CESifo Working Papers

www.cesifo.org/wp

Beyond Condorcet: Optimal Aggregation Rules Using Voting Records

Eyal Baharad
Jacob Goldberger
Moshe Koppel
Shmuel Nitzan

CESIFO WORKING PAPER NO. 3323
CATEGORY 2: PUBLIC CHOICE
JANUARY 2011

An electronic version of the paper may be downloaded
- *from the SSRN website: www.SSRN.com*
- *from the RePEc website: www.RePEc.org*
- *from the CESifo website: www.CESifo-group.org/wp*

CESifo *Working Paper No. 3323*

Beyond Condorcet: Optimal Aggregation Rules Using Voting Records

Abstract

The difficulty of optimal decision making in uncertain dichotomous choice settings is that it requires information on the expertise of the decision makers (voters). This paper presents a method of optimally weighting voters even without testing them against questions with known right answers. The method is based on the realization that if we can see how voters vote on a variety of questions, it is possible to gauge their respective degrees of expertise by comparing their votes in a suitable fashion, even without knowing the right answers.

JEL-Code: D700.

Eyal Baharad
Department of Economics
Bar Ilan University
Ramat Gan 52900
Israel
baharad@mail.biu.ac.il

Jacob Goldberger
School of Engineering
Bar Ilan University
Ramat Gan 52900
Israel
goldbej@eng.biu.ac.il

Moshe Koppel
Department of Computer Sciences
Bar Ilan University
Ramat Gan 52900
Israel
koppel@cs.biu.ac.il

Shmuel Nitzan
Department of Economics
Bar Ilan University
Ramat Gan 52900
Israel
nitzans@mail.biu.ac.il

1. Introduction

In Condorcet's (1785) paradigm of voting, some "correct" collective decision is presumed to exist, and some information regarding voters' skills at making that decision is assumed to be known. In this context, each individual voter is typically assigned a probability that represents his ability of making a correct decision (see, for example, Young (1988), (1995), (1996)). Given such probability assignments, we can compute the probability that some judgment aggregation rule (such as a majority rule) will actually yield the correct decision. In fact, given such probability assignments, the optimal aggregation rule can be identified (Nitzan and Paroush (1982), Shapley and Grofman (1984), Ben-Yashar and Nitzan (1997)).

Unfortunately, however, we generally do not have a firm foundation for estimating a voter's probability of making a correct decision. Even if a historical record of voters' individual decisions is available, we generally lack a definition of ground truth regarding the decided issues against which each voter's record could be compared. Some empirical attempts have thus been made to estimate voters' skills in medical and legal contexts, by comparing voter decisions to some exogenous proxy of the 'truth' (see, for example, Chapter 10 in Nitzan and Paroush (1985) and Karotkin (1994)). Such estimates of individual skills are, however, typically not sufficiently reliable for the purpose of determining an optimal aggregation rule.

A slightly more sophisticated estimation method has been proposed by Grofman and Feld (1983), an application of which to EU council voting has been considered by Nurmi (2002). They suggest to estimate a voter's decisional capability by the extent that his observed past decisions align with those of the majority. In other words, in this method the majority decision is considered as a plausible endogenous proxy for the 'truth'. This method is very interesting in that it anticipates the idea that voters' decisional skills can be somehow estimated using the track record of their decisions, even in the absence of knowledge of ground truth regarding those decisions.

In this paper, we show how this method can be generalized and optimized. To understand the sub-optimality of their method, notice that it can lead to internal inconsistency. A simplistic application of the method presumes the correctness of an outcome determined by simple majority rule (*SMR*) and assigns weights to voters accordingly. In turn, these assigned weights lead to a different outcome than has already been presumed, which might lead to different assigned weights. In this sense, the estimation procedure is usually inconsistent or 'unstable'; re-application of the estimation procedure yields different outcomes and, in turn, different estimation of voters' decisional skills.

This study considers the setting in which we have a track record of voters' individual decisions, but no given ground truth for the issues on which they voted. There are many real-world examples of repeated votes, in which competency varies among voters but remains invariant per voter, for which this framework is applicable. First, there are periodically repeated committee decisions regarding, for example, whether to change interest rates. Similar examples can be taken from the capital markets, court votes, medical expert decisions, firm theory (repeated decisions concerning quantity and prices), and more. Second, there are cases in which committee members are asked to vote on a range of similar but independent propositions. This has become quite common on the Internet, where site visitors are asked to offer judgments regarding repetitive tasks such as image tagging, expert identification and a variety of other tasks (hundreds of examples of which can be found, for example, at Amazon's Mechanical Turk site).

We provide an algorithm for finding consistent or “stable” evaluation of voters' skills – what we call *skill-evaluation equilibrium* – that ensures the assignment of optimal weights to the voters. Such consistency guarantees therefore the simultaneous fulfillment of the two necessary requirements an optimal skill-based aggregation rule should satisfy: estimating skills and aggregating accordingly. More specifically, we propose an iterative mechanism that allows skill evaluation based on repeated re-evaluation of the optimal voting outcomes. Any new such set of voting outcomes yields a new skill evaluation; it turns out that the simultaneous evaluation of skills and optimal outcomes converges, thus yielding internally consistent, stable skill-evaluation equilibrium, in which neither the outcomes nor the evaluated skills need to be changed. This unique property implies the attainment of optimal collective decision making as well as consistent, maximum-likelihood skill estimation.

The optimal weights assigned to the voters could, in principle, be equal, but they generally will not be. In the unlikely situation that they turn out to be equal, the applied rule is just the common *SMR*. By definition, the proposed optimal and consistent skill-based procedure outperforms any aggregation rule and, in particular, *SMR*, for any given number of voters, decision-making record and distribution of skills. In exhaustive experiments using simulated data, we demonstrate that the optimal procedure is substantially superior to *SMR*, under almost symmetric (uniform or normal), minimally competent skill distribution that allows unskilled decision makers. In fact, given sufficiently large voting records, the collective probability of making a correct choice converges to 1, even under conditions in which *SMR* achieves a correct decision with probability of about 0.7.

It is instructive to compare this result with that of Condorcet (1785). The main concern of Condorcet focused on the implications of 'equality' among voters for determining the socially most desirable aggregation rule, where voter equality is understood both in terms of assignment of equal voting rights and in terms of assignment of equal decisional skills. This main concern resulted in the celebrated Condorcet jury theorem (CJT) (Black (1958)). The theorem has two parts. The first part establishes the superiority of *SMR* over individual decision-making (the 'expert rule'), provided that voter skills are sufficiently competent and homogeneous, Condorcet (1785), Nitzan and Paroush (1982, 1985). (This result was subsequently extended to the case where there is uncertainty regarding individual decisional skills (Ben Yashar and Paroush (2000), Berend and Sapir (2005), Nitzan and Paroush (1985).) The second part establishes that *SMR* converges to a probability of 1 of making the correct decision as the number of voters grows, provided that the voters are sufficiently competent, individually or on average (Condorcet (1785), Grofman, Owen and Feld (1983), Owen, Grofman and Feld (1986), Berend and Paroush (1998), Paroush (1998)). Condorcet and his more recent followers do not take advantage of skill heterogeneity among the group members or of the track record of their decisions. In contrast, the optimality of our proposed aggregation rule is obtained by exploiting the diversity among group members that emerges from the track record of their decisions. Our work goes beyond Condorcet in the sense that we show how to obtain optimal decision-making without restrictive assumptions such as sufficiently high decision-making quality, skill homogeneity or existence of a sufficiently large group of decision makers. Instead, we assume only that a sufficient track record for voters' decisions is available, which can be exploited to optimally estimate their decisional skills and, in turn, optimally assign their decision-making weights, even in the absence of any knowledge of ground truth. We do in fact make one very weak assumption regarding voters' skills, namely, that a unanimous decision is correct with probability greater than 0.5. This is a far weaker assumption regarding voters' skills than required for CJT.

