CAST: Canvass Audit by Sampling and Testing

Philip B. Stark
Department of Statistics
University of California
Berkeley, CA 94720-3860

DRAFT 29 August 2008, 11:00am

Abstract

An election audit trail can reflect an electoral outcome that differs from the semi-official outcome. If so, the winner should be determined by a full manual count of the votes as recorded in the audit trail. CAST is a statistical method for deciding whether to certify the outcome of a race, or to count the entire audit trail by hand. It has a known, pre-specified chance of requiring a full hand count whenever that count would show an outcome different from the semi-official outcome. This limits the risk of certifying an incorrect outcome erroneously. CAST also allows some ballots to be selected for audit deliberately rather than randomly—"targeted" auditing. CAST requires:

(i) the desired minimum chance that if the preliminary outcome disagrees with the audit trail, CAST will require a full manual count; (ii) the maximum number of audit stages permitted before a full manual count; (iii) the semi-official results by precinct or "batch;" (iv) limits on the number of valid votes for any candidate that could have been cast in each batch. CAST collects data incrementally: if the the semi-official results overstate the margin by a sufficiently small amount in a sufficiently large sample, CAST says "certify." Otherwise, CAST says "audit more batches and check again." Eventually, either CAST says "certify," or there has been a full hand count. CAST is a refinement and simplification of the method iof Stark [1, 2].

1 Introduction

CAST, Canvass Audit by Sampling and Testing, is a refinement and simplification a method introduced by [1] for determining—by auditing a random sample of batches of ballots—whether to certify the semi-official outcome of an election. CAST has a large chance of requiring a full hand count whenever that count would show a different winner than the semi-official outcome.[1] When a full hand count would not change the outcome, CAST

[1] Of couse, even a full manual count might not identify the rightful winner. For example, the audit record could be incomplete or inaccurate: paper ballots can be lost before they are counted, ballot boxes can be "stuffed," etc. Moreover, it is impossible to audit direct recording electronic (DRE) voting machines that do not produce voter verifiable paper audit trails (VVPATs) without relying on electronic records. And agreement between

tries to minimize the amount of auditing effort before declaring "certify."

CAST is performed in stages. At each stage, the semi-official vote counts in a random sample of batches of votes[2] are compared with hand counts of the paper audit trail for those batches, for each race under audit. Additional batches may be selected for audit deliberately rather than randomly, either before the random selection begins, or between stages of random selection.[3] For each audited batch, errors that resulted in overstating the margin between any winner and any loser in a given race are expressed as a percentage of the semi-official margin between those candidates, adjusted for any batches that were hand counted at a previous stage. A calculation is performed to answer the question, "If the semi-official outcome disagrees with what a full manual count would show, what is the minimum chance that in at least one batch selected in the current stage, the error would have overstated the margin by more than it in fact did?" If that chance is large enough, the audit stops and the semi-official result is certified: it is likely that the audit would have uncovered more error if the outcome were wrong. If that chance is not large, the semi-official margin is adjusted to take into account the net error

hand counts of audit batches and semi-official results does not ensure that the batch-level results were aggregated correctly across the race; viz. the recent Diebold/Premier software "glitch."

[2]A batch could comprise all the votes cast in a precinct, the votes cast on a particular machine, or other convenient group for which a semi-official subtotal is available.

[3]The vote counts in any batches already audited, whether at random or deliberately, are treated as known perfectly. The reported margin is adjusted to reflect the manual counts in those batches, and those batches are excluded from future samples.

observed in the audit sample, a new sample is drawn from the batches not yet audited, and the calculation is performed again using the adjusted margin. After a pre-determined number of steps, either the semi-official outcome has been certified, or there has been a complete manual count of the audit trail, and the "correct" outcome is known.

If the thresholds for "large enough" chances at each stage are set correctly, CAST guarantees that the chance of a full manual count is large whenever a full manual count would contradict the semi-official outcome.[4] That controls

[4]Suppose we want an overall chance β (e.g., $\beta = 90\%$) that there will be a complete manual count if the semi-official results are wrong. We contemplate performing at most S audit stages before conducting a full manual count—if the outcome has not been certified by one of those stages. Suppose that at each stage, we test in such a way that, if the semi-official outcome is incorrect, the chance that the audit goes on to the next stage is at least β_0, given the results observed at all previous stages. Then the chance that the audit progresses to a full manual count is β_0^S. So, if we pick $\beta_0 = \beta^{1/S}$, the Sth root of β, the chance that the audit will lead to a complete manual count when the outcome is incorrect is at least β. For example, suppose we want a 90% chance ($\beta = 90\%$) of conducting a full manual count whenever it would show that the semi-official outcome is wrong, and we contemplate drawing an initial sample, a second sample if necessary, and then proceeding to a full manual count if we have not yet certified the outcome. Then $S = 2$ and $\beta_0 = \sqrt{90\%} = 94.9\%$.

