Auditing an Election Using Sampling: The impact of bin size on the probability of detecting manipulation[1]

Jonathan Wand[2]

First version: February 2, 2004
This version: February 17, 2004

Basic notation

There are a few basic variables that can describe simple models for how an election might be corrupted (intentionally or not), benefiting one candidate at the expense of another. Assuming a two candidate election, define the following symbols:

p	the probability of candidate R receiving a vote
p^*	the probability of an observed vote for candidate R
s_{RD}	probability that a machine i will be programmed to change votes, from R to D
q	probability that a specific vote for R will be changed to D in machine i

If there is no corruption we expect $p^* = p$. If there is unidirectional tampering, the probability of observing a Republican vote is,

$$p^* = (1 - s_{RD})p + s_{RD}p(1 - q)$$

The first term represent the "true" probability of voters casting ballots for candidate R using untampered machines. The second term is for those machines that are tampered, such that the observed vote probability is reduced in expectation by $0 \leq (1 - q) \leq 1$.

This setup covers a simple case where not all votes are potentially corrupted, but rather only vote for one party; for example, if there was a single partisan group seeking to alter the election. If there is a process that alters ballots for both candidates, either by errors in the voting process or through competing groups undertaking manipulations, then we have

$$p^{**} = (1 - s)p + s_{DR}[p + (1 - p)r] + s_{RD}p(1 - q)$$

[1]This is a preliminary draft. Please do not cite without permission of the author. I thank David Dill for a stimulating conversation which led to the start of this paper, and to Walter Mebane for comments on an earlier draft. All errors are mine.

[2]Assistant Professor, Department of Political Science, Encina Hall West, Stanford University. Stanford, CA 94305. `http://wand.stanford.edu`.

where

s_{DR}	probability that a machine i will be programmed to change votes, from D to R
r_i	probability that a specific vote for D will be changed to R in machine i
s	$s_{DR} + s_{RD}$

This equation is more appropriate for the case of errors that potentially affect both candidates, since with manipulation there is the possibility that the first side to manipulate a machine (e.g., changing code) may have their changes overwritten by the other side's subsequent manipulation, and thus a machine manipulated by both sides may end up with a unidirectional bias.

In this current paper, I explore only a few simple scenarios using the unilateral corruption model and restricting the analysis to two candidates R and D.

Bin size: ballots versus aggregate units

At two extremes, one can detect manipulation in the following ways. If there is a unique *anonymous* identifier that matches a paper ballot to a machine tabulated record (but not to the voter), then any discrepancy is evidence of manipulation. If no identifier exists, one must compare the total votes for each candidate (and other categories) from machine and hand counted tabulations. (It is also possible to do tests based on sampling of subsets of these larger units, as was mentioned in an earlier note; that method is not explored here). In between comparing individual ballots and comparing totals of a precinct, one might also be able to compare totals for individual machines. As we will see, the size of the bin for comparison does matter, larger bins require more recounts in order to verify the machine tabulation.

Define bin size as U, where $U = 1$ if there are unique anonymous identifiers, otherwise $U > 1$ is the size of the unit for which a distinct set of vote totals is available.

IMPORTANT: I assume throughout that there is *no* error in the recount process when $U > 1$. Obviously when $U = 1$, any discrepancy is evidence of a problem. If there is a discrepancy with $U > 1$ then it must first be established that the recount was done reliably.

Scenarios

The following scenarios illustrate how the bin size U is affects the probability of detecting a manipulation under different scenarios about how manipulations and votes are distributed.

Results are presented in terms of the probability of not detecting the manipulation, and therefore a probabilities closer to zero are better. I present the probabilities themselves, but also I generally refer to having confidence in detection when the probability of missing the manipulation is less than 1 in 10,000. It is important to keep in mind that these results are mainly illustrative and should at most be considered relative to one another rather than in absolute terms.

In addition to analytical results, I provide numerical examples of probabilities based on a statewide election with $T = 3,000,000$ voters. Two different machine bin sizes are considered, with 100 or 200 voters per machine. Two different precinct bin sizes are considered with 500, 1000, or 5000 voters per precinct. The "machine" and "precinct" labels are merely suggestive category labels, and have no analytical content other than their particular bin sizes U.

For the bin size of $U = 1$, it is possible that among the sampled ballots that there are a number of cases where a single ballot from a precinct is selected for an audit. Due to the likely non-trivial fixed cost of auditing one or more ballots in a precinct, a cluster sampling approach is also considered where $C = 100$ and 200 ballots are randomly audited within a sampled precinct.