Voter equality as implicit in *SMR* is justified in situations where decisional skills are all of sufficient quality and homogeneity. In this case, a sufficiently large number of voters results in convergence to the maximal collective performance. In contrast, optimal decision-making that is based on our proposed voting procedure is *always* warranted, deriving its superiority from its ability to identify skills through learning from experience. The merit of this procedure and its advantage over *SMR* are especially high when skills are sufficiently spread out and the track record of individual decisions is sufficiently abundant.

The next section describes our setting, presents the proposed judgment aggregation rule and results establishing that this skill-based rule is consistent and optimal. In the following three sections, we use simulated data to demonstrate the superiority of our approach over *SMR* and one other baseline method, relating the results to a variety of parameters: the number of decision makers (whom, for convenience, we call voters), the number of issues with respect to which decisions have been made and the form of the distribution of individual decisional skills. Further implications, possible applications and possible extensions of the results are discussed in the concluding Section 6.

2. An Algorithm for Optimal Judgment Aggregation

In this section we present our algorithm. The discussion here follows that of our companion paper (Baharad et al 2010), which deals with the algorithm's convergence properties.

Let N={1,…,n}, $n \geq 3$, denote a finite set of voters and let M denote a set of m distinct binary issues, $m \geq 2$. The judgment of voter $i \in N$ over issue $j \in M$, is denoted by a_{ij}. Unlike preferences, judgments are binary; thus, $a_{ij} \in \{0,1\}$. a denotes the entire set of judgments $\{a_{ij}\}$. The rows and columns of the matrix are associated, respectively, with voters and issues. Hence, a column a_j in the matrix a is the judgments profile on issue j; similarly, a row a_i is the judgment profile of voter i. We assume that the issues are independent of each other (so that the problem of inconsistent aggregation across issues (List and Petit 2002) does not arise here) and that each issue has some (unknown) "correct" resolution, denoted by t_j. In addition, each voter i is associated with an unknown probability p_i of making the correct decision. This somewhat simplified assumption assigns a probability to a voter, and not to the combination of voter-issue. It supports an implicit assumption, according to which a voter can be referred to as having some fixed reliability level in a *field* (that is a set of issues under a common topic), and not only on a specific issue. The set of individual probabilities $\{p_i\}$ is denoted by θ. For simplicity, we assume that, in the absence of any information, the two possible resolutions of an issue are equally likely; that is, for every j, the prior probability $p(t_j = 1) = ½$.

A *judgment aggregation rule V* is a mapping from the set of individual judgments $a = \{a_{ij}\}$ to a set of binary decisions in $\{0,1\}^m$. Our objective is to find an optimal judgment aggregation rule, given no information other than the set of judgments $a = \{a_{ij}\}$.

The suggested framework does not assume that the individual skills, $\{p_i\}$, and the correct resolution for each issue are (ex-ante) known; hence, one might wonder in what sense a decision method could be optimal. In principle, given $\theta = \{p_i\}$ and {$p(t_j$

= 1)}, we could compute the conditional probability of obtaining the set of judgments *a*. Thus, given some set of judgments *a*, optimality is obtained by the values of θ and $\{p(t_j = 1)\}$ that maximize the probability of *a*. As shown below, the values $\{p(t_j = 1)\}$ can be determined from *a* and θ. Thus, denoting by $p(a;\theta)$ the probability of *a* given the parameters θ, our objective is to maximize $p(a;\theta)$. The suggested iterative approach for finding this maximum is based on some initial estimate of p_i. These values are re-used to compute, for each issue *j*, the probability that $t_j = 1$. Moreover, once all the conditional resolution probabilities $p(t_j = 1 | a)$ are given, one is able to compute, for each decision maker i, the most likely value of p_i. This (temporarily) most likely value of p_i is referred to as p_i' and $\theta' = \{p_i'\}$ is said to be *induced* from $\{p(t_j = 1 | a\})$. The iterative procedure is incomplete so long as $\theta \neq \theta'$, i.e. $p_i \neq p_i'$ for at least one decision maker. The procedure is complete when $\theta = \theta'$, i.e. for all i, $p_i' = p_i$; at this stage a *skill-evaluation equilibrium* is obtained.

To summarize, we have the following hill-climbing procedure, Q, for finding a skill-evaluation equilibrium:

1) Choose some initial θ.
2) Using the current θ, compute $p(t_j = 1 | a)$ for each $j = 1,...,m$.
3) Replace θ with the induced θ'.
4) Repeat until convergence.

It remains only to show how $p(t_j = 1 | a)$ is obtained given $\theta = \{p_i\}$, and how the maximum-likelihood values of $\theta = \{p_i\}$ are computed, given $\{p(t_j = 1 | a)\}$.

First, to compute $p(t_j = 1 | a)$, recall that the prior probability $p(t_j = 1) = ½$. Thus, Bayes' rule implies that $p(t_j = 1 | a) = p(t_j = 1 | a_j) =$

$$\frac{p(a_j | t_j = 1) \cdot p(t_j = 1)}{p(a_j | t_j = 1) \cdot p(t_j = 1) + p(a_j | t_j = 0) \cdot p(t_j = 0)} = \frac{p(a_j | t_j = 1)}{p(a_j | t_j = 1) + p(a_j | t_j = 0)} \quad (1)$$

where a_j is the judgment profile on issue j. This can easily be computed by substituting

$$p(a_j | t_j = 1) = \prod_i p(a_{ij} | t_j = 1) = \prod_{a_{ij}=1} p_i \prod_{a_{ij}=0} (1 - p_i) \quad (2)$$

and

$$p(a_j | t_j = 0) = \prod_i p(a_{ij} | t_j = 0) = \prod_{a_{ij}=0} p_i \prod_{a_{ij}=1} (1 - p_i)$$

Given the values $\{p(t_j = 1 | a)\}$, they can be compared to the judgments of individual i, in order to compute the maximum-likelihood values of $\theta' = \{p_i^{'}\}$. Specifically, the maximum likelihood value of p_i is equal to the average (over j) probability that $a_{ij} = t_j$. Thus,

$$p_i^{'} = \frac{1}{m}\left(\sum_{a_{ij}=1} p(t_j = 1 | a) + \sum_{a_{ij}=0} p(t_j = 0 | a)\right) = \frac{1}{m}\sum_j (1 - | p(t_j = 1 | a) - a_{ij} |)$$

The procedure Q is a special case of the well-known EM algorithm (Dempster et al. 1977). Thus, it can be shown (Baharad et al (2010)) that this procedure converges to a skill-evaluation equilibrium, and almost always converges to a local maximum.

