{P}^{-1}_{\chi^2_{N_i-1}} \left(D(X_{i,1}^{Y/N}, \ldots, X_{i,N_i}^{Y/N}; \pi_i) \right)$$

which, under the Poisson model for the i-th mesa should be Unif(0,1). Note that these are the p-values for testing whether there is underdispersion in each mesa.

From the figures, there appears to be a slighly heavier left tail than would be expected under the Poisson model for both the YES and the NO votes.

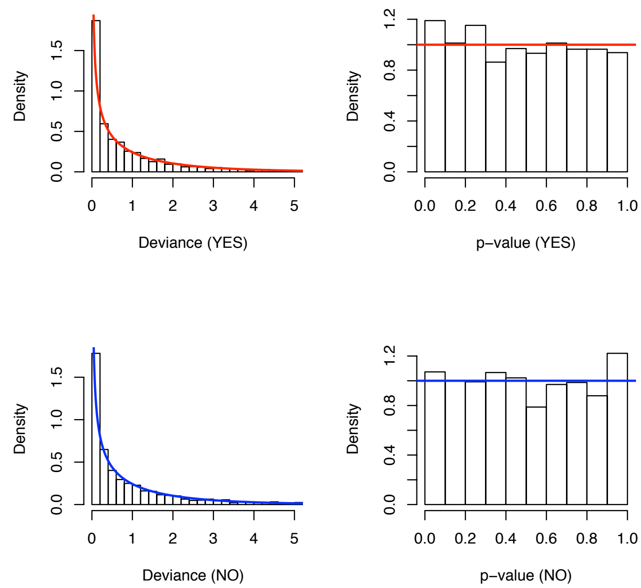

Figure 2: Deviance statistics for YES and NO votes in mesas with 2 machines

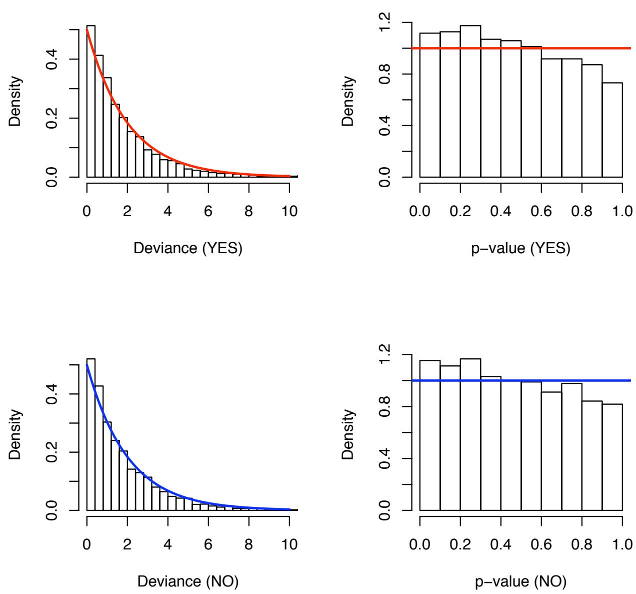

Figure 3: Deviance statistics for YES and NO votes in mesas with 3 machines

However, the tail is not extraordinarily different than what would be expected under the Poisson model. One way to get a handle on how different this is than what would be expected under the Poisson model is through a mixture: if some machines were not tampered with while the rest were not, what we would expect to see in the histogram of p-values is a mixture of Unif(0,1) and some alternative distribution F_1, or, in terms of densities

$$f(x) = \pi_0 \cdot 1 + (1 - \pi_0) \cdot f_1(x).$$

There are many ways to fit such a model, to estimate π_0 and the alternative density f_1 (though the pair (π_0, f_1) are unidentifiable, strictly speaking). There is actually much current research in such problems in statistics, specifically in the genomics, neuroimaging and other "high throughput" types of data. One popular approach to such multiple-comparison problems is the use of the False Discovery Rate (FDR) [1] or the local FDR [3]. We will not address the statistical issues here, but refer the readers to some of the relevant literature [3, 7, 8].

We will use the FDR analysis to focus on some properties of the alternative that can be inferred from the samples $\{U_i^{Y/N}, 1 \leq i \leq 6507\}$. Roughly speaking, if we run a "standard" tail FDR analysis on the samples $\{U_i^{Y/N}, 1 \leq i \leq 6507\}$ what is output is an FDR curve which, as a function of q which can be thought of as an estimate of the posterior probability that a given mesa follows the Poisson model for the YES/NO, given we had observed a YES/NO p-value of less than or equal to q for this mesa [7]. The local FDR [**?**] yields an estimate of the same probability, only it is conditioned on observing a YES/NO p-value of exactly q for this mesa.