Total samples of size 1000, 5000, 10000, 15000, 20000, 25,000, 30000, and 50000 are considered. The number of units audited will depend on the size of the unit and the total sample size, as shown in Table 1. Obviously if it is not necessary to count all the ballots in a precinct, such as with $U = 1$ and cluster sampling, it is possible to audit more precincts for the same total number of ballots audited.

I will use as a fixed reference point an election where the true expected vote proportion for R is $p = .505$ that is changed through manipulation to an observed outcome with $p^* = .495$. This will lead to a reversal of the probable outcome, and a margin of 1 percent between R and D ($|.495 - .505| = .01$). In Washington state, mandatory recounts occur when the margin is less than .005 for machine tabulations and .0025 for hand tabulations, and thus this manipulation is large enough to successfully avoid an automatic review under the current provisions. With $T = 3,000,000$ voters, the manipulation results in an expected change of $(0.505T - 0.495T) = 0.01T = 30,000$ votes.

	$U = 1$ Cluster Sample		$U > 1$ Unit Sample				
	Precincts		Machines		Precincts		
Sample size	C =100	200	U =100	200	500	1000	5000
1000	10	5	10	5	2	1	
5000	50	25	50	25	10	5	1
10000	100	50	100	50	20	10	2
15000	150	75	150	75	30	15	3
20000	200	100	200	100	40	20	4
25000	250	125	250	125	50	25	5
30000	300	150	300	150	60	30	6
50000	500	250	500	250	100	50	10

Table 1: Number of units audited by sampling scheme

Scenario 1: iid manipulation of all machines

In this case, assume $s = 1$, such that all machines are tampered. The probability of a ballot for R being altered is assumed to be independent across ballots, with a constant probability of q; this is termed the independent and identically distributed (iid) manipulation scenario.

For $s = 1$, consider a close race with true expected vote proportion $p = .505$, and $q = .0198$ (about two in a hundred votes for R are altered on average).

The number of "true" votes M for R in a random sample of ballots of size n has a binomial distribution, $M \sim \text{Binomial}(p, n)$. Conditional on the actual number of votes for R in the sample, $M = m$, the number of manipulated ballots X is also has a binomial distribution, $X \sim \text{Binomial}(q, m)$. The probability of not having a manipulated ballot in the sample is,

$$Pr(X = 0) = \sum_{m=0}^{n} \binom{n}{m} (1-q)^m p^m (1-p)^{(n-m)}$$

Scenario 1a: with iid votes

Making the further assumption that votes for R are also independently and identically distributed across all machines and precincts (i.e., a constant p at every level of aggregation for every unit), then the size of the bin U does not matter so long as the recount number is greater than U. With a total sample of ballots of N =5,000 one will miss the tampering with a probability of less than 1 in 10,000 for all $U \leq N$. However, this iid vote assumption is an empirically unrealistic assumption.

Scenario 1b: with clustering votes

The iid votes assumption is empirically not justifiable since votes for candidates tend to cluster together within localities. A generalization of p allows for clustering of votes for each candidate within precincts, such that across precincts (indexed by i) the expected true vote probability is p_i, with an average across all units weighted by the number of votes in each, equal to p.

Consider a case where there are two types of precincts, pro-R (p_i) and pro-D ($p_j = 1 - p_i$), and each is equally common. Given a sample a total sample of size N, it is possible to sample $V = N/U$ units out of a total of $W = T/U$ units (where $T = 3,000,000$ voters). The probability of not having a manipulated ballot in the sample is,

$$Pr(X=0) = \sum_{v=0}^{V} \left\{ Pr(X_i=0)^v Pr(X_j=0)^{V-v} \right\} \frac{\binom{W/2}{v}\binom{W/2}{V-v}}{\binom{W}{v}}$$

where

$$Pr(X_i=0) = \sum_{m=0}^{U} \binom{U}{m} (1-q)^m p_i^m (1-p_i)^{(U-m)}$$

With balanced clustering, as in this scenario, the probability of missing the manipulation decreases as p_i approaches 1. Since the detection actually improves in this scenario compared to (1a), again a relatively small total sample (N = 5000) is able to detect a manipulation.

The main illustrative point is that it is relatively easy to detect a manipulation if it is applied to all machines equally, regardless of the sampling scheme. As a rule of thumb for manipulators, tampering with all machines is best avoided since it requires extra effort and increases the likelihood of detection.