In the following sections, we will show empirically that, for properly chosen initial values, the local maximum to which Q converges is usually a global maximum. It should be noted, however, that there are skill-evaluation equilibria that are not local maxima. For example, the probability set $p_i = 0.5$, for all i, is an equilibrium that is not a local maximum. Such problematic sets are rare and should not to be chosen as initial values.

Note that we do make a minimal assumption regarding voters' competency, in the sense that if voters are unanimous on some issue, then their vote is correct with probability greater than ½. More formally, we assume that $\prod_i p_i > \prod_i 1 - p_i$.

The procedure converges for θ as well as for $\{p(t_j = 1 | a)\}$. While the procedure does not technically entail a final aggregation rule, one naturally follows from the procedure. For an aggregation rule *V*, let $V_j(a) \in \{0,1\}$ be the decision of *V* for issue *j*, and let $V_j(a) = 1$ if and only if the obtained $p(t_j = 1 | a) > 0.5$. It is now shown that this aggregation rule satisfies an optimality condition that is well known in the voting literature.

An aggregation rule *V* is *linear*, if there exist weights w_i and constant *c* such that, for every *j*, $V(a_j) = 1$ if and only if $\sum_{i=1}^{n} w_i a_{ij} \geq c$. By the main result in Nitzan and Paroush (1982) and Shapley and Grofman (1984), if voter skills $\{p_i\}$ are known, a linear aggregation rule is *optimal*, that is, yields the maximal collective probability of making the correct decision, if $w_i = log(p_i/1-p_i)$.

It should be noted that the judgment aggregation rule *V* implied by procedure *Q* is an optimal linear aggregation rule. From equations (1) and (2) above, it follows that

$$p(t_j = 1 | a) > 0.5 \Leftrightarrow \prod_{a_{ij}=1} p_i \prod_{a_{ij}=0} (1 - p_i) > \prod_{a_{ij}=0} p_i \prod_{a_{ij}=1} (1 - p_i) \Leftrightarrow \prod_i \left(\frac{p_i}{1 - p_i}\right)^{(2a_{ij}-1)} > 1$$

$$\Leftrightarrow \sum_i (2a_{ij} - 1) \cdot \log\left(\frac{p_i}{1 - p_i}\right) > 0 \Leftrightarrow \sum_i a_{ij} \cdot \log\left(\frac{p_i}{1 - p_i}\right) > \frac{1}{2} \sum_i \log\left(\frac{p_i}{1 - p_i}\right)$$

The last inequality is precisely the one required by the definition of optimality for linear aggregation rules.

It should be noted that although the estimation of decisional skills is based on complete information regarding the judgments of all *n* individuals on all *m* issues, our method is applicable even when voter records are incomplete, that is, when some voter's decision on some issues are not available. All steps in the algorithm can be carried out using partial information. Thus, in particular, the method can be applied in cases where voters are allowed to abstain.

We add one note on strategic voting in this context. Our setting assumes a common interest among voters to arrive at the truth. Such common interest would hopefully reduce a voter's temptation to vote strategically in order to maximize his own weight. The voters' best interest is, after all, to assign the correct weights to every voter, in order to reach the correct decision. The question of strategic voting, when voters are solely concerned by the common collective interest, was recently examined by Ben-Yashar and Milchtaich (2007). They established that under the 'first best' voting rule, the decision makers do not have an incentive to vote strategically and non-informatively. Fortunately, our setting proposes a mechanism that results in the use of the 'first best' voting rule and is therefore *immune* to strategic non-informative voting. Such strategy-proofness does not hold under 'second best' anonymous aggregation rules, as have been demonstrated by Austen-Smith and Banks (1996), Ben-Yashar and Milchtaich (2007), Feddersen and Pesendorfer (1998) and McLennan (1998). In fact, in this setting, effective deliberation prior to the vote is expected to take place, as established in Coughlan (2000). That is, every member of the group has an incentive to truthfully reveal his private information and then all group members unanimously vote for the collectively best alternative.

3. Simulation Design

As we noted above, for appropriate initialization, the *Q* procedure converges to a locally maximal skill-evaluation equilibrium and it implies a linear aggregation rule that is optimal for that skill evaluation. In this section, we outline our simulation method and in the following two sections, we will show that this equilibrium typically yields a global maximum of the probability function $p(a;\theta)$, as borne out by a range of empirical experiments.

To test the effectiveness of the proposed method relative to some baseline methods, we apply simulated scenarios. In each simulation, for each j, the "correct" value t_j associated with issue j is randomly sampled using a fair coin mechanism. The voters' skills parameters $\{p_i\}$ are sampled according to some distribution, as will be specified in context.

For every voter i and issue j we generate the vote $a_{ij} = t_j$ if $k_{ij} = 1$ and $a_{ij} = 1 - t_j$ if $k_{ij} = 0$, where k_{ij} is the result of a coin toss with a probability p_i for a "1" result. The object of our algorithms is to find t_j, given the values a_{ij}. Our first baseline is *SMR* the application of which does not require the estimation of the voters' decisional skills. A more sophisticated baseline is a weighted majority rule where the optimal weight of voter i is obtained by estimating p_i as the proportion of issues for which a_{ij} coincides with the majority vote. Formally, let $\hat{x}_j$ denote the majority of votes over $(a_{1j},...,a_{nj})$, for $j = 1,...,m$. Then the estimated p_i is $\hat{p}_i = \frac{1}{m} |\{ j \mid a_{ij} = \hat{x}_j \}|$ for i = 1,...,n. We refer to this latter method, which is proposed in Grofman and Feld (1983), as *GF*. Finally, we use the procedure *Q*, with p_i initialized as in the second baseline. (Other initializations are possible; for example, every voter could be initially assigned some fixed $p_i > \frac{1}{2}$. The results for such alternative initializations are not substantially different than those reported below.) When the decisional skills estimated by the procedure Q imply convergence to a value of $p(t_j = 1 | a)$ greater than ½, the collective decision is assumed to be $t_j = 1$. In this case, the optimal rule is the weighted majority rule corresponding to the decisional skills estimated by the procedure Q. For the sake of completeness, we also present the results for the optimal rule where the actual values of $\{p_i\}$ are known. This represents the upper bound on the accuracy that can be achieved on the basis of approximated values of $\{p_i\}$.

The methods can be evaluated by comparing the true value of t_j to the value provided by each algorithm; for a given matrix a, we measure the proportion of correct t_j values determined by each algorithm. Below, we compare the algorithms on a single example for illustrative purposes. In the following two sections, we consider systematically generated simulations under varying assumptions regarding voter competence. In each case, we will see that Q significantly outperforms both *GF* and *SMR*.

For clarity, let's consider a simple example. Suppose that there are five voters whose true decisional skills are given by $\theta_t = \langle 0.46, 0.54, 0.92, 0.20, 0.92 \rangle$ and ten issues for each of which the correct result is 1. (This is for expositional simplicity and implies no loss of generality.) Using the above-described coin toss method, we generate a 5×10 vote matrix $\{a_{ij}\}$ as shown in Table 1. The likelihood of voters with skills θ_t voting precisely as indicated in this matrix is $2^{-33.8}$.

