Evaluating Resistance to False-Name Manipulations in Elections

Bo Waggoner
School of Engineering
and Applied Sciences
Harvard University
Cambridge, MA 02138, USA
bwaggoner@fas.harvard.edu

Lirong Xia
School of Engineering
and Applied Sciences
Harvard University
Cambridge, MA 02138, USA
lxia@seas.harvard.edu

Vincent Conitzer
Department of Computer Science
Duke University
Durham, NC 27708, USA
conitzer@cs.duke.edu

Abstract

In many mechanisms (especially online mechanisms), a strategic agent can influence the outcome by creating multiple false identities. We consider voting settings where the mechanism designer cannot completely prevent false-name manipulation, but may use false-name-limiting methods such as CAPTCHAs to influence the amount and characteristics of such manipulation. Such a designer would prefer, first, a high probability of obtaining the "correct" outcome, and second, a statistical method for evaluating the correctness of the outcome. In this paper, we focus on settings with two alternatives. We model voters as independently drawing a number of identities from a distribution that may be influenced by the choice of the false-name-limiting method. We give a criterion for the evaluation and comparison of these distributions. Then, given the results of an election in which false-name manipulation may have occurred, we propose and justify a statistical test for evaluating the outcome.

Introduction

A commonly studied method of manipulating mechanisms, particularly those run on the Internet, is through the creation of false identities. In the systems literature, this is known as a "Sybil attack" (Douceur 2002). False-name manipulations have also been studied by AI researchers, from the perspective of incentives and mechanism design. This literature has focused on designing mechanisms that are *false-name-proof* (Yokoo, Sakurai, and Matsubara 2001; 2004), meaning that even if participating in the mechanism under multiple identities were possible, an agent would derive no benefit from doing so. An overview of this line of work is given by Conitzer and Yokoo (2010). They emphasize the mostly negative nature so far of results on mechanisms that are false-name-proof in the strict sense, though they hold out more hope for extended models, such as ones that can use social-network structure (Conitzer et al. 2010). Meanwhile, however, mechanisms that are vulnerable to false-name manipulation continue to be used in practice.

In this paper, we take a different approach and address the implications of operating mechanisms in which false-name manipulation may take place. We focus on *voting* settings, which are particularly challenging for the design of mechanisms that are completely false-name-proof; for example, with two alternatives, the best possible false-name-proof voting rule simply flips a fair coin to decide the winner if there is even the slightest disagreement among the voters (Conitzer 2008a). Restricting to single-peaked preferences may give a small improvement (Todo, Iwasaki, and Yokoo 2011). If the cost of voting is taken into account, somewhat more positive results can be obtained (Wagman and Conitzer 2008).

While it has proven difficult to prevent false-name manipulation from a traditional mechanism design point of view by incentivizing voters not to create false identities, a more commonly used method for preventing false-name manipulations in real life is to restrict voters' ability to create false identities. One prominent method is to use a CAPTCHA, which prevents automated voting and may reduce (but still allows) the casting of many votes by a single agent. The designer may use additional false-name-limiting methods in an attempt to prevent such manipulation: allowing only one vote from each IP address; requiring voter registration with a confirmed e-mail address, phone number, or credit card number; using a memory test (Conitzer 2008b); and perhaps (as an extreme example) charging a small fee to vote.

These methods might be effective in limiting the number of false identities created by a strategic voter, but unless we know that each voter will cast the same number of votes, we do not know for sure how much influence these false identities will have on the outcome of the election. Therefore, it is important to be able to assess the reliability of such an election's outcome. We would like to be able to answer, e.g., the following questions. Is a false-name-limiting method where a voter creates 1 identity 80% of the time and 2 identities 20% of the time better than another method where a voter creates 1 identity 90% of the time and 3 identities 10% of the time? If we run an election using one of these false-name-limiting methods and observe 100 votes for alternative A and 94 votes for B, how surely can we say that A would have won if no voter had created false identities? To the best of our knowledge, the question of how to approach and evaluate an election in which false-name manipulation may occur has not yet been addressed theoretically in the literature.

Our contributions. In this paper, we take a first step towards addressing these questions. A premise of this paper is that each of these false-name-limiting methods is characterized by a probability distribution (which we call the *individual vote distribution*) over the number of identities used (or votes cast) by each voter, and each voter generates a number of identities i.i.d. from this distribution. This distribution can be estimated from experiments, e.g., on Mechanical Turk.

We focus on single-peaked preferences, and for convenience present our results in the case of two alternatives; these results extend naturally to the single-peaked preferences setting. Let A and B denote the two alternatives, and suppose the majority rule is used to select the winner. Let n_A (respectively, n_B) denote the number of voters who support A (respectively, B).

First, we present a criterion by which to compare two individual vote distributions. Suppose we have two false-name-limiting methods (e.g., using CAPTCHAs and using memory tests) that are characterized by two individual vote distributions π_1 and π_2. Suppose w.l.o.g. that $n_A > n_B$ (if $n_A = n_B$, then we do not have a basis to compare the two false-name-limiting methods). We say that an election outcome is "correct" if A receives more votes than B. The first question we address, therefore, is how to evaluate these two distributions and select the one that is more likely to produce a correct outcome. We show that in large and relatively close elections, the method whose individual vote distribution has a higher [mean/standard deviation] ratio is more likely to produce a correct outcome. (If the election is not close, then any method gives a correct outcome with high probability, which trivializes the question.)

