

Statistical validation of EVMs via VVPAT

Rajeeva L. Karandikar

PROFESSOR, CHENNAI MATHEMATICAL INSTITUTE

rlk@cmi.ac.in, rkarandikar@gmail.com

Abstract

For over 20 years now, all Parliament and State assembly elections in India have been conducted using Electronic Voting Machines (EVMs). There have been questions raised on trustworthiness of EVMs. For the 2019 Parliamentary elections, the VVPAT (Voter Verifiable Paper Audit Trail) system was incorporated in the polling process on the basis of Supreme Court (SCI) directions. Based on commitments made by the elections commission (ECI) to the SCI, on the day of counting, from each of the 4025 assembly segments of the 543 Lok Sabha constituencies, five booths were chosen randomly and the VVPAT votes in these 20125 booths were counted manually and found to match with the corresponding EVM count. We will argue as to how the data based on VVPAT matching done in 2019 Parliamentary elections provides a strong statistical proof that EVMs are highly reliable.

1 History of EVM's and VVAPT

Till 2001, all the elections to Lok Sabha and Vidhan Sabha were conducted using paper ballot - ballot box. From newspaper reports and Television coverage, it can be seen that there were multiple instances of election fraud- called *booth capturing* which would mean that the ballot box would be stuffed by large number of paper ballots marked in favour of a specific candidate.

The elections Commission of India (ECI) along with teams of experts from IITs and Electronics Corporation of India (ECIL) developed the Electronic Voting Machines (EVMs) in India, beginning with 1977. The ECI worked on improving EVMs. Since The Representation of the People Act 1951 explicitly mentioned paper ballot ECI worked with government to have the Act amended (in 1989).

In 1999, the ECI made public its updated EVMs and it was used in Goa assembly poll in 1999. It was also used in 46 constituencies out of 543 in the 1999 Parliamentary elections. Based on the experience, EVMs have been used in all the constituencies in assembly elections since May 2001 and in all parliamentary elections since 2004.

It put an end to booth capturing, invalid votes and printing of large number of ballot papers and was widely welcomed.

After EVMs completely replaced paper ballot multiple cases have been filed under various grounds and have been dismissed by the courts.

In one of the petitions, it was proposed that ECI should incorporate a safeguard in the EVMs called “paper backup”, “paper receipt” or “paper trail” which would easily and cheaply meet the requirement of proof that the EVM has rightly registered the vote cast by a voter. The appellant has further highlighted that the “paper trail” system is to supplement the procedure of voting as in this procedure, after recording a vote in the EVM, a print out will come out which will appraise the voter that his vote has been rightly registered and the same will be deposited in a box which can only be used by the ECI in case of election dispute.

The Supreme Court (SC) was informed that ECI had been developing a VVPAT (Voter Verifiable Paper Audit Trail) system and conducting trials of the same with technical expert committee. In 2013, the technical committee approved a design of VVPAT and were prepared to introduce the same. In response to this, the SC said that *we are satisfied that the “paper trail” is an indispensable requirement of free and fair elections. The confidence of the voters in the EVMs can be achieved only with the introduction of the “paper trail”. EVMs with VVPAT system ensure the accuracy of the voting system. With an intent to have fullest transparency in the system and to restore the confidence of the voters, it is necessary to set up EVMs with VVPAT system because vote is nothing but an act of expression which has immense importance in democratic system.* The SC permitted the ECI to implement VVPAT in a phased manner, since ECI would require over 10 lakh VVPATs.

This involved huge expenditure and the central government allotted the required budget for procuring VVPAT machines, so that 2019 Lok Sabha election could be conducted using EVM-VVPAT. The total money spent on EVM-VVPAT before the 2019 parliamentary election was over Rs. 6,600 crores.

2 Important features of EVM

The ECI has published several manuals for use by various officials who are called for election duty and also for public. See:

<https://eci.gov.in/evm/>

<https://eci.gov.in/manuals/>

Here are some of the highlights.

- (A) The EVMs do not have any networking capability- no wired internet, no WiFi and no Bluetooth or any other wireless connectivity whatsoever. On the day of counting, a button is pressed and then the total votes for each candidate are displayed on the EVM and need to be copied by hand.
- (B) All ECI EVM units have a secure microcontroller where the main program of the machine is stored and the microcontroller is then One Time Programmed during the manufacturing of the machines. Thereafter the program cannot be modified at all.
- (C) The design of EVMs ensures that it does not permit more than 4 votes in a minute. This protects it from any large scale *booth capturing*

(D) The names of candidates appear on the ballots screen of ballot unit of EVM following the same scheme that existed when they were using paper ballot. This means that the names are divided in 3 groups: with first group consisting of candidates of recognized political parties, second group consisting of candidates of registered parties and the third group consisting of the independent candidates. Within each group, the candidates are listed in the alphabetical order of their names : **NOT in the order of the party name**. Thus the order in which names appear would be decided only after the last date of withdrawal of candidates, by when the EVMs would already be distributed to the constituencies.