	1	1	1	1	1	1	1	1	1	1
(0.46)	0	0	1	0	1	1	1	1	0	0
(0.54)	1	1	1	0	1	0	1	0	0	1
(0.92)	1	0	1	1	1	1	1	0	1	1
(0.20)	0	0	0	0	1	0	0	0	0	0
(0.92)	1	1	1	1	1	1	1	1	1	1

Table 1. The entries in the 5×10 matrix $\{a_{ij}\}$ represent the vote of voter *i* on issue *j*. The uppermost row represents the correct results for the issues and the leftmost column represents the voters' respective skills.

We are given only the matrix $\{a_{ij}\}$ and our challenge is to find the correct result for each issue. Note first that *SMR* yields the outcome vector (1,0,1,0,1,1,1,0,0,1), which is correct for only 6 of 10 issues. The *GF* algorithm assigns decisional skills to the voters by comparing their respective votes to the majority vote. This yields the respective vector of skills $\theta_I = \langle 0.7, 0.8, 0.8, 0.5, 0.6\rangle$. Now, computing for j = 1,…,10 the probability that t_j = 1, we obtain the vector

$$\{p(t_j = 1 \mid \theta_I)\}_{j=1}^{10} = \langle 0.91, 0.39, 0.98, 0.39, 0.98, 0.78, 0.98, 0.18, 0.39, 0.91\rangle .$$

Using 0.5 as a cutoff, the *GF* method yields the identical vector of outcomes as *SMR*; it is thus correct for only 6 of 10 issues. Note that the likelihood of voters with skills θ_I voting precisely as indicated in the matrix $\{a_{ij}\}$ is $2^{-39.8}$.

Finally, we apply the *Q* procedure. The $\{p(t_j = 1 \mid \theta_I)\}_{j=1}^{10}$ vector is used to update the skills vector and, in turn, the updated vector of decisional skills is used to update the probabilities, for j = 1,…,10, that t_j = 1. Ultimately, the procedure converges to the skill-evaluation equilibrium $\theta_* = \langle 0.45, 0.65, 0.83, 0.18, 0.91\rangle$, which, in turn, yields the vector

$\{p(t_j = 1 \mid \theta_*)\}_{j=1}^{10} = \langle 0.99, 0.95, 0.99, 0.99, 0.94, 0.99, 0.99, 0.81, 0.99, 0.99 \rangle$. Thus, the *Q* procedure gives the correct result for all ten issues. Moreover, the likelihood of voters with skills θ_* voting precisely as indicated in the matrix $\{a_{ij}\}$ is $2^{-32.9}$, which is actually greater than that of the true skills θ_t.

4. Simulation Results: Near-Symmetric Skill Distributions

In our first set of simulations, voters' skills are sampled uniformly in the range $[0,1]$, subject to the single weak constraint that $\prod_i p_i > \prod_i 1 - p_i$. This constraint requires only that the probability that a unanimous decision is correct is greater than 1/2.

For the first experiments, the number of issues is fixed and equal to 10 and 100, respectively, while the number of voters varies from 1 to 1000. Accuracy results are tabulated over 10,000 independent trials (every trial represents a new choice of $\{p_i\}$ and $\{t_j\}$ and, hence, of *a*). As can be seen in Figure 1, the results remain almost constant for sufficiently large *n*. In particular, unlike in the standard Condorcet case and its extensions, *SMR* does not converge to accuracy of 1 as the number of voters increases. More significantly, for $m = 10$ and sufficiently large *n*, *GF* outperforms *SMR* by approximately 0.12 and the procedure *Q* outperforms *GF* by approximately 0.03. Nevertheless, the limited number of issues prevents fine-grained approximation of the true $\{p_i\}$ values and even the *Q* procedure does not come close to the theoretical optimum. When the number of issues is increased to 100, we find that *GF* outperforms *SMR* by approximately 0.21 and the procedure *Q* outperforms *GF* by approximately 0.04, quickly reaching accuracy of 0.94.

