Myth and Reason in Designing Election Audits

By
Jerry Lobdill
September 10, 2006

There are at least two myths associated with conventional approaches to designing election audits. One is that laws mandating election audits ought to specify a specific percentage of the vote count to audit. I address this myth in a companion paper. This paper discusses the myth that all precincts in which a race appears on the ballot are equally likely targets for attack.

A useful metaphor for this discussion is the concept of objects. An "object" is a computer-programming concept. Objects of a certain class have certain properties in common, and may have other properties as well that distinguish class members.

For example, consider marbles as objects. They are spherical; they have a radius; they have a density, a color scheme, a surface texture. These are properties of marbles. Properties of objects become important in devising algorithms to sort objects or to otherwise deal with them.

One classic problem that illustrates objects and how their properties are considered is the "marbles in an urn" (MIAU) problem. In this problem marbles of two colors are contained in an urn. There are a certain number, n, of one color, and m marbles of another color in the urn. It is assumed that for the marbles in the urn all other marble properties are identical. If a sample of s marbles is drawn at random from the urn, what is the probability that there will be at least one marble of a specified color in the sample?

The famous French mathematician, Siméon-Denis Poisson, first published the solution to this problem some 170 years ago.

Let's explore object properties and how they might affect the analysis of a problem such as the MIAU problem. Suppose there are three colors of identically sized marbles. And suppose that there is a black dot on r of the marbles in the urn. It is known, let's say, that a large percentage of the black dots will be found on blue marbles, and the rest of the black dots are distributed evenly between the yellow marbles and the white marbles. The problem now is to devise a sample selection process by which at least one marble with a black dot is found and the number, s, selected is minimized. Suppose it is permitted to sort the marbles by color and then draw random samples from each color. Would it be best to ignore the colors and calculate the probability of finding a marble with a black dot on it in the total marble population of marbles in the urn? Or would it be best to first sort the marbles by color, and then sample? We'll return to these questions later.

The standard version of the MIAU problem specifies only two colors of marbles. This is the version that is normally used to discuss the use of Poisson's formula. The lore about this problem is so well known that the discussion usually neglects to emphasize that the essential assumption that makes the Poisson formula apply in the way described is that the marbles have only one distinguishing property, color. When there are only two colors, say, black and white, and no other distinguishing properties of the marbles, the Poisson formula gives the number of marbles, s, that must be drawn at random from the urn to ensure that the probability of drawing at least one black marble is some specified number.

If we used the standard MIAU model to calculate the sample size, s, in the problem where we have three colors of marbles we would be ignoring known properties of the marble population. Perhaps it would be wise to isolate the blue marbles, since we know they will have the largest percentage of black dots. Suppose there are 10 blue marbles and 100 each of yellow and white marbles. Just for an example let's say we know that if there are any black dots on any marbles in the urn, then on average 80% of the marbles with black dots will be blue marbles with 20% being evenly distributed between the yellow and white marbles.

Suppose we want to be 90% sure that, if there is even just one blue marble with a black dot, we won't miss it in our sample. Then, by the Poisson formula, we must draw nine blue marbles at random and examine them. Since it is no big cost to examine all 10 blue marbles instead of 9, we would choose to look at all 10 of the blue marbles.

Then, suppose we didn't find a black dot in that sample. In that case the probability that there are any yellow or white marbles with a black dot is small.[1] The sampling plan for the remaining 200 marbles will depend on our assessment of the penalty we may pay for failure to find a marble with a black dot among the remaining 200 marbles.

We might ask, if there are actually 10 marbles with black dots what is the probability that none of them are on blue marbles? This is the same as the probability that if we have an urn with, say, 10,000 marbles, 8,000 blue ones and 2,000 white ones, and we draw 10 marbles at random from the urn, there are no blue marbles in the sample. This probability is $(0.2)^{10} = 1x10^{-7}$. So, if we don't find any blue marbles with black dots the probability that there are any other marbles with black dots is vanishingly small assuming there are either 10 marbles with black dots in the urn or none. If there are actually more than 10 marbles with black dots the probability that there are no blue marbles with black dots is even smaller than $1x10^{-7}$. Under these conditions, if we find no blue marbles with black dots there is no need to sample the remaining marbles at all.

One may think that this discussion is leading up to a similar conclusion with respect to election audits. Not so. Read on.

Objects in Election Audits

Now, let's consider precinct vote count files (PVCF) for a race as objects. Assuming a race with two candidates, we have a number of pertinent properties for PCVFs. Some or all of the following precinct properties exist and are available at the time an audit must be designed:

- Number of registered Democrats
- Number of registered Republicans
- Number of registered voters
- Vote count for Candidate A (Republican)
- Vote count for Candidate B (Democrat)
- Number of undervotes
- Number of early votes
- Number of mail (absentee or senior) votes

The most important properties are probably the two vote counts for each precinct.