Figures 4 and 5 show, for both the YES and NO votes, a plot of the empirical FDR curve, a version of the FDR curve based on a smooth estimate of the distribution funcation, as well as the local FDR curve. If there were systematic departures from the Poisson model, we would expect that there would be non-trivial regions where the FDR or local FDR is very low.

As can be seen from the figures, they are relatively high for all values of q, with two exceptions. First of all, the empirical FDR curve goes to zero, but this curve has to go to zero. Second, the local FDR curve drops down for q-values near 1. The disagreement between the two FDR curves and the local FDR curve makes sense here because the FDR is sensitive only to changes at small q-values whereas the local FDR can in theory detect differences at all values of q. The fact that the local FDR curve drops to zero indicates

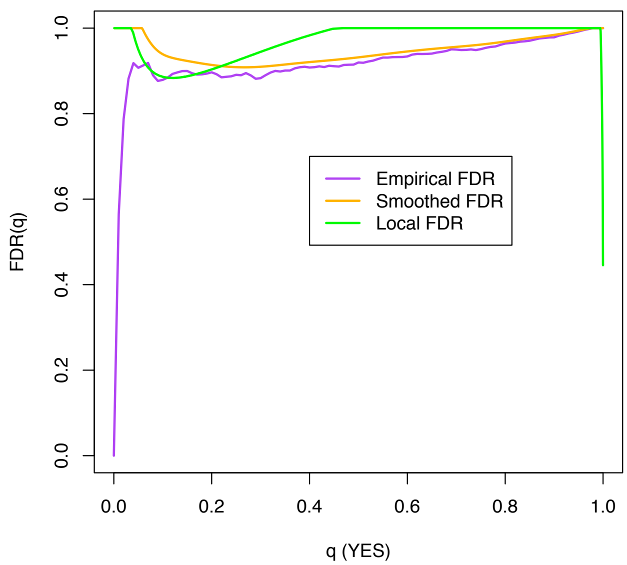

Figure 4: FDR curves for the YES votes

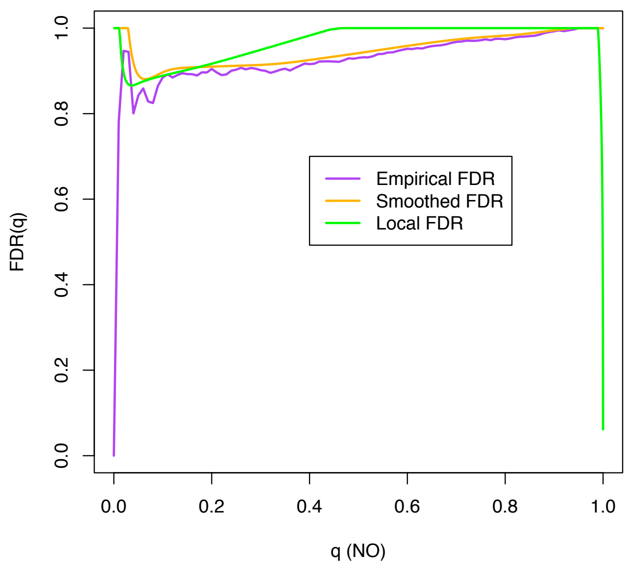

Figure 5: FDR curves for the NO votes

there are some mesas in both the YES and the NO where the counts are over dispersed compared to the Poisson. This departure from the Poisson model (in 8 mesas) is not consistent with widespread "vote-capping", and actually counters another YES side claim that the votes are underdispersed.

To further study the claims of underdispersion, we looked at the mesa level F statistics given by the ratio

$$F_i = \frac{D(X^N_{i,1}, \ldots, X^N_{i,N_i}; \pi_i)}{D(X^Y_{i,1}, \ldots, X^Y_{i,N_i}; \pi_i)}$$

and the p-values

$$U^F_i = 1 - \mathbb{P}^{-1}_{F_{N_i-1,N_i-1}}(F_i)$$

for testing whether the dispersion in the YES votes is larger than the NO votes, i.e. these p-values would be expected to be smaller if the the YES votes were underdispersed relative to the NO votes. As can be seen in Figure 6, the histogram of the p-values and Figure 7, the same 3 FDR curves as above, there is no evidence that the dispersion of the NO votes is any different than that of the YES votes. If "vote-capping" were taking place only in the YES votes, then this would not be expected to happen.