These points mentioned above and the information available at ECI published manuals makes it abundantly clear that if we accept that it is not possible to tamper an EVM remotely, **then it is definitely more secure than the old style paper ballot and ballot boxes it replaced**.

So the crux is then the question : *Is it possible to tamper the EVM remotely when all the steps described above are followed.*

3 Strength and Weakness of EVM

The ECI had announced in 2018 that one booth in each assembly segment will be chosen randomly on the counting day, and the VVPAT count of the chosen booth will be matched with the EVM count. This was proposed by ECI as a confidence building measure. There were 4025 assembly segments in 2019 elections. This was challenged in the courts and demands were made to match 50% of VVPATs with the corresponding EVMs.

In 2018, The ECI constituted a committee of 3 members, including me, and sought our opinion that assuming that all the 4025 randomly chosen EVM count and the corresponding VVPAT count match, can we statistically conclude that EVM design and implementation is good, in other words, can we take it as statistical validation of EVMs.

4 Statistical Hypothesis Tests

Suppose that indeed someone does manage to find a way of tampering with the internal memory of EVM where the voting outcome is stored remotely. Then it is safe to assume that the person(s) who can do so will not stop at one or few EVMs but would try to tamper several EVMs so as to modify the outcome of the National election. It can be seen that to do so, one would need to tamper a large number of EVMs. Thus the entity that wishes to tamper with outcome of the national elections would need to tamper with EVMS in, say at least 25 constituencies in such a way to change the outcome, which in turn would need tampering with say 20the EVMs in these constituencies. Of course the tampering itself would have to be done carefully for if it is found that a very large proportion of votes go to one candidate which is not expected, it will raise alarm and if this happens in several booths, the losing candidate can seek recounting of votes (using VVPAT) in the entire constituency.

Based on loose calculations, we concluded that the entity that wishes to tamper with the outcome of the national elections would need to tamper say 2% or more of the EVMs in the country. Of course if we see a mismatch between EVM count and VVPAT it could be due to mechanical defect or may have been tampered with. Let us denote by θ the proportion of EVMs that are defective (mechanical defect or have been tampered and manipulated). We need to explore if it is possible for a fraudster to tamper EVMs in $\theta = 1\%$ (or more) of the booths.

Of course the fraudster could (and would) choose the booths so as to minimise probability of detection. So we should explore the probability that indeed $\theta = 1\%$ or more EVMs are defective and yet a random sample of 1 booth from each of the assembly segments does not detect any tampering. How likely or unlikely is this?

We consider

Null Hypothesis $H_0: \theta \geq 1\%$

and

Alternate Hypothesis $H_1: \theta = 0$

and take level of significance of the test to be very very low $\alpha = 0.00006334248$

(this is the probability of observing a deviation of 4σ from a normal distribution).

H_0 is a composite hypothesis where we assume that the fraudster has chosen booths to tamper that minimises the probability of being caught (by mismatch of EVM-VVPAT).

Let the assembly segments be numbered $A_i, i = 1, 2, \dots, 4025$, with A_i having n_i booths, lets number them as $b_{ij}, 1 \leq j \leq n_i, i = 1, 2, \dots, 4025$. Let

$$a_i = \sum_{j=1}^{n_i} b_{ij}$$

be the total number of booths where there is possible mismatch between EVM and VVPAT count, due to malfunction of result of fraud or manipulation.

Under H_0 , mismatch, if any is more than 1% *i.e.*

$$\sum_{i=1}^{4025} \sum_{j=1}^{n_i} b_{ij} = \sum_{i=1}^{4025} a_i \geq \frac{1}{100} \sum_{i=1}^{4025} n_i.$$

Let $X_i = 0$ if at the randomly chosen EVM in A_i , the EVM count matches the VVPAT count and $X_i = 1$ if there is mismatch. Let $Z = \sum_{i=1}^{4025} X_i$ denote the total mismatches. We take

$$\{Z = 0\}$$

as the critical region.

Since H_0 is composite, we need to see if

$$P(Z = 0) \leq \alpha = 0.00006334248$$

for all possible choices of b_{ij} , $1 \leq j \leq n_i$, $i = 1, 2, \dots, 4025$ satisfying (4).

Indeed, this is true even if we take Null Hypothesis $H_0: \theta \geq 0.5\%$

Our conclusion was that

If EVMs count matches the corresponding VVPAT count
at all 4025 EVMs (randomly chosen as prescribed by EC) ,
then errors in EVM, if any, are less than 0.5%

Our report was part of the ECI response to SC. The SC rejected the petition on the basis of these statistical findings.