Second, we consider the analysis of the results of an election involving false-name manipulation. Given the individual vote distribution of the method, we assume that the only data we can observe is the number of votes cast for each alternative. That is, we know neither the actual numbers of supporters n_A and n_B nor their prior distributions. Because of this, and the fact that only one data point (election outcome) is observed, we are unable to apply standard statistical significance tests to this setting. Instead, we propose a novel, to our knowledge, statistical test for obtaining a *p*-value for these election results. To do so, we give a formula for computing a test statistic (referred to as the *adjusted margin of victory*) from the votes cast and derive the best choice of the parameter. Our simulations show that our proposed adjusted margin of victory formula is indeed a good measurement in this context.

Preliminaries

We focus on the case of two alternatives, denoted by A and B. Let n_A and n_B denote the number of supporters ("true" identities) for A and B, respectively; we refer to (n_A, n_B) as a *supporter profile* and let $n = n_A + n_B$. Each supporter creates zero or more identities and uses them to cast votes. (The creation of zero identities models the decision not to participate in the election, possibly due to the cost or hassle of participation.)

In this paper, a false-name-limiting method is characterized by a distribution π on the number of identities created by each individual supporter. We assume that, given the election mechanism, the number of identities created by any given supporter is independent and identically distributed according to π. We call π an *individual vote distribution*.

Throughout this paper, we will consider only individual vote distributions π which have mean $\mu > 0$, variance $\sigma^2 > 0$, and absolute third moment $\rho < \infty$. If $\mu = 0$, then no participant ever casts any votes. If $\sigma^2 = 0$, each individual casts the same number of votes, and the outcome always exactly reflects the supporter profile. Finally, a distribution with an infinite absolute third moment would seem to be unlikely in the voting setting: It must not only have no bounds on the number of votes that can be cast by a single individual, but must moreover be heavy-tailed.

Given an individual vote distribution π, we let V_A and V_B denote two random variables representing the number of votes received by A and B respectively. Let $V = V_A + V_B$, the total number of votes cast, and let $D = V_A - V_B$, the number of votes by which A exceeds B. (D may be negative.) We use v_A and v_B to denote realizations of V_A and V_B.

By the Central Limit Theorem, as n_A grows large, the distribution of V_A (resp., V_B) approaches that of a Gaussian V_A' (resp., V_B') with mean μn_A (resp., μn_B) and variance $\sigma^2 n_A$ (resp., $\sigma^2 n_B$). In the same way, the distribution of V approaches that of a Gaussian V' with mean $\mu(n_A + n_B)$ and variance $\sigma^2(n_A + n_B)$. Finally, the distribution of D approaches that of a Gaussian D' with mean $\mu\,(n_A - n_B)$ and variance $\sigma^2\,(n_A + n_B)$.

Evaluation of Individual Vote Distributions

In this section, we consider the comparison of two false-name-limiting methods. Consider individual vote distributions π_1 and π_2 corresponding to two different methods. π_1 and π_2 have respective means μ_1, μ_2, variances σ_1^2, σ_2^2, and absolute third moments ρ_1, ρ_2. Suppose that the supporter profile is fixed and without loss of generality that $n_A > n_B$; therefore, a "correct" outcome is one in which A receives more votes ($D > 0$). Under π_1, we denote the random variable for the difference in votes $V_A - V_B$ by D_1; under π_2, we denote it by D_2. Similarly, define the corresponding Gaussian variables D_1' and D_2'. Recall that $\Pr[D_1 > 0]$ is the probability of a correct outcome under π_1.

We show that for close elections, as the total number of supporters grows large, these distributions may be effectively evaluated on the basis of a single statistic: the ratio of the mean to the standard deviation. To do so, we consider a sequence of elections in which the total number of supporters n is strictly monotonically increasing, and in each of which $n_A > n_B$. We will be concerned with such sequences in which the outcomes are close ($n_A - n_B$ is $O(\sqrt{n})$), though not "dead even" ($n_A - n_B$ is $\omega(1)$), as n increases.

Theorem 1 *Suppose that, as $n \to \infty$, $n_A - n_B$ is $\omega(1)$ and $O(\sqrt{n})$. If $\frac{\mu_1}{\sigma_1} > \frac{\mu_2}{\sigma_2}$, then there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $Pr[D_1 > 0] > Pr[D_2 > 0]$.*

That is, for any sequence of increasingly large, close elections, a higher ratio of mean to standard deviation is sufficent

to imply that, eventually, π_1 will be more likely to produce a correct outcome than π_2.

