TRABAJOS DE ESTADISTICA Y
DE INVESTIGACION OPERATIVA
Vol. 33, Núm. 3, 1982, pp.119 a 122

PROBABILITY OF REVERSAL IN AN ELECTION WITH MORE THAN TWO CANDIDATES*

V. K. Rohatgi
Bowling Green State University

Consider an election with three candidates A_1, A_2 and A_3. Suppose that N is the total number of votes cast of which A_1 receives a_1 votes, A_2 receives a_2 votes and A_3 receives $a_3 = N - (a_1 + a_2)$ votes. We assume without loss of generality that $a_1 > a_2 > a_3$. Suppose further that n votes are irregular or suspect. If these votes are removed it is possible that the result of the election may be reversed.

Example: Let $N = 2550$, $a_1 = 900$, $a_2 = 850$, $a_3 = 800$ and $n = 200$. It is then possible that of the 200 suspect votes more than 100 were for A_1 and not more than 50 for A_2 so that their elimination would result in a reversal in favor of A_2. Similarly, if more than 100 suspect votes were for A_1 and none of A_3, then there would be a reversal in favor of A_3. Does such a possibility preclude the determination of the rightful winner without holding a new election?

We compute the probability of reversal under the assumption that the suspect votes form a random sample of size n (without replacement) from the total of N votes cast. This assumption is reasonable as long as no fraud is involved or charged.

Let X_i = number of suspect votes for A_i, $i = 1, 2, 3$. Then $X_1 + X_2 + X_3 = n$ and X_1, X_2 have a multivariate hypergeometric distribution with mass function given by

* Supported by NSF grant no. MCS78 - 01338.

$$P(X_1 = x_1, X_2 = x_2) = \frac{\prod_{i=1}^{3} \binom{a_i}{x_i}}{\binom{N}{n}}, \quad a_1 + a_2 + a_3 = N \qquad (1)$$

$$x_1 + x_2 + x_3 = n$$

It is easily seen (see Rohatgi (1976), p. 119) that X_i has a marginal hypergeometric distribution with mean

$$\mathcal{E}\, X_i = \mu_i = \frac{n}{N}\, a_i \qquad (2)$$

and variance

$$\sigma_i^2 = \frac{n\, a_i}{N^2\,(N-1)}\,(N - a_i)\,(N - n) \qquad (3)$$

Moreover,

$$\mu_{ij} = \text{Cov}\,(X_i, X_j) = -\frac{n\,(N-n)}{N^2\,(N-1)}\, a_i\, a_j, \quad i \neq j \qquad (4)$$

Now there will be a reversal if and only if

$$\max\,(a_3 - X_3, a_1 - X_1) \leqslant a_2 - X_2 \quad \text{or} \quad \max\,(a_2 - X_2, a_1 - X_1) \leqslant a_3 - X_3$$

Hence

$$P(\text{Reversal}) \leqslant P(X_1 - X_2 \geqslant a_1 - a_2 \quad \text{or} \quad X_1 - X_3 \geqslant a_1 - a_3) \qquad (5)$$

which may be computed from (1). Since, however, N is usually large we find an approximation for the probability in (5). Clearly

$$P(\text{Reversal}) \leqslant P(X_1 - X_2 \geqslant a_1 - a_2) + P(X_1 - X_3 \geqslant a_1 - a_3)$$

We will use the fact that $\dfrac{X_i - X_j - \mathcal{E}(X_i - X_j)}{\sqrt{\mathrm{Var}\,(X_i - X_j)}}$ is approximately a standard normal random variable. Clearly from (2) through (4)

$$\mathcal{E}\,(X_i - X_j) = \frac{n}{N}\,(a_i - a_j)$$

and

$$\sigma_{ij}^2 = \mathrm{Var}\,(X_i - X_j) = \frac{n\,(N-n)}{N^2\,(N-1)}\,[a_i\,(N - a_i) + a_j\,(N - a_j) + 2\,a_i\,a_j]$$

$$= \frac{n\,(N-n)}{N^2\,(N-1)}\,[N\,(a_i + a_j) - (a_i - a_j)^2]$$

$$= \frac{n\,(N-n)}{N^2\,(N-1)}\,[4\,a_i\,a_j + a_k\,(a_i + a_j)]$$

where $\{i, j, k\} = \{1, 2, 3\}$. Hence

$$P(X_i - X_j \geqslant a_i - a_j)$$

$$= P\left(\frac{X_i - X_j - \mathcal{E}(X_i - X_j)}{\sigma_{ij}} \geqslant \frac{(a_i - a_j) - \frac{n}{N}\,(a_i - a_j)}{\sqrt{\frac{n\,(N-n)}{N^2\,(N-1)}\,[4\,a_i\,a_j + a_k(a_i + a_j)]}}\right)$$

$$\doteq P\left(Z \geqslant \frac{(a_i - a_j)\sqrt{N-n}}{\sqrt{\frac{n}{N}\,[4\,a_i\,a_j + a_k\,(a_i + a_j)]}}\right) \tag{6}$$

where we have replaced $N - 1$ by N in the denominator and Z is standard normal. Note that

$$4\,a_i\,a_j < (a_i + a_j)^2 < N^2$$

and

$$4\,a_k\,(a_i + a_j) < (a_i + a_j + a_k)^2 = N^2$$

so that

$$P(\text{Reversal}) < P(Z > 2\,(a_1 - a_2)\sqrt{(N-n)/5\,N\,n}) \qquad (7)$$

$$+\,P(Z > 2(a_1 - a_3)\sqrt{(N-n)/5\,N\,n})$$

In the example above

$$P(\text{Reversal}) < P(Z > 2(50)\sqrt{(2550-200)/2550\,(200)\,(5)})$$

$$+\,P(Z > 2(100)\sqrt{(2550-200)/2550\,(200)\,(5)})$$

$$= P(Z > 3.07) + P(Z > 6.13) < .0012$$

The result (7) can be extended in a straightforward manner to more than three candidates. One can also use this analysis to compute the probability that the ranking of candidates will be preserved after the removal of irregular votes. It is interesting to note that the probability of reversal depends essentially on the plurality and the number of irregular votes.

In the case of two candidates we can get a nicer expression for the probability of reversal. In this case we note that $a_3 = 0$, $X_3 = 0$ so that $X_2 = n - X_1$, $a_2 = N - a_1$ and there is a reversal if and only if

$$X_1 \geqslant \frac{a_1 - a_2 + n}{2}$$

Returning to (6) and substituting $a_k = 0$ we get

$$P(\text{Reversal}) < P(Z \geqslant (a_1 - a_2)\sqrt{(N-n)/N\,n})$$

which is precisely the expression obtained in Finkelstein (1978), pp. 128-129.

REFERENCES

M. O. Finkelstein (1978). *Quantitative Methods in Law,* Free Press, New York.

V.K. Rohatgi (1976). *An Introduction to Probability Theory and Mathematical Statistics,* Wiley, New York.