Title

Introduction

Due to the coronavirus pandemic in 2020, the 46th U.S. presidential election was situated in unprecedented territory. The Center for Disease Control and other health professionals highly discouraged large groups of people congregating in public at the time (CDC 2023). In-person voting was rendered a hazardous activity. As a result, the 2020 elections saw a massive uptick in absentee voting via mail-in ballots (Scherer 2021). On the one hand, the ability for the electoral system to accommodate this form of voting can be seen favorably. Without absentee voting, voters who were uncomfortable with voting in person still had a way to make their voices heard. However, absentee voting has consistently raised skepticism over its resilience to fraud (Ballotpedia, n.d.). This skepticism was on full display in the aftermath of the of 2020 presidential election, in which democrat Joseph Biden took home the victory. Members of the Republican party, including their nominee Donald J. Trump, questioned the legitimacy of the election, specifically citing the use of fraudulent absentee ballots to manipulate the votes in Biden's favor (Bekiempis 2020).

Despite official investigations finding no evidence of fraud, advocates of this corruption theory point towards a well-established number phenomenon to justify their beliefs: Benford's law (Campaign Legal Center, n.d.; Reuters 2020). This law, which states that the first digit of naturally occurring numbers has a non-uniform distribution, is commonly used to raise red flags about potential fraud in accounting and other domains. Neigh Sayers of the election adopted this law to analyze the first digit of votes for Trump and Biden by precincts in states like Wisconsin and Georgia. *Investigators* discovered while votes for Trump followed the Benford's law, the data for Biden strongly deviated from the expected distribution (Reuters 2020). Clearly something was up with Biden's votes, right? Can detecting election fraud with Benford's law be this easy?

Walter R. Mebane, a political science academic who formalized using Benford's law for election fraud detection, condemned the use of the law in this way (Walter R. Mebane 2020). It is clear using Benford's law as a red flag test for election is not as cut and dry as the modeling the distribution of the first digits and assessing whether they fit a specified distribution. Namely, issues regarding the appropriateness of Benford's law in this context has questioned what deviations from the law really mean, how to test for deviations, and even if the assumptions of the numbers themselves are valid (Deckert, Myagkov, and Ordeshook 2011; L. Pericchi and Torres 2011; Walter R. Mebane 2011). Many of these questions have been addressed by Mebane in his previous work, but what remains is a full assessment of the 2020 election data to "prove" once and for all if Biden's victory was fictitiously curated through fraudulent absentee ballots.

This work is meant to examine the application of Bedford's law, as described by (Walter R. Mebane 2006), on the 2020 presidential election data. The paper will establish Mebane's 2BL framework, assess additions to and criticisms of this framework, and finally apply the framework to a convicted case of absentee ballot fraud (Bladen County, NC, 2018 General Election) and an alleged case (the 2020 Presidential Election). Overall, this work is meant to underscore the potential of Benford's Law to detect fraud in a changing election landscape.

Background

Benford's Law

The Math behind Benford's Law

Benford's Law, also known as the Newcomb-Benford Law or the first digits law, is an observation that the first digit of naturally occurring numbers do not follow a uniform distribution. Instead, these numbers follow a distribution modeled by $Pr(D_1 = d) = log_{10}(\frac{d+1}{d})$ for all d = 1,..,9 Hill (1995). While this equation may seem *absurd, it tracks mathematically.*

Consider a natural number X that is written in scientific notation so $X = r*10^n$. The first digit of r is the first digit of X and n is the magnitude of X. Taking the log_{10} of X gives $log_{10}(X) = log_{10}(r) + n$. For the first digit of X to be 1, $r \in [1, 2)$ meaning $log_{10}(r) \in [log_{10}(1), log_{10}(2))$. So, the first digit of X is 1 when $log_{10}(1) + n \leq log_{10}(r) + n < log_{10}(2) + n$. The inequality simplifies to $n \leq log_{10}(r)+n < 0.301$, so the first digit of X is 1 when $log_{10}(r) \in [0, 0.301)$. Since $log_{10}(r) \in [0, 1)$ for $r \in [1, 10)$, the probability that the first digit of X is 1 is approximately 30.1%. For the probability that the first digit of X is any number i for i=1,2,..,9, one would need to subtract $log_{10}(i + 1)$ from $log_{10}(i)$, leading to the formula from above ("A Simple Explanation of Benford's Law," n.d.).