In summary, looking at the FDR analyses:

> *If this data came from an exploratory analysis of a scientific experiment, the conclusion would almost certainly be that there really are no clear departures from the Poisson model, i.e. that there are no "interesting findings" in the data.*

5 Benford's Law

Benford's Law, a model for significant digits in a dataset, can be stated as:

$$\mathbb{P}\left(D_1 = d_1, \ldots, D_k = d_k\right) = \log_{10}\left(1 + \left(\sum_{i=1}^{k} d_i \cdot 10^{k-i}\right)^{-1}\right). \qquad (6)$$

It is commonly used as a tool for fraud detection in various scenarios, such as tax evasion [6]. There is considerable empirical evidence that if one studies the digits of large diverse datasets, then their digits obey Benford's Law. Formal proofs of this fact are based on an assumption of mixing and scale

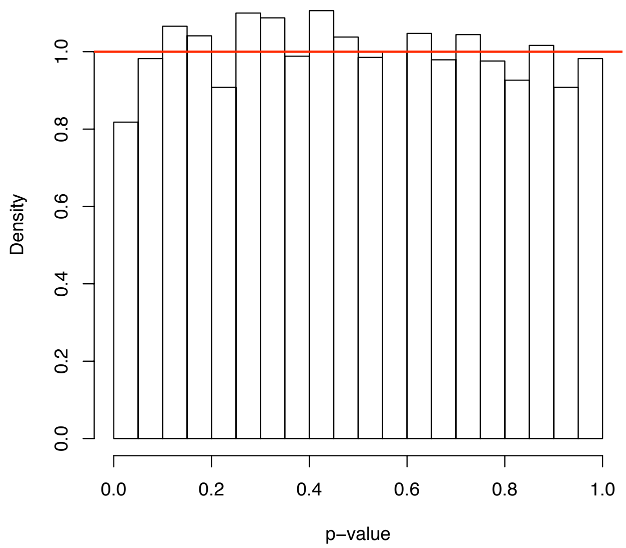

Figure 6: Histogram of the p-values of the F-tests

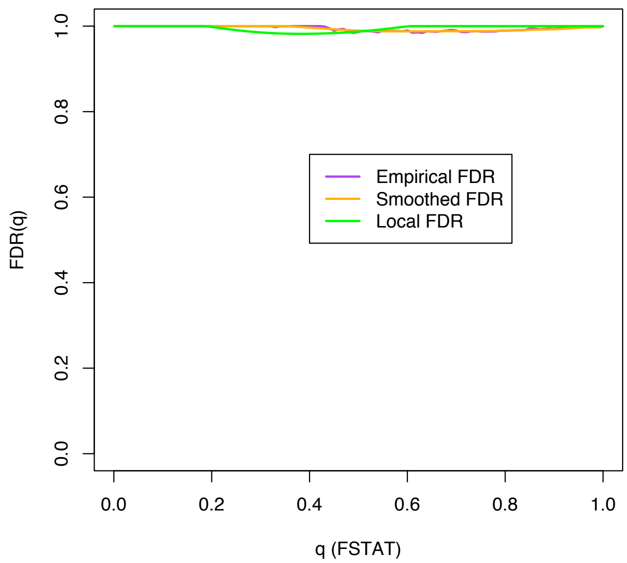

Figure 7: FDR curves for the F-statistics votes

invariance, which may not be true in any given election. Generally speaking, the datasets in which Benford's Law appear are *supermixing*, that is, mixtures of data from many different sources. That is, you should probably expect to see Benford's Law if you combine data from

- baseball, such as individual players' batting averages;
- sizes of stars in various galaxies;
- diameter of trees in a boreal forest;
- diameter of clams in the Pacific;
- ...

To see why mixing is important, consider the first example above: baseball players' batting averages. Most players averages fall between 0.100 and 0.399, so if we looked at the empirical distribution of the first significant digit, it would have almost no mass on $\{1,4,5,6,7,8,9\}$ (0 is not counted as a 1st significant digit for obvious reasons), while Benford's Law places mass

$$\log_{10}\left(1+\frac{1}{d_i}\right), 1 \le i \le 9$$

on each digit.