This idea can be generalized: let β_s be the minimum conditional probability that the outcome is not certified at stage s given everything the audit has uncovered up to stage s, assuming that the outcome is in fact wrong. Then, provided $\prod_{s=1}^{S} \beta_s \geq \beta$, the chance of conducting a full count when it would show a different outcome is at least β. Taking β_s smaller for early audit stages and larger for later stages may reduce the audit burden when the semi-official outcome is correct.

the risk of erroneously certifying and incorrect outcome: Let β denote the minimum chance of a full manual count when a full manual count would show a different outcome. Then if an outcome is incorrect, the chance that it will be certified anyway is at most $(100\% - \beta)$.

2 CAST Step-by-step

This section gives the steps in CAST. It assumes that semi-official results are available for all batches of ballots before the audit starts. Batches can be grouped into *strata*[5] for convenience, with random samples of batches drawn independently from different strata.[6] For example, one stratum might comprise ballots cast in-precinct on election day; another might comprise vote-by-mail (VBM) ballots; and a third might comprise provisional ballots. Batches might be stratified by county, if a race crosses county lines.

Batches should never be selected from a stratum before the semi-official counts have been announced for all batches in that stratum. To do otherwise is to invite fraud.

The same argument shows that it is fine to allow targeted auditing before and between stages of random audit, provided that at stage s the test that is used guarantees that the conditional probability is at least β_s that the audit will not certify the outcome if the outcome is wrong, given everything the targeted and random audit has discovered prior to stage s.

[5]Strata must be *mutually exclusive* (no ballot can be in more than one stratum) and *exhaustive* (every ballot must be in some stratum).

[6][1] gives two methods for dealing with stratification in the probability calculations. Many others are possible; current research seeks optimal approachs.

Some batches might be selected for audit deliberately rather than randomly. For example, laws and regulations might permit the candidates in a race each to select for audit a few batches they find suspicious; this is sometimes called a *targeted audit* in distinction to a *random audit*. There can be rules such as "if the targeted audit of a batch reveals a discrepancy of more than $X\%$ of the margin, perform a full manual count." A rule of that form can only increase the chance of a full manual count when such a count would show a different outcome from the semi-official outcome, so it can be used with CAST.

2.1 Step 1: pick the chance β of catching an incorrect semi-official result

The first step is to pick the minimum chance of a full manual count when the semi-official outcome differs from the outcome a full manual count would show. Typically, this is a matter of legislation or administrative rule. To limit auditing burden, it can be desirable to choose β smaller for small races than for countywide or statewide races.[7]

2.2 Step 2: pick the maximum number of stages S

The second step is to pick the maximum number of sampling stages before a full manual count. If the audit cannot confirm the semi-official outcome

[7] If β is large, it will often be necessary to hand count most of the ballots in small races, even when the semi-official outcome is correct.

after the first stage of sampling, a second sample will be drawn, and so on. If the audit cannot confirm the semi-official outcome at stage S (because the audit has found too much error at every stage), there will be a full manual count.

2.3 Step 3: select the subtotals that comprise batches and the strata

The third step is to define the batches of votes from which the audit sample will be drawn, and to partition those batches into strata for convenience. Generally, the fewer votes each auditable batch contains, the smaller the audit effort required to confirm the semi-official outcome if that outcome is correct. Batches must satisfy two requirements, though: (i) semi-official counts for each batch in a stratum must be published prior to drawing the sample from that stratum. (So, if a jurisdiction does not publish semi-official totals by machine within a precinct, but only for precincts as a whole, batches must comprise at least entire precincts.) (ii) There must be an upper bound on the number of valid votes in the batch for any candidate or position in the race. (Generally, such bounds are available only at the precinct level and above; see section 2.4.)