Proof: First, we note the following consequence of the Berry-Esséen Theorem (Durrett 1991):

$$\left|\Pr[D > 0] - \Pr[D' > 0]\right| \leq \frac{c\rho}{\sigma^3\sqrt{n}} \quad (1)$$

for a constant c. That is, given an individual vote distribution and supporter profile, we can bound the difference between the true probability of a correct outcome and its Gaussian approximation; morever, this bound is proportional to one over the square root of the number of supporters. This implies that

$$\Pr[D_1 > 0] - \Pr[D_2 > 0] \geq \Pr[D_1' > 0] - \Pr[D_2' > 0] - \frac{c}{\sqrt{n}}\left(\frac{\rho_1}{\sigma_1^3} + \frac{\rho_2}{\sigma_2^3}\right). \quad (2)$$

We wish to show that there exists an N so that, whenever $n \geq N$, $\Pr[D_1 > 0] > \Pr[D_2 > 0]$. Therefore, by Equation (2), it suffices to exhibit an N for which $n \geq N$ implies

$$\Pr[D_1' > 0] - \Pr[D_2' > 0] > \frac{c}{\sqrt{n}}\left(\frac{\rho_1}{\sigma_1^3} + \frac{\rho_2}{\sigma_2^3}\right). \quad (3)$$

$D_1' \sim N(\mu_1(n_A - n_B)), \sigma_1\sqrt{n})$, and similarly for D_2'. So for the left side of this inequality, we substitute

$$\Phi\left(\frac{\mu_1}{\sigma_1} \cdot \frac{n_A - n_B}{\sqrt{n}}\right) - \Phi\left(\frac{\mu_2}{\sigma_2} \cdot \frac{n_A - n_B}{\sqrt{n}}\right) \quad (4)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\left(\frac{\mu_2}{\sigma_2}\right)\frac{n_A - n_B}{\sqrt{n}}}^{\left(\frac{\mu_1}{\sigma_1}\right)\frac{n_A - n_B}{\sqrt{n}}} e^{-t^2} dt \quad (5)$$

$$> \left(\frac{\mu_1}{\sigma_1} - \frac{\mu_2}{\sigma_2}\right)\frac{n_A - n_B}{\sqrt{2\pi n}} e^{-\left(\frac{\mu_1}{\sigma_1}\frac{n_A - n_B}{\sqrt{n}}\right)^2}. \quad (6)$$

We can think of Expression (6) as a Riemann rectangle lower-bound on Expression (5). Since $n_A - n_B$ is $O(\sqrt{n})$, there exists an N_1 so that, for all $n \geq N_1$, for some constant k_1, $e^{-\left(\frac{\mu_1}{\sigma_1}\frac{n_A - n_B}{\sqrt{n}}\right)^2} \geq e^{-\left(k_1\frac{\mu_1}{\sigma_1}\right)^2}$ and so, for some constant $k_2 > 0$, Expression (6) becomes $\geq k_2\frac{n_A - n_B}{\sqrt{n}}$. Now, we have $k_2\frac{n_A - n_B}{\sqrt{n}} \geq \frac{c}{\sqrt{n}}\left(\frac{\rho_1}{\sigma_1^3} + \frac{\rho_2}{\sigma_2^3}\right)$ whenever

$$n_A - n_B \geq \frac{c}{k_2}\left(\frac{\rho_1}{\sigma_1^3} + \frac{\rho_2}{\sigma_2^3}\right). \quad (7)$$

Since $n_A - n_B$ is $\omega(1)$, there must exist an N_2 such that Inequality (7) holds for all $n \geq N_2$. Thus, we have that, whenever $n \geq \max\{N_1, N_2\}$, Inequality (3) holds and therefore $\Pr[D_1 > 0] > \Pr[D_2 > 0]$. ■

Discussion. This result is significant because it gives a designer a simple criterion for evaluating an individual vote distribution: taking the ratio of the mean to the standard deviation.

Excluded by Theorem 1's assumptions are two cases. The first consists of extremely (or "unnaturally") close elections: those in which the difference in votes received by the two alternatives remains constant (or even shrinks!) as the number of supporters diverges. The second consists of elections which are not close at all, in which any choice of π is likely to produce a correct outcome.

If we look again at the Gaussian approximation to the probability of a correct outcome, it is immediately clear why the statistic $\frac{\mu}{\sigma}$ is so significant:

$$\Pr[D' > 0] = \Phi\left(\frac{\mu}{\sigma} \cdot \frac{n_A - n_B}{\sqrt{n}}\right).$$

We can interpret the term $\frac{\mu}{\sigma} \cdot \frac{n_A - n_B}{\sqrt{n}}$ as a measurement of the probability of the correct outcome occurring ($V_A > V_B$).

The statistic $\frac{\mu}{\sigma}$ has intuitively appealing behavior. For instance, as σ approaches 0, leaving μ fixed, the correct outcome becomes certain regardless of the value of μ. As another example, suppose we change π so that every realization is doubled—i.e., everyone gets twice as many votes as before. This change in units does not affect the voting outcome; and indeed, it does not affect $\frac{\mu}{\sigma}$, either.

On the other hand, $\frac{\mu}{\sigma}$ also gives us some idea what changes to the distribution *would* improve the chances of a correct outcome. For example, it will be beneficial, if possible, to increase the mean number of identities while holding the standard deviation fixed. Such intuition may lead to novel false-name-limiting methods for voting mechanisms.

Evaluation of Election Results

In this section, we focus on evaluating the confidence that an observed election outcome (v_A and v_B) accurately reflects the preferences of the voters (n_A and n_B). Throughout the entire section, we assume that the election is run with a particular false-name-limiting method that has a fixed, known individual vote distribution π. But even though π is known, it is not obvious how to evaluate an election outcome such as, for example, 92 votes cast for A and 80 votes cast for B. In general, it is impossible to know how many of these votes were cast by false identities.