Again, I emphasize that these results are for the case where there is assumed to be no random errors in the counting process itself. Stealing a few votes in each precinct may become attractive if small deviations between the vote totals and audit totals might be attributed to random errors; this is not considered in the current paper.

Sample size	$U = 100$	200	500	1000	5000
1000	0.34863	0.59045	0.80998	0.90000	
5000	0.00513	0.07163	0.34839	0.59027	0.90000
10000	0.00003	0.00511	0.12115	0.34810	0.80985
15000	0.00000	0.00036	0.04205	0.20509	0.72859
20000	0.00000	0.00003	0.01457	0.12072	0.65537
25000	0.00000	0.00000	0.00504	0.07099	0.58939
30000	0.00000	0.00000	0.00174	0.04171	0.52996
50000	0.00000	0.00000	0.00002	0.00492	0.34575

Table 2: Probability of not detecting manipulation with $U > 1$, sampling by units

Scenario 2

Consider the case where 1 in 10 machines are tampered with, $s = 1/10$. Again, assume a close race with true expected vote proportion $p = .505$, but now $q = .198$ (on average about 20 in a hundred ballots cast for R are changed in the subset of tampered machines). This larger probability of tampering within the subset of affected machines is required to produce the same change in observed outcome, $p^* = .495$. For the current scenario, this would require tampering with 1500 machines for $U = 200$, and for each machine switching on average 20 of the expected 101 ballots cast for R.

The probability of not detecting the manipulation for $U > 1$ is shown in Table 2. In this case, all ballots within a sampled unit are recounted and compared to the original tabulation. If the bin size is 500 ballots or larger, a recount of almost 50,000 ballots is required in order to have confidence that one will not miss a manipulation of the type described in this scenario.

When tabulations are done at a centralized location, identifiers give power to detect a manipulation even with relatively small samples, by sampling a subset of ballots within a randomly selected unit. As noted before, by clustering of samples by by a unit, it is possible to improve power of detection without the cost of counting every ballot within a unit. Table 2 the probability of not detecting the manipulation for $U = 1$ for different numbers of ballots randomly recounted C within sampled units of size U'. The number of units sampled is $V' = N/C$. The probability of detection are not surprisingly very similar to those for equivalent sized bins, $U = C$. Most of the power of detection is coming from the larger number of units of size U' which are having a random subset of ballots audited, and the

	$U' = 500$		$U' = 1000$		$U' = 5000$	
Sample size	$C = 100$	200	100	200	100	200
1000	0.34840	0.59038	0.34811	0.59027	0.34576	0.58939
5000	0.00504	0.07139	0.00492	0.07099	0.00405	0.06780
10000	0.00002	0.00504	0.00002	0.00492	0.00001	0.00405
15000	0.00000	0.00035	0.00000	0.00033	0.00000	0.00021
20000	0.00000	0.00002	0.00000	0.00002	0.00000	0.00001
25000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
30000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
50000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Table 3: Probability of not detecting manipulation with $U = 1$, cluster sampling ballots by unit of size U'

assumption that manipulations iid within and across tampered machines. Relatedly, there is a increasing advantage of sampling within larger units U' as this also enables a larger number of units to be audited.

Further comments

Having centralized tabulations of votes leads to very difficult and costly auditing requirements due to the large sample size required for verifying the vote totals. With individual identifiers enabling an audit to link an electronic record with a hand checked vote record, this problem is removed, and indeed using a sample within units of size C is better in terms of efficiency than a situation which relies on verifying equivalent sized small units, $1 < U \leq 200$.

With the introduction of measurement error, the advantages of linking a particular electronic record with a hand verified paper record becomes even more important for identifying the source and of observed discrepancies.

What should be done if a manipulation is believed to have been observed? In unambiguous cases, a full recount investigation is likely warranted. If there is some ambiguity, such as only a small fraction of discrepancies in a small number of units, then the sampling should be expanded sequentially to gain further evidence to resolve the ambiguity; this sequentially process is not covered here.

IMPORTANT: The main interest is in the relative size of the probabilities, rather than in the magnitudes under any given scenario or sampling scheme. These scenarios and calculations are solely illustrative of some scenarios under very restrictive assumptions, and excluding many other important considerations. The actual sample size and sampling design needed for detecting manipulation in a particular state or locality depends on a variety of features particular to it's elections and its administrative setup, only some of which is captured in these idealized examples.