The strata must be defined so that every batch is in one and only one stratum. Generally, strata should not cross jurisdictional boundaries, so that a single jurisdiction can carry out the audit of all the batches selected from a given stratum. Let P denote the total number of batches of ballots across all strata; let C denote the total number of strata; and let B_c denote the number of batches in stratum c, for $c = 1, \ldots, C$. Then

$$B_1 + B_2 + \cdots + B_C = P. \tag{1}$$

2.4 Step 4: find upper bounds on the number of votes per candidate per batch

To establish a limit on the extent to which error in each batch could possibly overstate the semi-official margin (see section 2.7), one needs to know the maximum number of votes each candidate or position could possibly get, batch by batch. Upper bounds on the number of votes a candidate or position could get in any precinct can be derived from voter registrations, pollbooks, or an accounting of ballots.[8] For example, if an accounting of ballots confirms that b_p ballots were voted in precinct p, then any candidate or position could receive at most b_p votes in that precinct. See [1].

2.5 Step 5: set initial values of the variables

Set $s = 1$; s represents the current stage of the audit. Let P_s be the number of as-yet-unaudited batches at stage s. If there has been no targeted auditing

[8] An accounting of ballots compares the number of ballots sent to a precinct with the number returned voted, unvoted and spoiled to account for each piece of paper. An accounting of ballots is generally impossible for DRE voting machines. An accounting of ballots should be performed whenever possible, to ensure that ballots have neither disappeared nor materialized.

so far, then $P_1 = P$.

2.6 Step 6: calculate all pairwise margins

The sixth step is to find the all margins of victory for the race, according to the semi-official results and what the audit has discovered so far. Let P_s denote the number of batches that have not yet been audited at stage s.

For each semi-official winner w and each semi-official loser ℓ in the race, calculate the margin of victory in votes:

$$V_{w\ell} = \text{ (votes for winner } w) \; - \; \text{(votes for loser } \ell). \tag{2}$$

In this calculation, use the semi-official results for the P_s batches that have not yet been audited, and the audit results for the $P - P_s$ batches that have already been counted by hand.

If any of these margins $V_{w\ell}$ is now zero or negative, the audit has already found so much error that the list of winners has changed. If that occurs, abort the audit and do a full manual count.

In a winner-take-all race with K candidates in all, there are $K - 1$ such margins of victory: the apparent winner is paired with each of the remaining candidates in turn. Some races, such as city council races, can have several winners. For example, suppose a race allows each voter to vote for up to 3 of 7 candidates, and the three candidates receiving the largest number of votes are the winners. Then the 3 semi-official winners each have a margin of victory over each of the 4 semi-official losers, and there are $3 \times 4 = 12$ margins of victory $V_{w\ell}$.

2.7 Step 7: find upper bounds on the maximum overstatement of pairwise margins

In each batch not yet audited, the semi-official counts in each batch, together with the bounds on the number of valid votes per candidate in each batch (section 2.4), limit the amount that error in a given batch could have overstated the margin between any semi-official winner and any semi-official loser.

Let candidate w be one of the semi-official winners of the race and let candidate ℓ be one of the semi-official losers of the race, as above. For each batch p that has not been audited, compute

$$u_{pw\ell} = \frac{(\text{votes for candidate } w \text{ in batch } p) - (\text{votes for candidate } \ell \text{ in batch } p) + b_p}{V_{w\ell}}, \tag{3}$$

according to the semi-official results for batch p. Compute the largest value of $u_{pw\ell}$ for all pairs (w, ℓ) of semi-official winners and losers. Call that number u_p. Then u_p is the most by which error in counting the votes in batch p could have overstated the margin between any apparent winner and any apparent loser, expressed as a fraction of the margin of victory between those two candidates, adjusted for any errors that were discovered in batches already audited. See [2].

2.8 Step 8: targeted audits

At this point, if there are a few un-audited batches p for which u_p is much larger than the rest, auditing those batches deliberately can reduce substantially the sample size required in the random audit to follow. If additional batches are selected for targeted audit, count them and return to step 6.

2.9 Step 9: select the desired threshold for "escalation"

The next step is to set the tolerable level of error, t, a number between 0 and 1. If any margin is overstated by t or more, the audit will progress to the next stage. Generally, the larger the value of t, the larger the sample size will need to be. If t is chosen so large that the sum over all as-yet-unaudited batches of the smaller of u_p and t is 1 or larger, a full manual count will be required to confirm the election. The smaller the value of t, the smaller the sample size will need to be at the first stage, but when t is small there is also generally a greater chance that the audit will progress to the next stage. For example, if $t = 0$, the audit will have to go to the second stage if the first sample finds even one discrepancy that overstates any margin. It is permissible to change the value of t from stage to stage of the audit, provided the sample size for stage s is calculated using the value of t for stage s.