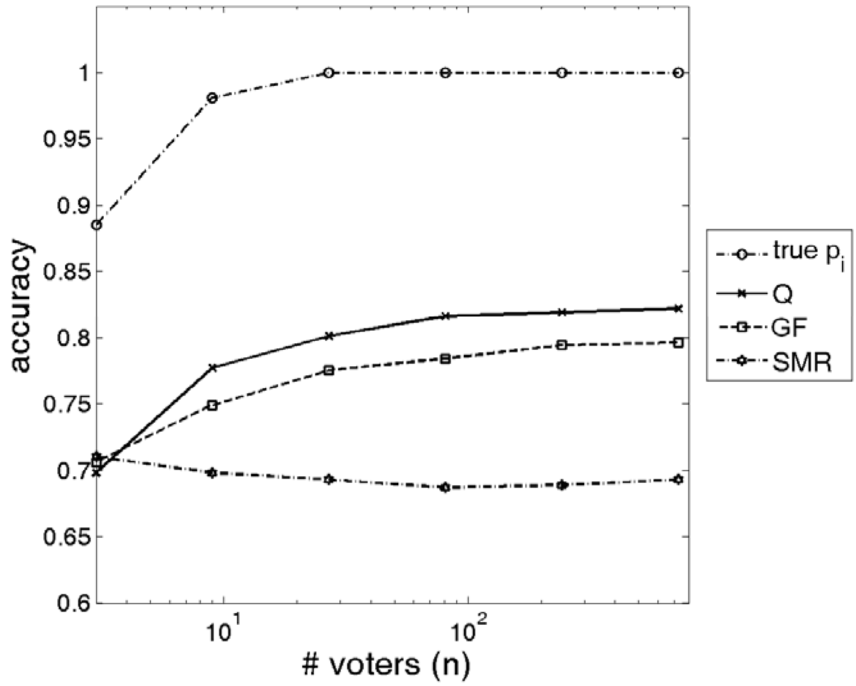

Figure 1a: $m = 10,\ p \sim U[0,1]$

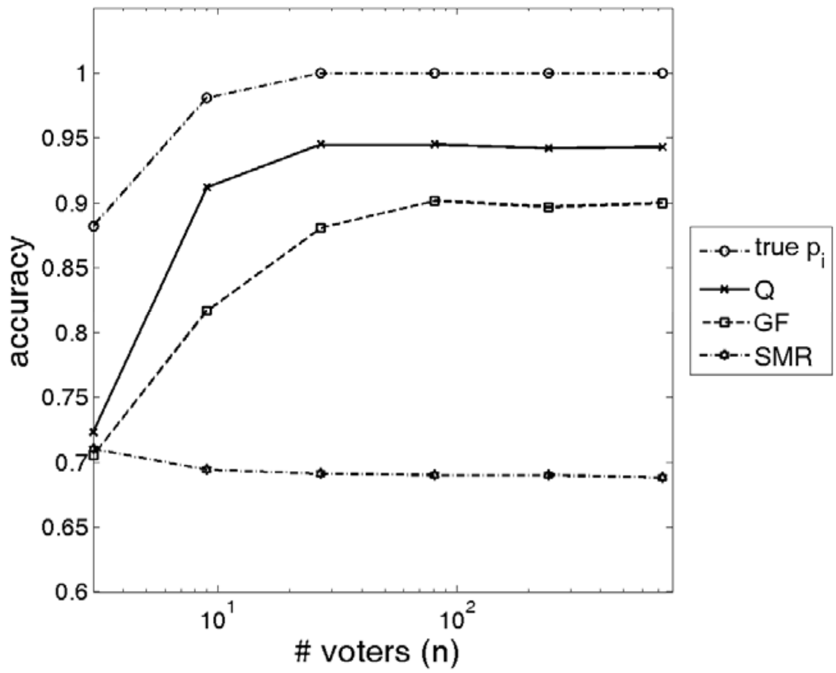

Figure 1b: $m = 100,\ p \sim U[0,1]$

The next simulation presents the effect of the number of issues, varying this number from 1 to 1000. The number of voters is assumed to be constant equal to 5 (Figure 2a) and 500 (Figure 2b) – representing committees and constituencies, respectively. Since SMR is independent of the number of issues, its performance remains constant as *m* grows. As was already evident in Figures 1a and 1b, for both *n*=5 and *n*=500, it holds constant at approximately 0.69. By contrast, the accuracy of both the *GF* and *Q* procedures is positively affected by increasing *m*, due to the increased ability to estimate voters' skills. Moreover, the greater the number of voters the greater the ability of the *GF* and *Q* procedures to leverage differential voter skills.

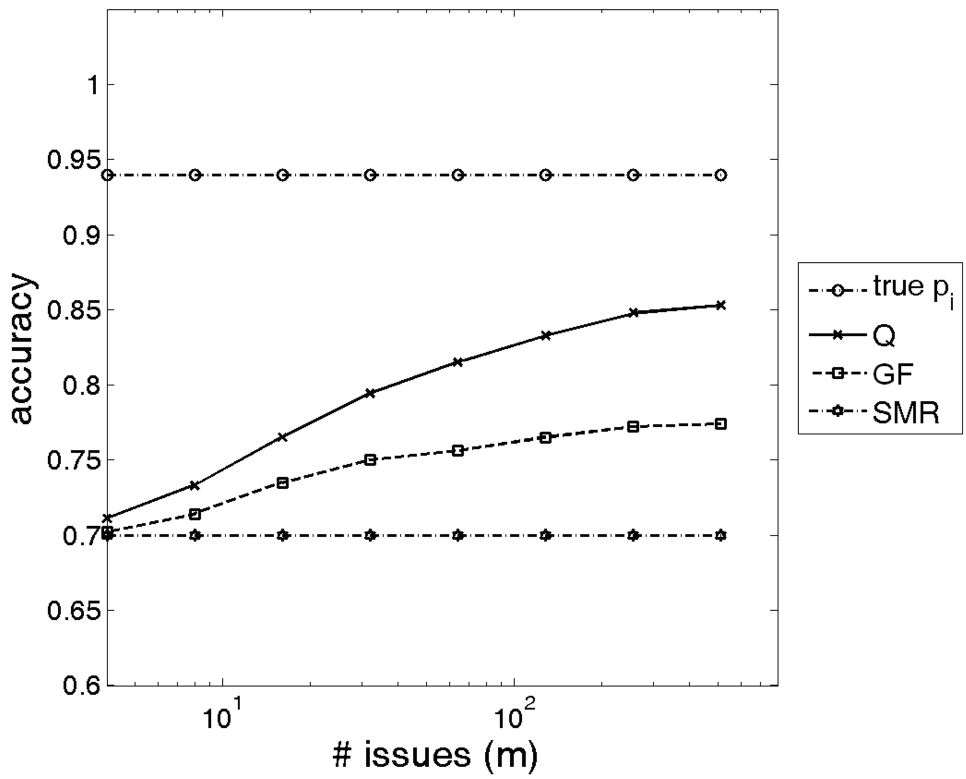

Figure 2a: $n = 5,\ p \sim U[0,1]$

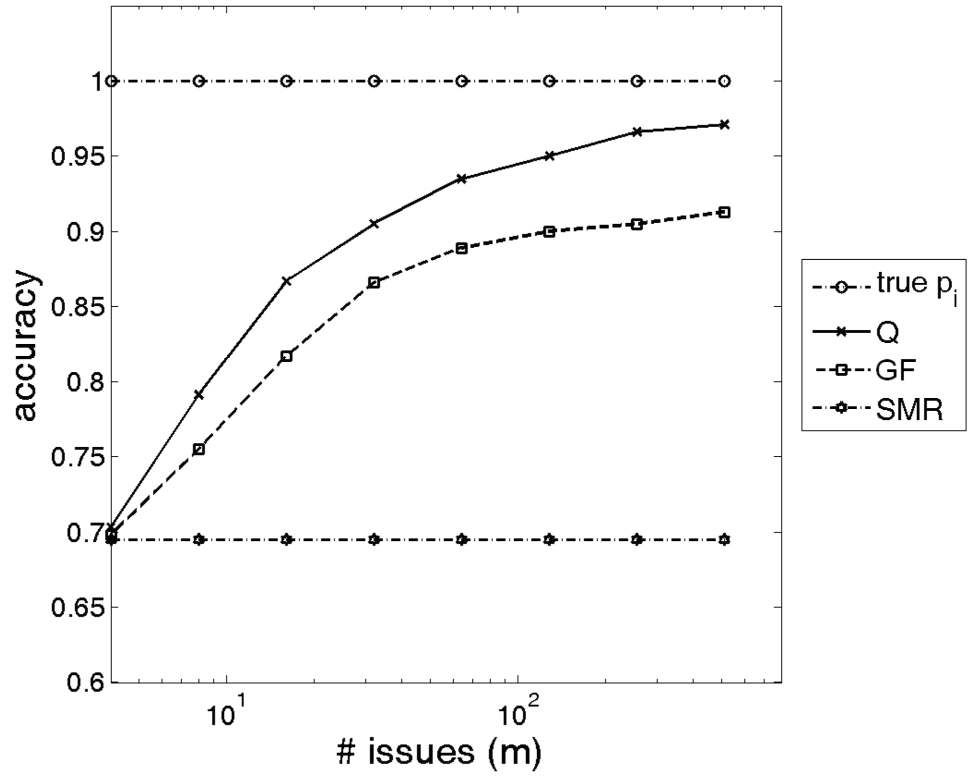

Figure 2b: $n = 500,\ p \sim U[0,1]$

In each of the above cases, the voters' skills parameters were sampled uniformly in the range [0,1], subject to the unanimity constraint. However, in many cases skills are normally distributed. In Figure 3, we show the results for *Q*, as in Figure 2b, comparing sampling of p_i from a uniform distribution in [0,1] with sampling from a normal distribution (truncated with lower and upper bounds of 0 and 1, respectively) with mean of ½ and a variety of variances. As might be expected, the greater the variance the faster the convergence as m increases, since *Q* is able to exploit the differences in skills to improve prediction. Note that a variance of 0.3 yields slightly better results than the ones obtained under a uniform distribution.