Often, electoral districts are chosen to have roughly the same number of voters in each region. For concreteness, let us suppose that each region in our hypothetical region has

$$X_i \sim \text{Poisson}(\lambda), 1 \le i \le n$$

voters. If, further, the proportion of voters in each region favoring the YES side was approximately p across regions, then the number of YES votes in each region would be close to

$$X_i^Y \sim \text{Poisson}(p\lambda)$$

and the significant digit distribution we would observe would be that arising from a Poisson model and not Benford's Law.

Real elections, of course, are more complicated so perhaps the significant digit distribution we observe might be a mixture of those arising from

Poisson's and how close it is to Benford's Law depends on how *mixing* the mixture is. Here, we use the multinomial model (4) of a "fair election" and find that its significant digit distribution is virtually identical to the observed distribution, which is different than Benford's Law. Specifically, Figures 8, 9 and 10 show the significant digit distribution of the observed YES votes; a constant 40 % proportion / varying number of voters Poisson model for the YES votes[4]; and the multinomial model (4).

As all three figures are virtually identical, Benford's Law is of little use in fraud detection in this instance: two of the three models above are fair election models.

6 A scenario of undetectable fraud

Based on the FDR analysis presented above, we can conclude that there is no widespread departure from the Poisson model, and can reject the YES side's claim that "vote-capping" took place.

In this section, we describe a simple (possibly the simplest) scenario of fraud that is undetectable based just on the count behaviour. That is, looking at the number of ties, or the deviance statistics, we would be unable to consistently decide between two competing scenarios: one in which the electoral authorities commited fraud, and one where no fraud has been committed. The most reliable way that such fraud can be detected is through a paper trail where each individual piece of paper has been verified by the voter, as was done in the Venezuela election. This is in contrast to many of the electronic voting machines to be used in the upcoming election. We emphasize that the results in this section do not constitute any evidence of fraud in the Venezuelan elections, they simply offer a scenario in which fraud is, provably, statistically undetectable.

Suppose then, that the true votes cast in each voting machine in the election were independent Poisson random variables with means $\widetilde{\lambda}_i^{Y/N}, 1 \leq i \leq 7913$, (i.e. means constant across machines within a mesa, or, weighted according to the proportion of people assigned to that machine) and suppose that the machines were programmed in such way as to randomly change YES votes to NO votes with probability p_i, independently across machines and mesas. Then, the apparent counts registered by the machines would also be

[4]Formally, $X_i^Y \sim \text{Poisson}(0.4 \cdot T_i^{Y+N}/N_i)$.