In such situations, we identify two possible tools that may be used to establish a confidence level: Bayesian analysis and statistical hypothesis testing. In our setting, Bayesian analysis, although a potentially natural approach, was ultimately rejected for the following reasons.

In a Bayesian analysis, the designer would explicitly state a prior distribution over supporter profiles (n_A, n_B), then use the observed votes (v_A, v_B) to update to a posterior. Then some method must be specified that selects a winning alternative and confidence level based on the posterior.

The first difficulty with such an approach is that formulating a prior distribution may be difficult. In some settings, information about the number of voters and their preferences may be costly or impossible to obtain. We prefer not to assume access to this information.

Additionally, relying upon a prior introduces the possibility of manipulating the outcome (or confidence level) by manipulating the prior. Especially in a voting setting where the election designer is supposed to be neutral, but the prior must necessarily be mostly subjective, this may be unacceptable.

Instead, we evaluate the election outcome using statistical hypothesis testing. The idea is this. First, we observe an outcome. In our case, this is the number of votes (v_A, v_B).

This suggests that there is a difference in parameters; here, for instance, $v_A > v_B$ suggests that $n_A > n_B$. However, an alternative explanation is that the outcome is due to chance rather than a difference in parameters. We test this explanation by fomulating a *null hypothesis* that is neutral between parameters and compute the probability of observing a result as extreme as (v_A, v_B). In our case, this null hypothesis is that $n_A = n_B$.[1]

When applied to voting, a statistical hypothesis test will take the form outlined in Algorithm 1. Each outcome (v_A, v_B) we might observe is associated with a *test statistic* β that summarizes the outcome; more extreme outcomes are associated with larger values of β. Suppose that the outcome we actually observe is associated with the value $\hat{\beta}$. Then we term *p-value* the probability, given a neutral null hypothesis $n_A = n_B$, of observing some result associated with a $\beta > \hat{\beta}$. If this probability is high, then we cannot reject the null hypothesis, and so our election outcome is inconclusive. If, on the other hand, this *p*-value is very low – lower, say, than some *confidence level* R – then our results would be very unlikely to arise due to chance. In this case, we may reject the null hypothesis and have confidence in our election results.

Algorithm 1 Statistical hypothesis test for elections

1. Select a significance level R (*e.g.*, 0.05).
2. Observe the election outcome (v_A, v_B); WLOG suppose $v_A > v_B$.
3. Compute a test statistic $\beta(v_A, v_B)$.
4. Assume as a null hypothesis that $n_A = n_B$.
5. Compute a *p*-value for observing a test statistic $\geq \beta$ given the null hypothesis.
6. If the *p*-value is below R, reject the null hypothesis. That is, accept alternative A as the winner of the election.

To apply Algorithm 1 in our specific setting, we must select, first, a formula for computing our test statistic β (step 3); and second, a procedure for computing a *p*-value (step 5). Our setting has several properties that differentiate it from those in which common statistical tests such as the *t*-test are used. The most important is that we observe only one data point (that is, we only operate the election and collect votes once).

Another important property of our setting is that we actually have many null hypotheses. Simply stating that $n_A = n_B$ is not specific enough; we get a different p-value for each possible assignment $n_A = n_B = 1, 2, \cdots$.

In the absence of a specific prior on the number of voters (some difficulties of which are mentioned above), we would prefer to be as neutral as possible when selecting a p-value. We consider two natural options as follows:

- The *max-p* option. For any β, its max-p value is the supremum *p*-value taken over all possible scenarios satisfying the null hypothesis (that is, $n_A = n_B = 1, 2, \cdots$).
- The *min-p* option. For any β, its min-p value is the infimum *p*-value taken over all possible scenarios satisfying the null hypothesis.

These two values can be used in combination in the following way. Let $\hat{\beta}$ denote the observed β value computed from an election outcome (v_A, v_B). If max-$p(\hat{\beta})$ is smaller than a preset significance level R (e.g., 5%), then we know that for *every* prior distribution over the supporter profile (n_A, n_B), the *p*-value of $\hat{\beta}$ is smaller than R. In this case, we can safely reject the null hypothesis, asserting that the election outcome is "correct" with high confidence. Similarly, if min-$p(\beta)$ is larger than R, then for every possible prior, we would have low confidence; we should assert that the election results are inconclusive.

This procedure provides an approach for executing Step 5 in Algorithm 1. In what follows, we propose a formula for completing Step 3 (computing a test statistic β), and prove theoretically that it is optimal among a natural family of test statistic formulae for elections. Then, we demonstrate through simulated elections that our statistical test seems to have in practice the nice intuitive behavior suggested by the theory. We shall see that, although the max-p and min-p phases of the test are each extreme-case analyses, our simulation results suggest that the gap between the max-p value and the min-p is quite small, making it potentially practical as a statistical hypothesis test of election outcomes.

To illustrate the statistical hypothesis testing procedure and motivate our choice of test statistic formula, we first consider the following example of an election.