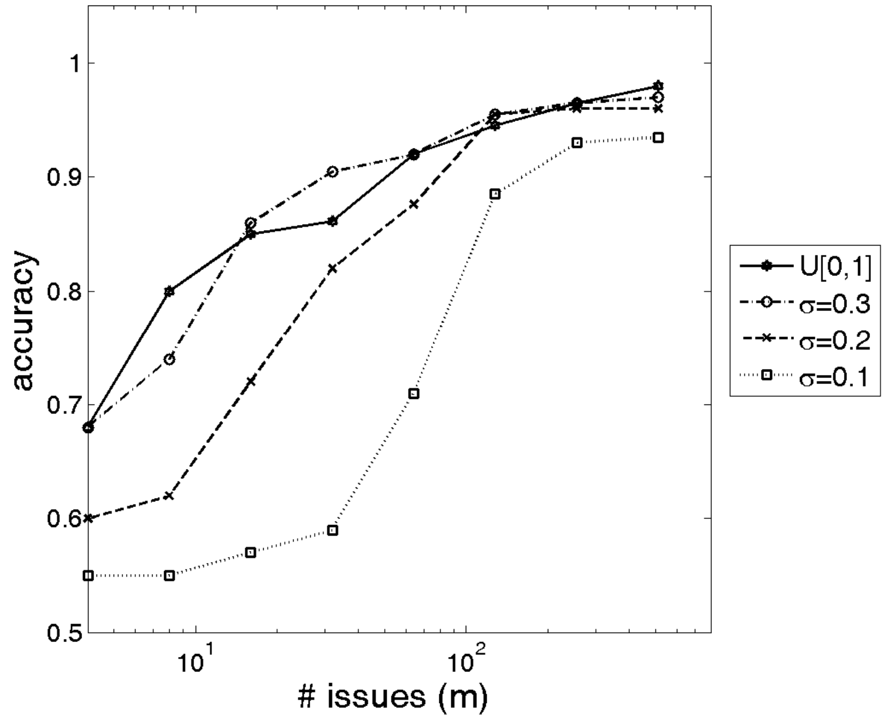

Figure 3: $n = 500$, *p* distribution varies as shown

Recall that the key to procedure *Q* is its ability to find values of θ that maximize the probability function $p(a;\theta)$. To gauge the extent to which *Q* succeeds in maximizing this function, we compare the value of $p(a;\theta)$ using the value θ_* obtained by *Q* with the value of $p(a;\theta_t)$ using the true value θ_t used to generate *a*. For the arbitrary parameter settings, *m* = 100 and *n* = 500, over 10,000 trials, we find that $p(a;\theta_*) > p(a;\theta_t)$ in 99.1% of the trials. This strongly suggests that *Q* is succeeding in finding global maxima in most cases.

This begs the question of why the accuracy of *Q* falls short of that obtained where the true value of θ is known. Note that for every assignment of probabilities, $\theta = \{p_i\}$, there is a dual assignment, $\overline{\theta} = \{1 - p_i\}$, such that $p(a;\overline{\theta}) = p(a;\theta)$. Thus, for every ``sensible'' solution, θ, there is a counter-intuitive one, $\overline{\theta}$. Examination of the errors made by *Q* reveals that most of the errors are in trials for which the dual solution of the true (actual) one is found. In these trials, every issue is assigned the wrong value. This systematic error occurs because of the weakness of our symmetry-breaking assumption requiring that $\prod_i p_i > \prod_i 1 - p_i$. When the difference between $\prod_i p_i$ and $\prod_i 1 - p_i$ is small, a slight discrepancy in estimating the values of {p_i} can result in the dual solution being chosen instead of the actual solution.

The weakness of our assumption can be demonstrated empirically; it can be verified that if the p_i values are chosen out of a uniform distribution in [0,1] subject solely to our assumption, then for every n the mean value of {p_i}, $\overline{p_i}$, satisfies $\overline{p_i} - 0.5 \approx \frac{c}{\sqrt{n}}$, where $c \approx 0.2$. It has been shown, however, by Berend and Paroush (1998) that a necessary (and sufficient) condition for Condorcet's Jury Theorem to hold is that $\lim_{n \to \infty} \left(\overline{p_i} - 0.5\right)\sqrt{n} = \infty$. Thus, by that result, Condorcet's Jury Theorem does not hold for a skill distribution such as we consider here, a fact we have already noted empirically. It might, therefore, be of interest to consider a stronger symmetry-breaking assumption.

5. Simulation Results: Asymmetric Skill Distributions

Consider now the case in which the {p_i} values are chosen from a uniform distribution in [0.1,1] instead of in [0,1], which implies that there are no voters who are (almost) always wrong. In this case, the mean value of {p_i} will typically be approximately 0.55. Figures 4a and 4b show the simulation results for m = 10 and m = 100, respectively, as n varies. As can be seen, *SMR* converges to 1 as n grows. This is expected by an extension of Condorcet's Jury Theorem. Note, however, that even in this case, *SMR* is dominated by the Q procedure, which converges to accuracy of 1 considerably faster than *SMR*.