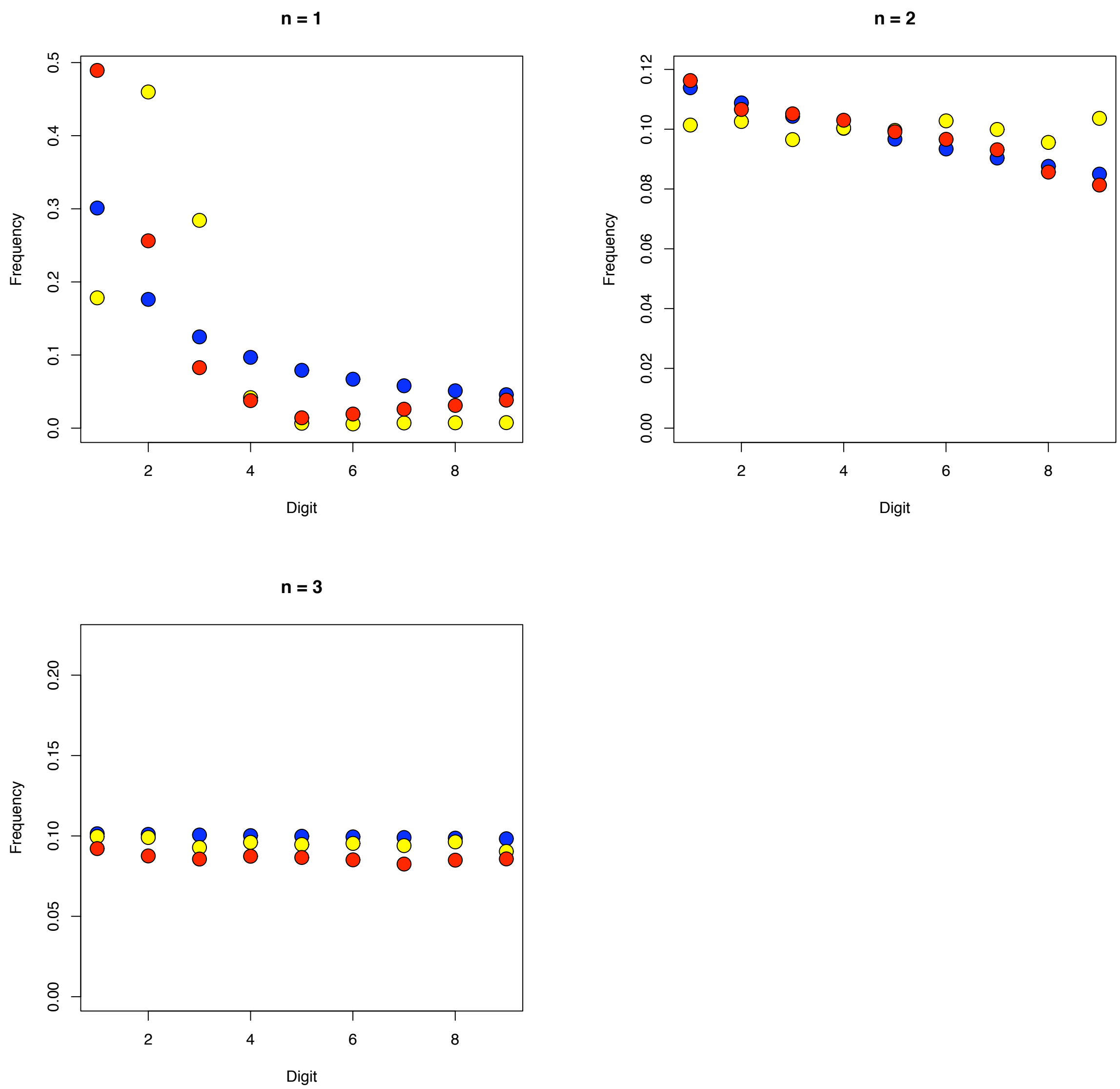

Figure 8: Digit distributions for first three digits of observed data: ●, YES; ●, NO; ●.