Benford's Law also extends to the second digit which is modeled by $Pr(D_2 = d) = \sum_{j=1}^{9} log_{10}(1 + \frac{1}{10j+d})$.

Table 1: Distribution of the First Digit by Benford's Law

digit	1	2	3	4	5	6	7	8	9

probability	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

Table 2: Distribution of the Second Digit by Benford's Law

digit	0	1	2	3	4	5	6	7	8	9
probability	0.120	0.114	0.109	0.104	0.100	0.097	0.093	0.090	0.088	0.085

Characteristics of "Benford-like" Data

Engineer and physicist Francis Benford, who formalized Benford's Law, found over 20,000 diverse sets of numbers that followed this first digit rule. These sets ranged from baseball statistics to heats of electrical compounds (Hill 1995). However, not all sets of numbers are expected to be "Benford-like." Phone numbers, identification numbers, and random draws from a uniform distribution are just some examples of data that will likely not follow the law. Moreover, there is no theory behind Benford's Law and therefore no consensus on why some sets of numbers follow this rule while others do not (Boyle 1994). Ideas of factors effecting the Benford-ness of sets of numbers are numerous, but there are some consistent findings across inquiries in diverse fields.

For one, mathematician Theodore Hill, PhD, has hypothesized that numbers which follow Benford's Law come from many different random distributions. These numbers can be thought of as coming from second generation distributions. Moreover, he argues that the initial distribution which form the secondary distribution do not have to produce Benford-like data themselves. Instead, it is the combination of the distributions that leads to the final numbers reaching a Benford distribution (Hill 1995). This facet helps frame why human-made numbers, like phone numbers, and numbers from singular distributions are not expected to follow this law. Hill's explanation of the data generating processes that adheres to Benford's Law appears to be widely accepted in applications of the law Durtschi, Hillison, and Pacini (n.d.) .

Following from Hill's work, it has also been theorized that Benford-like data arises from the multiplication and division of random variables (Boyle 1994). Random numbers assigned to higher and higher powers also appear to be more Benford-like (Adhikari and Sarkar 1968). Like with Hill's conception, these also assert that the numbers being manipulated do not have to follow Benford's Law themselves (Boyle 1994). Moreover, Boyle, PhD, argues that once a set of numbers reaches the Benford distribution, it remains Benford-like even with further mathematical manipulations. He also asserts that all naturally occurirng numbers can be thought of as products or quotients of random variables (Boyle 1994).

Other considerations to whether a set of numbers will follow Benford's law revolve around peculiarities of how the data is collected and maintained. In accounting data, scholars note that distinct transaction patterns may case an account not to follow Benford's law. Durtschi, Hillison, and Pacini (n.d.) use an example of a medical account that failed to conform to the expected distribution due to a large number of repeated transactions of the same supplies. Also from Durtschi, Hillison, and Pacini (n.d.)'s evaluation of Benford's Law in regards to accounting data, they find that account, or generally sets of numbers, that have a minimum or maximum value should also not be expected to follow the law. Finally, it has been noted that data will appear more Benford-like when the data represents a larger range of magnitudes and when the number of observations increases Durtschi, Hillison, and Pacini (n.d.).

Applications of Benford's Law

As alluded to previously, Benford's Law has applications in various fields. Its most common use in these fields is to detect fraud or otherwise undue manipulation to naturally occurring numbers. The idea behind this application is if a set of naturally occurring numbers do not follow Benford's distribution, then there is evidence that the data was manipulated. The use of this law in fraud detection is common in accounting, *()*, election forensics, which will be expanded upon later, and more Durtschi, Hillison, and Pacini (n.d.).