Example 1 Suppose we set a significance level of 0.05 and we observe 92 votes for A and 80 for B. For our test statistic formula, let us simply take the difference in votes, so that $\hat{\beta} = 92 - 80 = 12$. Our null hypothesis is that alternative B has the same number of supporters as A; if we can safely reject this hypothesis, we will be confident that A is the correct winner. Let us begin by computing the max-p-value: the supremum, over all possible scenarios in which $n_B = n_A$, of obtaining a test statistic ≥ 12. We can easily imagine such a scenario. For example, suppose that $n_B = n_A = 100000$. Since votes are being drawn probabilistically, we would expect close to a 50% probability that A wins by 12 or more. Therefore, the max-p value will be far above the significance level of 0.05. So we can already see that we will not reject the null hypothesis, and thus we will not state that the outcome is correct with high confidence.

We could also have defined the test statistic to be $\beta = \frac{v_A - v_B}{v_A + v_B}$. In the observed election, A won by 12 votes out of the 172 cast, so $\hat{\beta} \approx 0.07$. This time, let us consider the min-p value. If we again consider $n_A = n_B = 100000$, we can tell that we would need a very extreme outcome to obtain a ratio higher than 0.07, and the min-p value will be lower than the significance level of 0.05. So in this case, we

[1] An alternative approach is to attempt to rule out all other parameters; in this case, letting the null hypothesis be $n_A \leq n_B$ and computing the probability of a result as extreme as (v_A, v_B). We decided against this hypothesis because it seems less informative when used with our approach: If we reject $n_A = n_B$, then we will also reject $n_A \leq n_B$; however, it seems difficult to say anything about when to accept the latter, whereas we will see a natural condition for accepting the former.

can already see that we will not accept the null hypothesis, so we will not state that the outcome is inconclusive. ■

Both of the above formulae are commonly known as the *margin of victory* in elections. Both are special cases of the following form, which we refer to as an *adjusted margin of victory* and is parameterized by a number α:

$$\beta_\alpha = \frac{v_A - v_B}{v^\alpha}\ , \tag{8}$$

where $v = v_A + v_B$ is the total number of votes cast. In this paper, we consider the family of test statistics of this form. Although we are not aware of a statistical significance test which fits the requirements of our setting, the adjusted margin of victory formula is similar in form to other known test statistics. Here, our goal is to find the optimal choice of formula among this family, and then evaluate its performance as a statistical hypothesis test.

Optimal Adjusted Margin of Victory

In this section, we show that, when using the max-p and min-p approach, the optimal choice of α is 0.5. More precisely, we will show that for any $\alpha < 0.5$, the max-p value is always very high, so we will never reject the null hypothesis regardless of the the state of the world (leading to what is called a *Type II error* in statistical hypothesis testing). On the other hand, for any $\alpha > 0.5$, the min-p value will be very low, so we will never accept the null hypothesis (leading to a *Type I error*).

Theorem 2 *For any $\alpha < 0.5$, any observed adjusted margin of victory $\hat{\beta} > 0$, and any $\delta > 0$, there exists a supporter profile with $n_A = n_B$ such that Pr$[\frac{V_A - V_B}{V^\alpha} \geq \hat{\beta}] \geq 0.5 - \delta$.*

That is, when we use β_α (with $\alpha < 0.5$) as the test statistic, whatever its value $\hat{\beta}$ for the actual election outcome, the max-p value is always arbitrarily close to 0.5. This means that Algorithm 1 *never* rejects the null hypothesis. Therefore, we are prone to Type II errors.

Proof: Let n be the total number of voters, and select a profile in which $n_A = n_B = n/2$. Recall that the total number of votes is $V = V_A + V_B$ and the difference in votes is $D = V_A - V_B$. We are to show that there exists an n for which $\Pr[\frac{D}{V^\alpha} \geq \hat{\beta}] \geq 0.5 - \delta$. Verbally, we will exhibit a total number of voters n such that, when exactly half support each candidate, there is close to 50% probability that A wins with an adjusted margin of victory of at least $\hat{\beta}$.

We will use the following two lemmas. The first lemma directly follows from the Central Limit Theorem.

Lemma 1 *For any $\epsilon, \delta_1 > 0$, there exists an N s.t. for all $n \geq N$, $V \in [(1-\epsilon)\mu n, (1+\epsilon)\mu n]$ with probability at least $1 - \delta_1$.*

Lemma 2 *For any $k > 0, \alpha < 0.5$, and $\delta_2 > 0$, there exists an N s.t. for all $n \geq N$, $D \geq kn^\alpha$ with probability at least $0.5 - \delta_2$.*

Proof: Using the Berry-Esséen Theorem, for a given positive value such as $\delta_2/2$, we can select an N_1 so that, for all $n \geq N_1$, $|\Pr[D \geq kn^\alpha] - \Pr[D' \geq kn^\alpha]| \leq \delta_2/2$.