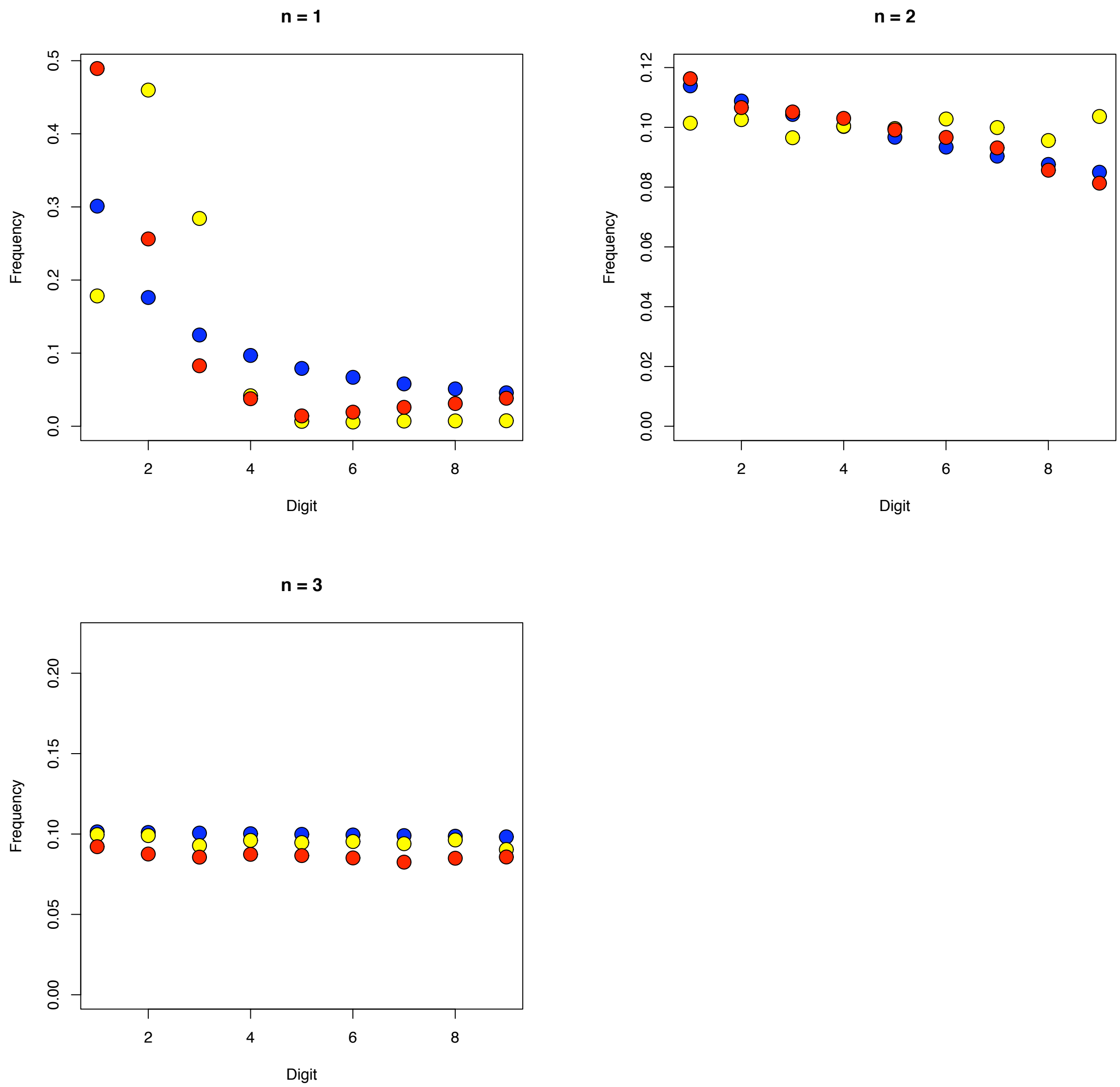

Figure 9: Digit distributions for first three digits of constant 40 % proportion / varying number of voters Poisson model: •, YES; •, NO; •.