For the complete district of the office in question there are N precincts. The purpose of the audit is to detect corruption of the vote that could produce a reversal in the outcome of the election, thereby triggering a complete recount of the vote across the election district.

If we are to make intelligent use of the properties of PVCFs we need to consider the probable threats that could produce the total vote margin observed and how the threats would affect the properties and metrics that can be calculated from the properties.

The most complete threat analysis available at this writing is contained in the Brennan Report of June 28, 2006.[2] A discussion of the threats was given in a previous paper.[3] There it was concluded that to maximize the chances of altering the outcome of a large district race a wholesale[4] attack would be needed. The Trojan Horse software needed would have protections that would detect when the equipment is being tested, when it is being used in the election, and include an algorithm to monitor the number of votes cast and an algorithm to switch votes for the race in question after the close of the polls. This Trojan Horse software would operate automatically without need for conspirators at polling places to intervene by any of the possible methods described in Brennan. The vote-switching algorithm could (and probably would) include controls that would prescribe the percent of the vote count that would be switched, as a cautionary measure to attempt to avoid suspicious results at any polling place. Howard Stanislevic of VoteTrustUSA has suggested that, based on an analysis of vote shifts from the 2000 to the 2004 election in the 1250

[1] This example assumes that there will either be no black dots at all or at least 10 marbles with black dots.
[2] The Machinery of Democracy: Protecting Elections in an Electronic World, by Brennan Center Task Force on Voting System Security, Lawrence Norden, Chair, June 28 2006
[3] "Designing Mandatory Election Audits", by Jerry J. Lobdill, August 20, 2006
[4] Brennan Report terminology- indicates that all electronic voting equipments have the same Trojan Horse software installed prior to the election.

national exit poll precincts, 20% of the total vote per precinct should be the assumed upper limit of votes switched.[5] The other essential feature of the attack is that it would be designed to cause the reversal using as few precincts as possible (given the lower limit on percent of the vote count required to permit vote switching) so that the number of corrupt PVCFs is minimized. The simplest way to accomplish this is to set a vote count threshold above which votes would be switched.

Implications of the Assumed Attack on Audit Design

We need now to discuss the implications of this threat. This threat is defined so that each precinct has its own instance of the Trojan Horse software which operates independently based on the information generated by voting activity at that precinct. The more voters there are above the threshold the larger will be the number of votes switched. This rule implies that the largest vote count precincts will definitely be attacked, and the attack will target every precinct whose vote count is above a certain threshold. So we know that if this wholesale attack is in progress the precincts are attacked not at random but according to vote count size.

The attacker is limited in her control over the countywide margin actually produced by the Trojan Horse because the total vote count and distribution of votes by precinct is not known in advance. The attacker cannot afford to randomize the attacks on the precincts, choosing to attack some but not others without regard to the vote count, because this plan would risk a loss for the desired winner.

Possibilities For Other Attacks

The attack postulated here takes place at each precinct. The same result could possibly be obtained by attacking the central tabulating software with a Trojan Horse in every county that has precincts within the race jurisdiction. In this case, at close of polls, the MBB cards from the precincts would be read and the vote switching would be done in the central tabulator in each county. In this case the attacker's Trojan Horse has additional information available. The actual true vote counts are available for all of the county's precincts involved in the race. Switching votes at this level could be done in a more sophisticated way that might make the attack more difficult to detect. Perhaps the data on the MBB cards could be altered as well as changing the outcome of the countywide vote count.

Other attack strategies could be rationalized. The attacker might throw caution to the winds with regard to a switching percentage on grounds that she could attack fewer precincts and decrease her probability of being exposed in this way. That is a trade-off between the possibility of creating a vote margin that someone in the know might think is suspicious and the possibility of getting caught in a randomized audit plan because too many precincts had to be altered to achieve the desired result.

I do not see any sensible way for the attacker to make precincts equally likely to be attacked and have any confidence in the desired election outcome.

This analysis begs the question, "Given that the attack is going to involve methodical attacks on precincts does it make sense to ignore this knowledge in planning an audit?" If the answer comes back, "We must not have any preconceptions," then we must ask how this conclusion has been reached. Is it because we can't know exactly what the attacker will do? Is there any inclination to be politically correct and avoid any sort of accusation that political profiling has guided the audit plan? If there are good technical reasons to treat all precincts as equally probable targets of attack even though we know that is not true, should we recommend to the IRS that they select tax returns to audit on this same basis? Should we advise a bettor at the Kentucky Derby to pick his winner using the theory that all horses have an equal chance to win?

[5] Private communication, and http://www.votetrustusa.org/pdfs/VTTF/EVEPAuditing.pdf