Recalling that D' is a Gaussian with mean $\mu(n_A - n_B) = 0$ and variance $\sigma^2 n$, we have that $\Pr[D' \geq kn^\alpha] = 1 - \Phi(\frac{kn^\alpha}{\sigma\sqrt{n}}) = 1 - \Phi(\frac{k}{\sigma n^{0.5-\alpha}})$. Since k and σ are constant, and $\alpha < 0.5$, there exists an N_2 for which, whenever $n \geq N_2$, $\Phi\left(\frac{k}{\sigma n^{0.5-\alpha}}\right) \leq 0.5 + \delta_2/2$, and so $\Pr[D' \geq kn^\alpha] \geq 0.5 - \delta_2/2$. Thus, when $n \geq \max\{N_1, N_2\}$, $\Pr[D \geq kn^\alpha] \geq 0.5 - \delta_2$. ■

We can now apply these lemmas to prove our result: that there exists an n for which $\Pr[\frac{D}{V^\alpha} \geq \hat{\beta}] \geq 0.5 - \delta$. First, fix some small ϵ. By Lemma 1, we know there is an N_1 so that, when $n \geq N_1$, there is at most a $\delta/2$ probability that V is outside the range $[(1-\epsilon)\mu n, (1+\epsilon)\mu n]$.

Conditional on V being in the range, we can assert that

$$\frac{D}{V^\alpha} \geq \hat{\beta} \tag{9}$$

$$\text{is true if} \qquad D \geq \hat{\beta}(1+\epsilon)^\alpha \mu^\alpha n^\alpha\ . \tag{10}$$

By Lemma 2, we know there is an N_2 so that, when $n \geq N_2$, Equation (10) holds with probability at least $0.5 - \delta/2$.

Now select some $n \geq \max\{N_1, N_2\}$. We have shown that V falls outside the range given above with probability no more that $\delta/2$ and that Equation (10) fails to hold with probability no more than $0.5 + \delta/2$. Therefore, by a union bound, the probability that at least one of these events occurs is at most $0.5 + \delta$. Thus, the complement – that V falls within the given range and that Equation (10) holds – occurs with probability at least $0.5 - \delta$. But both of these occurring implies that Equation (9) holds, so the proof is complete. ■

We now state the corresponding result for the min-p value. Now, the issue is that for any β_α (with $\alpha > 0.5$), no matter what its value $\hat{\beta}$ for the actual election outcome is, the min-p is always 0, which means that we never accept the null hypothesis. Therefore, we are prone to Type I errors. The proof is analogous and is omitted.

Theorem 3 *For any $\alpha > 0.5$, any observed adjusted margin of victory $\hat{\beta} > 0$, and any $\delta > 0$, there exists a supporter profile with $n_A = n_B$ such that Pr$[\frac{V_A - V_B}{V^\alpha} \geq \hat{\beta}] \leq \delta$.*

These results show that values strictly above or below $\alpha = 0.5$ are poor choices in at least some settings. On the other hand, we now show that $\alpha = 0.5$ is a good choice in that it is susceptible to neither Type I nor Type II errors in the limit.

Theorem 4 *For any significance level R, there exists $b > 0$ such that, for any observed adjusted margin of victory $\hat{\beta} \geq b$, $\sup_{n_A = n_B}$ Pr$[\frac{V_A - V_B}{V^{0.5}} \geq \hat{\beta}] \leq R$.*

Proof: Let the significance level R be given. To prove the inequality holds for the supremum p-value over all $n = n_A = n_B$, we prove that the inequality holds for every n. To do so, we find some N and show the following: First, we find a b_1 so that, taking $\hat{\beta} \geq b_1$, the inequality holds for all $n \geq N$; and second, we find a b_2 so that, taking $\hat{\beta} \geq b_2$, the inequality holds for each $n < N$. Taking $b = \max\{b_1, b_2\}$ will complete the proof.

First, we fix some small ϵ and use Lemma 1 to select an N_1 such that, for all $n \geq N_1$, $V \in [(1-\epsilon)\mu n, (1+\epsilon)\mu n]$

with probability $1 - R/3$. When V is in this range, we wish to show for all large enough n that $\Pr[D \geq \hat{\beta}\sqrt{(1+\epsilon)\mu n}] \leq \frac{2}{3}R$.

Using the Berry-Esséen Theorem, we can select an N_2 such that, for all $n \geq N_2$, $|\Pr[D \geq x] - \Pr[D' \geq x]| \leq \frac{1}{3}R$ for all x. We also have that

$$\Pr[D' \geq \hat{\beta}\sqrt{(1+\epsilon)\mu n}] = \Phi\left(-\frac{\hat{\beta}\sqrt{(1+\epsilon)\mu n}}{\sigma\sqrt{n}}\right) = \Phi\left(-\frac{\hat{\beta}\sqrt{(1+\epsilon)\mu}}{\sigma}\right). \quad (11)$$

There exists a b_1 such that, for all $\hat{\beta} \geq b_1$, Expression (11) is $\leq \frac{1}{3}R$. Therefore, when $n \geq \max\{N_1, N_2\}$ and $\hat{\beta} \geq b_1$, $\Pr[\frac{V_A - V_B}{V^{0.5}} \geq \hat{\beta}] \leq R$.

Now, we consider the case of $n < \max\{N_1, N_2\}$. For each n, $\Pr[\frac{V_A - V_B}{V^{0.5}} \geq \hat{\beta}]$ is strictly decreasing in $\hat{\beta}$, so for each n, there exists a b_n such that, for all $\hat{\beta} \geq b_n$, $\Pr[\frac{V_A - V_B}{V^{0.5}} \geq \hat{\beta}] \leq R$. Take b_2 to be the maximum over these finitely many b_n, and the proof is complete. ■

Thus, $\beta_{0.5}$ is not susceptible to Type II errors in the limit. As the adjusted margin of victory grows large, the computed max-p-value goes to zero, so the null hypothesis will be eventually rejected. We now give the complementary result: that as the adjusted margin of victory goes to zero, the min-p value goes to $\frac{1}{2}$, which implies that $\beta_{0.5}$ is not susceptible to Type I errors in the limit. Again, the proof is similar and is omitted.