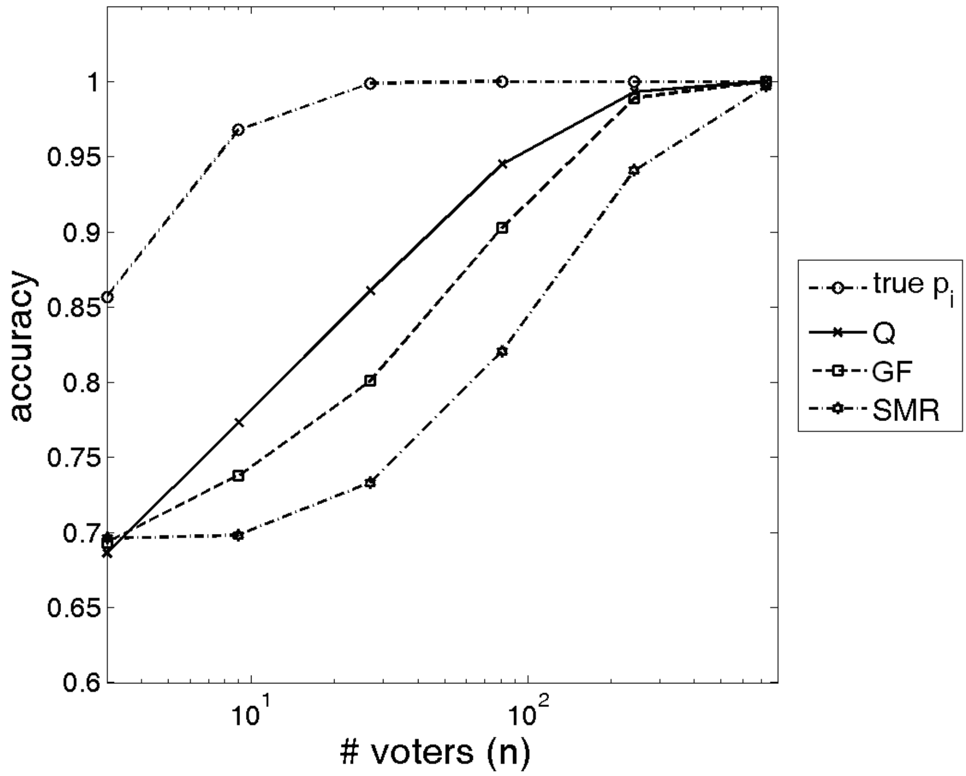

Figure 4a: $m = 10,\ p \sim U[0.1,1]$

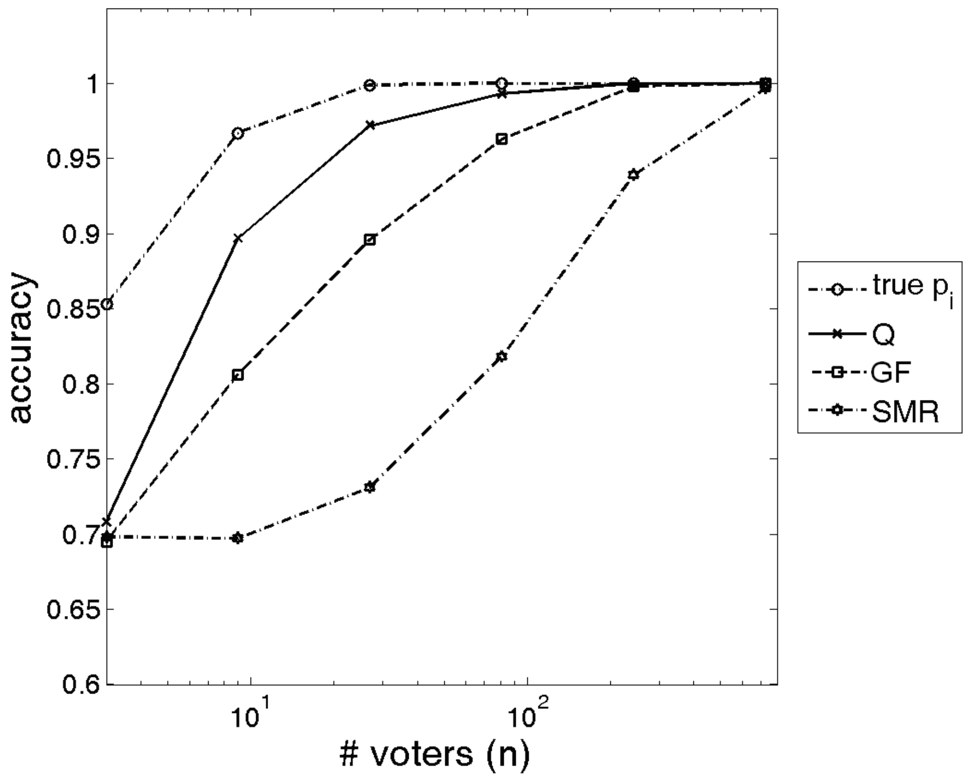

Figure 4b: $m = 100,\ p \sim U[0.1,1]$

Finally, we re-examine the results presented in Figures 2a and 2b, by considering the uniform distribution in [0.1,1]. As in the case of the uniform distribution in [0,1], *SMR* is unaffected by increasing *m*. However, the accuracy of the *GF* and *Q* procedures increases with *m*, converging quite dramatically to near perfect accuracy in the case of *n* = 500.