However, sets of numbers failing to adhere to this law are not always indicative with fraud. For instance, considering the medical account from Durtschi, Hillison, and Pacini (n.d.) discussed earlier, the data did not appear Benford-like due to valid transaction patterns, not fraud. Some instances of fraud may also not be detected by analyzing first digit distributions, such as if *idk*. Additionally, the tests used to determine Benford-ness may be flawed in their use and interpretations, which will be discussed in subsequent sections L. Pericchi and Torres (2011). In general, there are high consequences for false positives (detecting fraud when there is none) and false negatives (not detecting fraud when there is). Scholars in the field have therefore advocated for Benford's Law's use as a red flag test for detecting fraud and *not the final say* (Durtschi, Hillison, and Pacini, n.d.).

Election Forensics with Benford's Law

Overview

Elections play a crucial role in fairly gauging the opinions of a population within various political systems. However, when elections are manipulated, the very foundation upon which they are built can crumble, leading to severe consequences for the affected populations. Election forensics emerges as a discipline utilizing statistics "to analyze numerical electoral data and detect where patterns deviate from those that should occur naturally, following demonstrated mathematical principles" (Hicken and Mebane 2017, 1). This method offers distinct advantages over traditional approaches like in-person monitoring and parallel voting tabulations. It uses data at granular data levels, follows a systematic approach, and incorporates measures of uncertainty (Hicken and Mebane 2017). Given election forensics' *framework*, scholars have justified using Benford's Law as another framework to detect election fraud.

Second Digit Distribution

Mathematician Luis Raùl Pericchi, PhD, and David Torres (**can't find anything on him, nvm found [this](https://www.linkedin.com/in/davidtorresv/)**) introduced assessing the second digit distribution on voter counts against the Benford Law distribution in 2004. They showed law could be applied to the 2004 Venezuelan recall referendum to remove acting president Hugo Chávez and found statistical anomalies with the votes that served to keep Chávez in power. (L. R. Pericchi, n.d.). Pericchi and Torres findings from Benford's law were supported by other analyzes that found systematic errors in the election (**wiki2004venrecallref?**).

Political scientist Walter Mebane, PhD, expanded Pericchi and Torres' application of Benford's Law by explaining how and why the *law was a fit for election forensics*. To begin, he specifically advocates for analyzing the distribution of the second digit of vote counts to detect fraud. He refers to this method to second Benford's Law, or 2BL, as distinct from Benford's Law, BL, which refers to the first digit. Mebane argues 2BL is more appropriate for vote counts, especially at the precinct level, as most precincts are designed to have a similar number of constituents. Therefore, if the percentage of constituents in support of a specific candidate is similar across precincts, most of the vote counts will have similar first digits (Walter R. Mebane, n.d.). For example, if there is approximately 5,000 constituents per precinct and around 10% of each precinct supports a specific candidate, many of the vote counts for said candidate will start with 4, 5, and 6. Similar to Durtschi, Hillison, and Pacini (n.d.) minimum-maximum argument, constraints to precinct size clearly influences first digits. Therefore, Mebane suggests that the second digit is less influenced by size constraints and therefore is more reasonable to apply Benford's Law (Walter R. Mebane, n.d.).

To account for 2BL's appropriateness for vote count data, Mebane also calls for a rethinking of vote count data as just the sum of random coin flips representing a population's probability of voting one way or another. If this was the case, vote counts would not be complex enough for Benford's law to apply. Instead, Mebane describes vote counts as the sum of the probability a person votes multiplied by the conditional probability that they vote a certain way. In this way, vote counts meet the criteria of coming from a mixture of random distributions instead of just one (Walter R. Mebane 2006). Mebane argues that this way of modelling vote counts is aligned with how voters act behaviorally, and this way of thinking is also replicated by Pericchi and Torres' original framework [Walter R. Mebane (2006), L. R. Pericchi (n.d.)].