Theorem 5 *For any significance level $R < 0.5$, there exists $b > 0$ such that, for any observed adjusted margin of victory $\hat{\beta} \leq b$, $\inf_{n_A = n_B} Pr[\frac{V_A - V_B}{V^{0.5}} \geq \hat{\beta}] \geq R$.*

Experimental Results

In the previous section we have shown that if we use both max-p and min-p values to test the null hypothesis $n_A = n_B$, then $\beta_{0.5}$ is the only useful choice. However, this does not necessarily mean that $\beta_{0.5}$ is a good choice by itself. Our experimental studies in this section confirm that other choices are poor and support the conclusion that $\beta_{0.5}$ is a good choice for the max-p/min-p test, because the gap between the max-p and min-p values is small.

We provide results from two sets of simulations. In each of them, the individual vote distribution was $\pi(0) = 0.1, \pi(1) = 0.5, \pi(2) = 0.2$, and $\pi(3) = 0.2$. We have also tested other configurations, and have observed similar results.

• In Figure 1, we consider an election in which alternative A received 41 votes and B received 35. We plot, for each of the three choices of α (0.2, 0.5, and 0.8), the p-value at each number $n = 2n_A = 2n_B$ of total voters. Figure 1 provides an intuition for what happens when we select a value of α, compute the p-value w.r.t. β_α for each n, and then use the max-p value or min-p value to test the null hypothesis $n_A = n_B$. The same intuition underlies the proofs of Theorems 2 and 3. As n grows larger, we observe trending

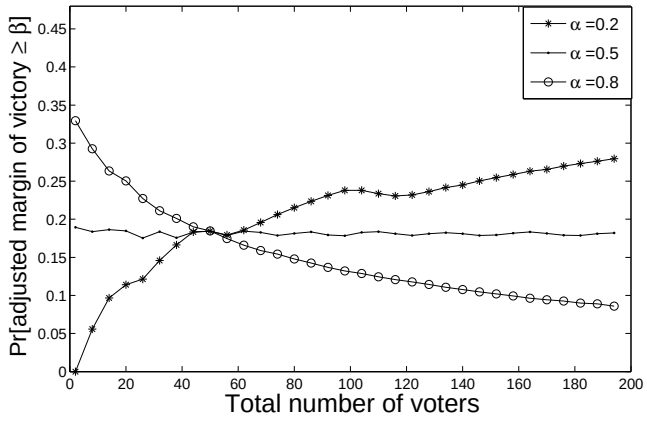

Figure 1: On the x-axis are null hypotheses given by $n = n_A + n_B$. For each α, the y-axis represents the p-value: the probability, given n, that a randomly generated election outcome has β_α value greater than the observed $\hat{\beta}_\alpha$.

toward an extreme for values of $\alpha < 0.5$ or $\alpha > 0.5$; furthermore, these trends occur not only for extremely large values of n, but also at relatively small values as well. Figure 1 also suggests that, when $\alpha = 0.5$, the p-value is stable w.r.t. n, meaning that the max-p and min-p approaches produce very similar results. This point is further illustrated in Figure 2.

• In Figure 2, we compare the behavior of the adjusted margin of victory β_α for the three different values of α (0.2, 0.5, and 0.8). For each α, we plot the max-p and min-p values. All choices of $\alpha < 0.5$ will exhibit a max-p value of 0.5; all choices of $\alpha > 0.5$ will exhibit a min-p value of 0. Thus, as seen in the plot, these values of α have widely differing results. For $\alpha = 0.5$, however, the max-p and min-p values are very similar. This shows that $\beta_{0.5}$ is quite discriminative if we combine the result of hypothesis tests using max-p and min-p.

Recall that if the max-p value for the observed $\hat{\beta}$ is smaller than the significance level R, then for *every* prior distribution over the total number of supporters n, we should reject the null hypothesis $n_A = n_B$; therefore, we should assert that the election outcome is "correct" with high confidence. If, on the other hand, the min-p value for $\hat{\beta}$ is greater than the significance level R, then for *every* prior distribution over n, we cannot reject the null hypothesis; therefore, we should assert that we do not have enough confidence to say that the election outcome is "correct". Finally, if the max-p value for $\hat{\beta}$ is larger than R and the min-p value for $\hat{\beta}$ is smaller than R, then our test cannot make any assertion. In our experiments, we observe that the last case seems to be rare for $\alpha = 0.5$. For example, in Figure 2, it happens only for $\hat{\beta} \in [1.24, 1.39]$ for $R = 5\%$.

Finally, we show a running example of Algorithm 1 combined with our max-p and min-p analyses.