However, SCI suggested that 5 booths instead of 1 booth be chosen randomly in each assembly segment and EVM VVPAT matching be done. The ECI accepted the suggestion and instructed the returning officers in each assembly segment to pick randomly five booths and verify if EVM count and VVPAT count match, and report the outcome.

When we are having random sample of size 5 from each assembly segment, we can consider

Null Hypothesis $H_0: \theta \geq 0.25\%$

and

Alternate Hypothesis $H_1: \theta = 0$

The $p - value$ of observing no mismatch is very very low, less than 0.000000001.

(it is 0.0000000000320973024519772 or 0.000,000,000,032,097,302,451,977)

In 2019, all 5×4025 EVM-VVPAT matched. Moreover there was no recount request that was filed in a court, (which needs to be accompanied by reasons).

Problem:

Given k , $\{n_i : 1 \leq i \leq k\}$, integers and $p \in (0, 1)$

Compute

$$\theta = \max \left\{ \prod_{i=1}^k \left(1 - \frac{a_i}{n_i} \right) : 0 \leq a_i \leq n_i, \quad \forall 1 \leq i \leq k, \quad \sum_{i=1}^k a_i \geq p * \sum_{i=1}^k n_i \right\}.$$

Writing $x_i = n_i - a_i$, $1 \leq i \leq k$, we see that

$$\theta = \max \left\{ \prod_{i=1}^k \frac{x_i}{n_i} : 0 \leq x_i \leq n_i, \quad \forall 1 \leq i \leq k, \quad \sum_{i=1}^k x_i \leq (1-p) * \sum_{i=1}^k n_i \right\}.$$

or writing $\Gamma = \prod_{i=1}^k n_i$

$$\theta = \frac{1}{\Gamma} \max \left\{ \prod_{i=1}^k x_i : 0 \leq x_i \leq n_i, \quad \forall 1 \leq i \leq k, \quad \sum_{i=1}^k x_i \leq (1-p) * \sum_{i=1}^k n_i \right\}.$$

Without loss of generality, we assume that $n_1 \geq n_2 \geq \dots \geq n_k$.

Suppose that the maximum is attained at $\{x_i^* : 1 \leq i \leq k\}$. For any $1 \leq i < j \leq k$ we can assume that $x_i^* \geq x_j^*$, for if $x_i^* < x_j^*$ we can swap x_i^* and x_j^* . Thus we assume $x_1^* \geq x_2^* \geq \dots \geq x_k^*$.

Now we will show that for $1 \leq i < j \leq k$, and $x_j^* < n_j$ implies $x_i^* \leq x_j^* + 1$. For, if $x_i^* \geq x_j^* + 2$, let $\tilde{x}_i = x_i^* - 1$, $\tilde{x}_j = x_j^* + 1$, and $\tilde{x}_u = x_u^*$ for all remaining u . Then $0 \leq \tilde{x}_u \leq n_u$ for all $1 \leq u \leq k$ and

$$\begin{aligned}\tilde{x}_i \tilde{x}_j &= (x_i^* - 1)(x_j^* + 1) \\ &= x_i^* x_j^* - 1 + (x_i^* - x_j^*) \\ &\geq x_i^* x_j^* - 1 + 2 \\ &\geq x_i^* x_j^* + 1\end{aligned}$$

and thus $\left(\prod_{u=1}^k \tilde{x}_u\right) > \left(\prod_{u=1}^k x_u^*\right)$.

Thus we conclude that we must have $c + 1 = x_1^* = x_2^* = \dots = x_u^*$; $c = x_{u+1}^* = x_{u+2}^* = \dots = x_v^*$ and $x_{v+1}^* = n_{v+1}$, $x_{v+2}^* = n_{v+2}$, $\dots x_k^* = n_k$ and

$$\sum_{i=1}^k x_i^* \leq (1 - p) * \sum_{i=1}^k n_i \leq \sum_{i=1}^k x_i^* + 1.$$

The required quantity θ then can be computed by the following python code given the fraction p and the vector $N = \{n_i\}$, ordered in decreasing size.

```

def prob(p):
    total=0
    x=[]
    a=[]
    for z in range(len(N)):
        x.append(N[z])
        a.append(0)
        total=total+N[z]
    ac=int(total*p)+1
    for j in range(ac):
        for k in range(len(N)):
            if(x[k]>x[k+1]):
                x[k]=x[k]-1
                a[k]=1
                break
    Q=1.0
    for k in range(len(N)):
        if(a[k]>0):
            Q=Q*x[k]/N[k]
            Q=Q*(x[k]-1)/(N[k]-1)
            Q=Q*(x[k]-2)/(N[k]-2)
            Q=Q*(x[k]-3)/(N[k]-3)
            Q=Q*(x[k]-4)/(N[k]-4)
    print(total,ac,p,Q)
prob(p)

```