Like with BL, there is no theory behind 2BL and this method also employs no covariates, implying that **Something**.

Controversy of its Use

Some have questioned 2BL's usefulness in detecting election fraud. Primary ctitiques are that the law is not based in theory,

Benford's Law and the 2020 Election

"A Simple Explanation of Benford's Law." n.d. https://www.tandfonline.com/doi/epdf/10.1198/tast.2009.0005?needAccess=true.

Adhikari, A. K., and B. P. Sarkar. 1968. "Distribution of Most Significant Digit in Certain Functions Whose Arguments Are Random Variables." *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)* 30 (1/2): 47–58. https://www.jstor.org/stable/25051623.

Ballotpedia. n.d. "Arguments for and Against No-Excuse Absentee/Mail-in Voting." https://ballotpedia.org/Arguments_for_and_against_no-excuse_absentee/mail-in_voting.

Bekiempis, Victoria. 2020. "70." https://www.theguardian.com/us-news/2020/nov/10/election-trust-polling-study-republicans.

Boyle, Jeff. 1994. "An Application of Fourier Series to the Most Significant Digit Problem." *The American Mathematical Monthly* 101 (9): 879–86. https://doi.org/10.2307/2975136.

Campaign Legal Center. n.d. "Results of Lawsuits Regarding the 2020 Elections." https://campaignlegal.org/results-lawsuits-regarding-2020-elections.

CDC. 2023. "CDC Museum COVID-19 Timeline." https://www.cdc.gov/museum/timeline/covid19.html.

Deckert, Joseph, Mikhail Myagkov, and Peter C. Ordeshook. 2011. "Benford's Law and the Detection of Election Fraud." *Political Analysis* 19 (3): 245–68. https://doi.org/10.1093/pan/mpr014.

Durtschi, Cindy, William Hillison, and Carl Pacini. n.d. "Effective Use of Benford's Law."

Hicken, Allen, and Walter Mebane. 2017. "A Guide to Elections Forensics."

Hill, Theodore P. 1995. "A Statistical Derivation of the Significant-Digit Law." *Statistical Science* 10 (4): 354–63. https://doi.org/10.1214/ss/1177009869.

Horton, Joanne, Dhanya Krishna Kumar, and Anthony Wood. 2020. "Detecting Academic Fraud Using Benford Law: The Case of Professor James Hunton." *Research Policy* 49 (8): 104084. https://doi.org/10.1016/j.respol.2020.104084.

Mebane, Walter R. 2006. "Election Forensics: The Second-Digit Benford's Law Test and Recent American Presidential Elections," November.

———. 2020. "Inappropriate Applications of Benford's Law Regularities to Some Data from the 2020 Presidential Election in the United States."

———. n.d. "Election Forensics: The Meanings of Precinct Vote Counts' Second Digits."

Mebane, Walter R. 2011. "Comment on "Benford's Law and the Detection of Election Fraud"." *Political Analysis* 19 (3): 269–72. https://www.jstor.org/stable/23011437.

Pericchi, Luis Raul. n.d. "La Ley de Newcomb-Benford y sus aplicaciones al Referendum Revocatorio en Venezuela. Reporte T´ecnico no-definitivo 2a. versio´n: Octubre 01,2004."

Pericchi, Luis, and David Torres. 2011. "Quick Anomaly Detection by the Newcomb–benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela." *Statistical Science* 26 (4): 502–16. https://doi.org/10.1214/09-STS296.

Reuters. 2020. "Fact Check: Deviation from Benford's Law Does Not Prove Election Fraud." *Reuters*, November. https://www.reuters.com/article/uk-factcheck-benford-idUSKBN27Q3AI.

Scherer, Zachary. 2021. "Majority of Voters Used Nontraditional Methods to Cast Ballots in 2020." https://www.census.gov/library/stories/2021/04/what-methods-did-people-use-to-vote-in-2020-election.html.