Example 2 *Suppose the individual vote distribution π is the one defined in the beginning of this section. Suppose we observe that B gets 100 votes. If A gets at least 121 votes, then $\hat{\beta}_{0.5} \geq \frac{121-100}{\sqrt{121+100}} = 1.41$. From Figure 2 we can see*

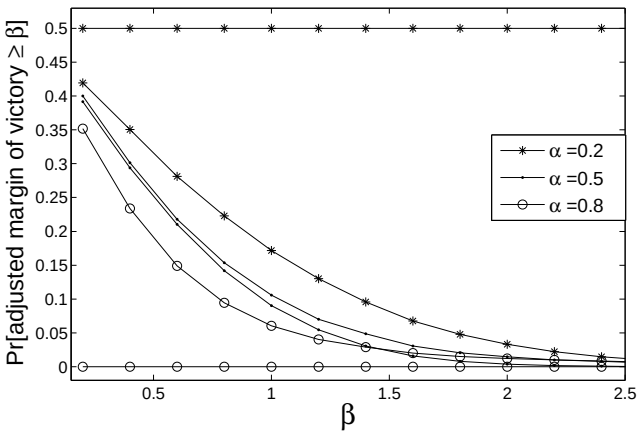

Figure 2: On the x-axis lie values for the adjusted margin of victory, from less extreme to more. For each α, we plot two lines: the minimum and maximum p-value at a given β. The maximum p-value of $\alpha = 0.2$ is always $\frac{1}{2}$; similarly, the minimum p-value of $\alpha = 0.8$ is always 0. The maxima and minima of $\alpha = 0.5$ lie close together.

that the max-p value for $\hat{\beta}_{0.5} = 1.41$ is smaller than 5%. In this case we should assert that A is the "correct" winner at the 5% significance level. If A gets at most 118 votes, then $\hat{\beta}_{0.5} \leq 1.22$, and the min-p value for $\hat{\beta}_{0.5} = 1.22$ is larger than 5%. In this case we cannot reject the null hypothesis and should declare the election outcome inconclusive at the 5% significance level. The above two claims hold for any prior distribution over the number of supporters. Only when A gets 119 or 120 votes are we unable to make any assertion.

Future Work

We have extended these results to the setting of single-peaked preferences with more than two alternatives. The intuition for this extension can be gained by grouping together an alternative and all those to its left as alternative A, and grouping together all those to its right as alternative B.

Many directions remain unexplored in the domain of mechanisms that are susceptible to false-name manipulation. One way to build on our work will be to investigate how specific false-name-limiting methods produce individual vote distributions. This work could proceed by means of experiments in online elections; it could also model agents with incentives for achieving certain results and mechanisms which impose a cost c_i for creating i different identities. Another direction to consider is a more sophisticated voter model in which numbers of votes are not necessarily drawn i.i.d.

An intriguing but potentially difficult open question is how to design false-name-resistant mechanisms in more general domains. This question may be most interesting in domains where strategic manipulation is also a concern; for example, when preferences are not single-peaked and there may be many alternatives. Analysis of results in such a case may be correspondingly difficult.

Pragmatically, we believe that our results can be applied to gain insight into a variety of practical voting settings. For instance, the voter turnout problem, in which supporters vote either once or zero times, is a special case of this work. These results may also be applied in various settings involving voting on the Internet.

Acknowledgements

The authors thank NSF Awards IIS-0812113, IIS-0953756, and CCF-1101659, as well as an Alfred P. Sloan fellowship, for support. Lirong Xia is supported by NSF under Grant #1136996 to the Computing Research Association for the CIFellows Project. We thank all anonymous AAAI-12 reviewers, participants of Dagstuhl seminar 12101 "Computation and Incentives in Social Choice", and Harvard EconCS Seminar attendees for helpful comments and suggestions.

References

Conitzer, V., and Yokoo, M. 2010. Using mechanism design to prevent false-name manipulations. *AI Magazine* 31(4):65–77. Special Issue on Algorithmic Game Theory.

Conitzer, V.; Immorlica, N.; Letchford, J.; Munagala, K.; and Wagman, L. 2010. False-name-proofness in social networks. In *Proceedings of the Sixth Workshop on Internet and Network Economics (WINE)*, 209–221.

Conitzer, V. 2008a. Anonymity-proof voting rules. In *Proceedings of the Fourth Workshop on Internet and Network Economics (WINE)*, 295–306.

Conitzer, V. 2008b. Using a memory test to limit a user to one account. In *Agent-Mediated Electronic Commerce (AMEC) workshop*, 60–72.

Douceur, J. R. 2002. The Sybil attack. In *First International Workshop on Peer-to-Peer Systems*, 251–260.

Durrett, R. 1991. *Probability: Theory and Examples*.

Todo, T.; Iwasaki, A.; and Yokoo, M. 2011. False-name-proof mechanism design without money. In *Proceedings of the Tenth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 651–658.

Wagman, L., and Conitzer, V. 2008. Optimal false-name-proof voting rules with costly voting. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 190–195.

Yokoo, M.; Sakurai, Y.; and Matsubara, S. 2001. Robust combinatorial auction protocol against false-name bids. *Artificial Intelligence* 130(2):167–181.

Yokoo, M.; Sakurai, Y.; and Matsubara, S. 2004. The effect of false-name bids in combinatorial auctions: New fraud in Internet auctions. *Games and Economic Behavior* 46(1):174–188.