2.10 Step 10: find sample sizes for the next random sample

The next step is to calculate the incremental number of batches to be selected at random from each stratum. First the total additional sample size is calculated; then that number is allocated across strata in proportion to the number of batches in each stratum. Other choices are possible; see [1]. We shall assume that the semi-official counts are available for all batches in all strata.[9]

Define $\beta_0 = \beta^{1/S}$, the S^{th} root of the probability of a full manual count when that would change the outcome. To find the overall sample size, we define a new list of numbers. Let t_p be the smaller of t and u_p, for the P_s batches p not yet audited. and let $\tilde{u}_p = u_p - t_p$ for those batches. Let T be the sum of $\tilde{u}_p$ for all as-yet-unaudited batches p.

1. Starting with the largest value of $\tilde{u}_p$, add successively smaller values of $\tilde{u}_p$ just until the sum of those values is $1 - T$ or greater. Let q denote the number of terms in the sum.

2. Find the smallest value n such that

$$\frac{(P_s - n)!(P_s - q)!}{P_s!(P_s - q - n)!} \leq 1 - \beta_0. \tag{4}$$

3. For $c = 1, \ldots, C$, the sample size n_c for stratum c is

$$n \times \frac{\#\text{unaudited batches in stratum } c}{P_s}, \tag{5}$$

rounded up to the nearest whole number. Thus,

$$n^* = n_1 + n_2 + \cdots + n_C \geq n. \tag{6}$$

[9] If not, it is possible to proceed with some reasonable but arbitrary initial choice. However, if the initial sample size is too small, the audit will proceed to the next stage even if the first stage finds no error whatsoever. Step 13 then needs to be modified. See [1] for more discussion.

[1] proves that these sample sizes guarantee that the chance of certifying erroneously at stage s is at most $1 - \beta_0$.

2.11 Step 11: draw the next sample and count votes

The next step is to draw the samples of batches for audit. Select batches using a transparent, mechanical, verifiable source of randomness, such as fair 10-sided dice. Computer-generated "pseudo-random" numbers are not appropriate, because it is essentially impossible for the public to verify whether the algorithm is fair or has been tampered with. For each stratum $c = 1, \ldots, C$, draw a random sample of n_c batches from the as-yet-unaudited batches in stratum c, and count the votes for each candidate in each batch by hand.

2.12 Step 12: calculate the maximum pairwise overstatement

For each of the n^* batches p just audited in this stage, calculate the largest value of

$$
\begin{aligned}
e_{pw\ell} \;=\; & \frac{(\text{reported votes for } w \text{ in batch } p) - (\text{reported votes for } \ell \text{ in batch } p)}{V_{w\ell}} \\
& - \frac{(\text{audited votes for } w \text{ in batch } p) - (\text{audited votes for } \ell \text{ in batch } p)}{V_{w\ell}}
\end{aligned}
$$

for all pairs (w, ℓ) of semi-official winners w and losers ℓ. Then find the largest value of those n^* values for all batches just audited. Call that value t_s.

2.13 Step 13: certify, perform a full count, or proceed to the next step

If $t_s < t$ certify the election and stop.[10] If not, and we are at stage S, perform a full manual count. Otherwise, add one to s; perform any additional desired targeted auditing; set P_s to be be the number of batches not yet audited; and return to step 6.

3 Technical notes

CAST differs from the method of [1, 2] in important ways. Instead of using Bonferroni's inequality to bound the overall chance of certifying an incorrect outcome, CAST controls the probability by designing the test at each stage to control the conditional probability of certifying erroneously.

Second, CAST conditions on the audit results at previous stages, rather than treating the sample at stage s as a "telescoping" sample that includes previous stages as in [1]. This has a number of benefits. First, if an early stage of audit finds a large discrepancy, the test statistic at later stages is not necessarily large, because at each stage only the incremental sample enters the test statistic. Second, margin overstatement errors discovered at one stage do not lead to more than one step of escalation if there are canceling margin understatement errors: while only overstatement errors are involved

[10] This can be refined slightly; it can be the case that $t_s > t$ but that the chance of observing a pairwise margin overstatement even larger is at least β_0 if the outcome is wrong.

in the test statistic at a given stage, the margin is adjusted sequentially to account for both overstatement and understatement errors. Third, this makes it easy to incorporate "targeted" sampling.

References

[1] P.B. Stark, *Conservative statistical post-election audits*, Ann. Appl. Stat. **2** (2008), 550–581.

[2] ______, *A sharper discrepancy measure for post-election audits*, Ann. Appl. Stat. **in press** (2008), in press.