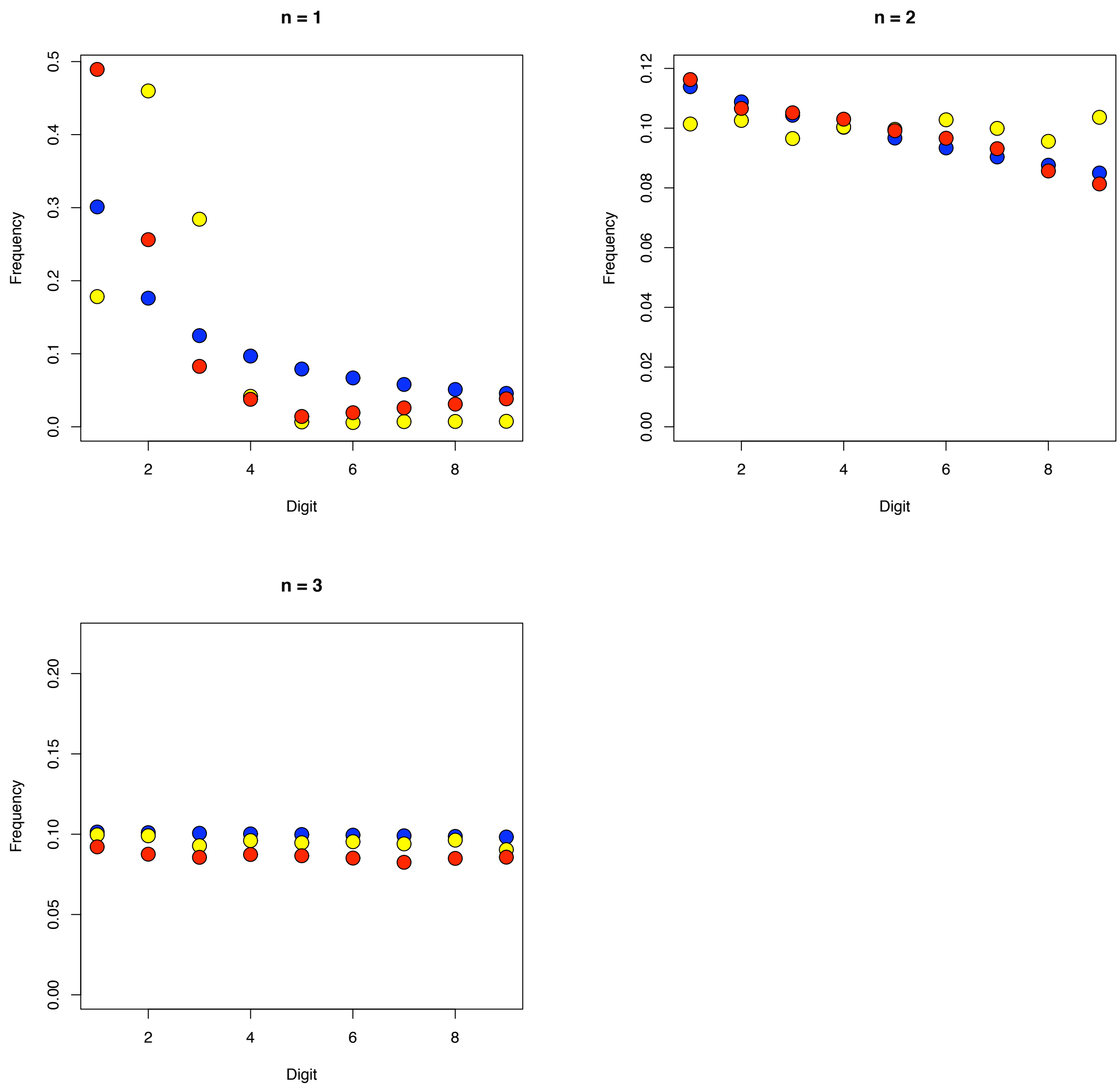

Figure 10: Digit distributions for first three digits of multinomial model (4): ●, YES; ●, NO; ●.

independent Poisson random variables with means

$$\lambda_i^Y = \widetilde{\lambda}_i^Y + p_i \cdot \widetilde{\lambda}_i^N.$$

Any test statistic based on just counts which seeks to reject the Poisson model would then be incapable of distinguishing between fraud and no fraud because for any values of p_i, the distribution of the counts follows a Poisson model.

It should be noted that, if this indeed did happen in Venezuela, then, given the Carter Center's work manual audits, roughly 1/4 of the people who voted YES in the mesas who that were audited would have their paper result disagree with their electronic result. This proportion would have to increase if we accept the YES side's claim that approximately only 1/3 of the centers were "eligible" to be audited according to the CNE's random number generator.

The main point here is that constructing a test to reject the Poisson model cannot detect this very simple form of fraud. We also emphasize the importance of having a paper receipt that is independently verified by the voter. Without this, this very simple fraud scenario is statistically undetectable based solely on the count data.

7 Conclusion

The results of this paper, both using the expected number of ties and the more powerful χ^2 tests show that the global Poisson model is not correct. However, this is not evidence of fraud, or even of widespread departures for the Poisson model. To investigate how widespread the departure from the Poisson model was, we carried out an FDR analysis of the deviance statistics at each mesa. We also studied claims of fraud based on Benford's Law, finding this technique unable to distinguish between the observed data and two fair election models.

The results of this analysis are essentially:

> *It is statistically impossible to distinguish between the results of a mesa chosen at random from the results of the Venzuelan recall election, and the results of a mesa that really did follow the Poisson model.*

In other words, these count data do not show any systematic signs of underdispersion relative to the Poisson: this counters both the claims of "vote-capping" and "tightly clustered votes" of the YES side.

References

[1] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc., Ser. B.*, 85:289–300, 1995.

[2] Persi Diaconis and David Freedman. On rounding percentages. *J. Amer. Statist. Assoc.*, 74(366, part 1):359–364, 1979.

[3] B. Efron. Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *Journal of the American Statistical Association*, 99:96–104, 2005.

[4] E. W. Felten, A. D. Rubin, and A. Stubblefield. Analysis of voting data from the recent venezuela referendum. Available at `http://venezuela-referendum.com`.

[5] Theodore P. Hill. A statistical derivation of the significant-digit law. *Statist. Sci.*, 10(4):354–363, 1995.

[6] M. Nigrini. A taxpayer compliance application of Benford's law. *J. Amer. Tax. Assoc.*, 18:72–91, 1996.

[7] J. D. Storey. The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics*, 31:2013–2035, 2003.

[8] J. D. Storey, J. E. Taylor, and D. O. Siegmund. A unified estimation approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, 66:187–205, 2004.