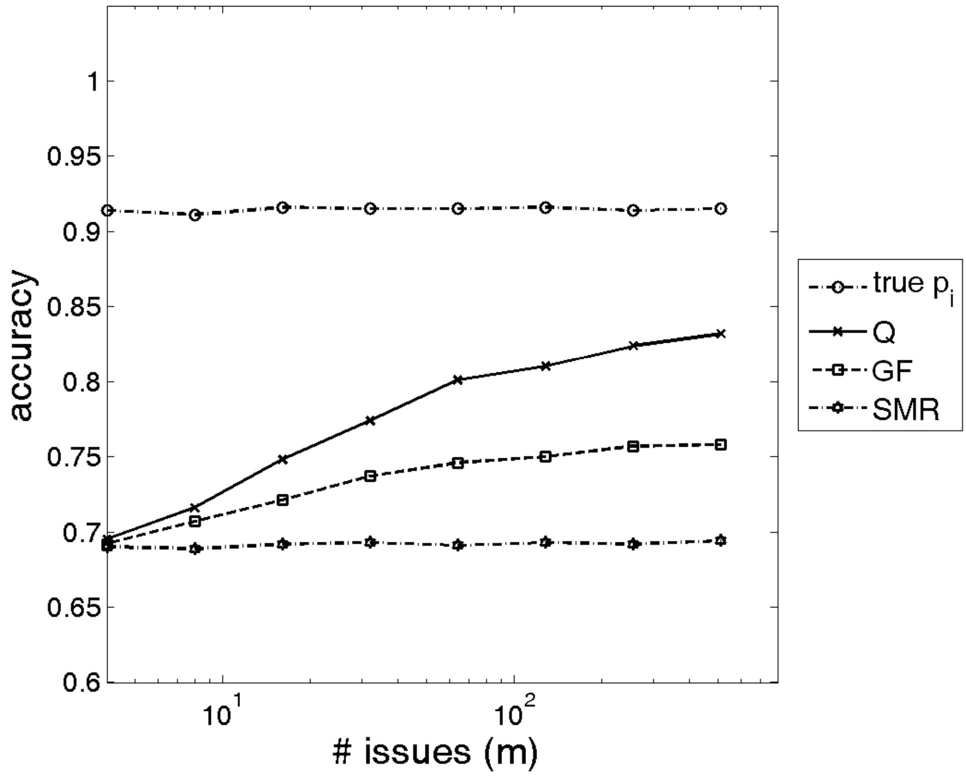

Figure 5a: $n = 5,\ p \sim U[0.1,1]$

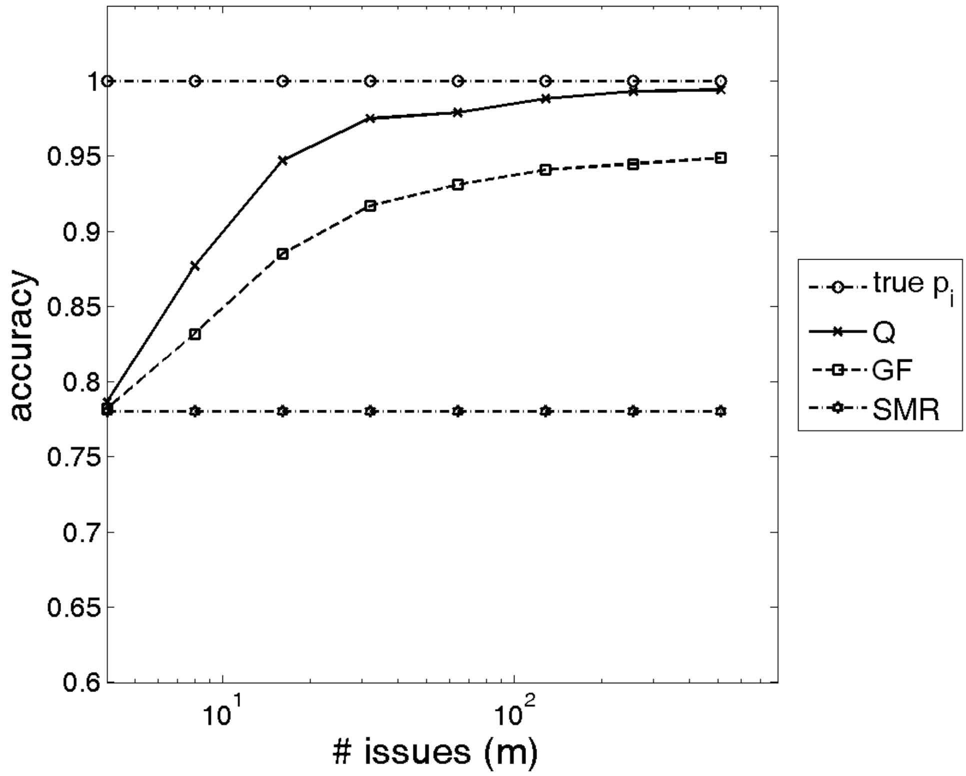

Figure 5b: $n = 500,\ p \sim U[0.1,1]$

To sum up, we find that our method is effective even under the most minimal assumption regarding voter skills: a unanimous vote must be more likely right than wrong. This assumption is considerably weaker than the voter skills assumption made by Condorcet and also weaker than the assumptions subsequently shown to be necessary and sufficient for the Condorcet Jury Theorem to hold (Berend and Paroush 1998). Moreover, we have shown that the *Q* procedure outperforms *GF* and *SMR* both under sufficient skills assumptions where Condorcet's Jury theorem holds and under our assumptions where it does not hold.

6. Conclusions

This paper lies at the confluence of two fields: voting theory and stochastic optimization theory. The *Q* procedure is, as we have noted, an instantiation of the EM algorithm, one of the best-known tools in stochastic optimization theory. The EM procedure is notably useful in data clustering, medicine, natural language processing, and a variety of other fields. It has been shown that the framework in which EM is used – optimizing values where information is missing – is appropriate to the problem of collective-decision making. In this case, the values to be optimized are estimations of voter skills and the missing information are the true outcomes. In short, EM can be profitably applied to the theory of voting and collective decision-making, where it can be used to achieve optimal resource utilization through maximum-likelihood experience-based estimation of the probabilities reflecting individual decisional skills and the assignment of the corresponding appropriate decisional weights. The proposed application of the EM procedure fills a significant gap in the literature on optimal collective judgment aggregation that was stimulated by Condorcet's (1785) approach and his celebrated jury theorem (Black 1958). When there are only some lower bounds on voters' skills that do not differentiate among individual voters, Condorcet's Jury theorem deals with the optimal aggregation rule, namely, *SMR*. However, when information regarding the respective probabilities of individual voters' making correct judgments is available, this additional information can be exploited through the optimal assignment of weights to individual voters according to their skills (Nitzan and Paroush 1982; Shapley and Grofman 1984). We have shown that optimization is possible even when the unrealistic assumption that voters' skills are known does not hold. By our main results, when voters' decisions on a multiplicity of issues are known, there is no need to know the "correct" decision in order to estimate individual voter skills. It is thus possible to overcome the informational obstacle in optimal collective decision methods which can be invoked without making unrealistic demands for information. This brings the generalization of

Condorcet's approach to a “happy end”: the optimal rule is identifiable even under common circumstances.

Note that in Condorcet's setting, the maximum likelihood estimation is applied to the likelihood of their single observed set of judgments relative to the two (unobserved) states of the world, assuming that voters' decisional skills are equal. From these estimated likelihoods, we infer that the most likely state is the one that would have produced the observation with the higher probability (Young 1988, 1995). The same approach is used in the more general settings of Condorcet's followers, making alternative assumptions on the voters' decisional skills. In our more general setting, the maximum likelihood estimation is applied given the observed judgments of the voters on *m* issues. The estimation consists of two stages. In the first stage, the unobserved voters' decisional skills are estimated and in the second stage, the likelihood of their single, m^{th} observed set of judgments is estimated in the two (unobserved) states of the world, assuming that voters' probabilities to make the correct decision are those obtained in the first stage. From these estimated likelihoods we infer that the most likely state is the one that would have produced the m^{th} observation with the higher probability.

The situation in which voter decisions on multiple issues are available is a common one both in the context of voting and in the context of expert judgments. Although the natural setting for our approach is the one in which voting is "a collective quest for truth" (Young 1995), we can also apply our algorithm to cases in which voting is "a compromise between conflicting values" (Young 1995), that is, where we wish to reach consensus among voters whose individual preferences are expressed by approving or rejecting each of a multiplicity of proposals or candidates. Thus, for example, every instance in which voters are asked to simultaneously approve or reject each of a multiplicity of proposals or candidates is amenable to the analysis presented here and to the application of the proposed optimal aggregation method based on the *Q* procedure. This is the balloting method used in approval voting (Brams and Fishburn 1978, 2005), in which the object is to determine some set of winners when voters' dichotomous preferences are defined on the set of alternative candidates. Thus, like Brams et al. (2007), we are, in effect, proposing an alternative aggregation function for the approval voting balloting method. In this context, however, the fact that voters are assigned different weights, and that some might even be assigned negative weights, plainly runs into political, institutional, structural, psychological and cultural difficulties.

Similarly, this method can be used in voting bodies, such as parliaments, where open voting takes place on a regular basis. Moreover, the use of the Internet for assembling positive or negative user judgments on products has become commonplace. The *Q* procedure can be used to optimally aggregate these individual judgments into an optimal overall judgment. More broadly, the method is applicable wherever ongoing decisions are required in uncertain dichotomous choice settings, in legal, medical, economic, political (Miller 1996) and other contexts.

References

Austen-Smith, D., and J.S. Banks. 1996. "Information Aggregation, Rationality and the Condorcet Jury Theorem." *American Political Science Review* 90(1): 34-45.

Baharad, E., J. Goldberger, M. Koppel, and S. Nitzan. 2010. "Distilling the Wisdom of Crowds: Weighted Aggregation of Decisions on Multiple Issues" *Journal of Autonomous Agents and Multi-Agent Systems,* to appear

Ben-Yashar, R., and I. Milchtaich. 2007. "First and Second Best Voting Rules in Committees." *Social Choice and Welfare* 29(3): 453-80.

Ben-Yashar, R., and S. Nitzan. 1997. "The Optimal Decision Rule for Fixed-Size Committees in Dichotomous Choice Situations: The General Result." *International Economic Review* 38(1): 175-86.

Ben-Yashar, R., and J. Paroush. 2000. "A Non-asymptotic Condorcet Jury Theory." *Social Choice and Welfare* 17: 189-99.

Berend, D., and J. Paroush. 1998. "When is Condorcet's Jury Theorem Valid." *Social Choice and Welfare* 15: 481-88.

Berend, D., and L. Sapir. 2005. "Monotonicity in Condorcet Jury Theorems." *Social Choice and Welfare* 24 (1): 83-92.

Black, D. 1958. *The Theory of Committees and Elections*. Cambridge: Cambridge University Press.

Brams, S., and P. Fishburn. 2005. "Going from Theory to Practice: The Mixed Success of Approval Voting." *Social Choice and Welfare* 25: 457-74.

Brams, S., and P. Fishburn. 1978. "Approval Voting." *American Political Science Review* 72(3): 831-47.

Brams, S.J., D.M. Kilgour and M.R. Sanver. 2007. "A Minimax Procedure for Electing Committees." *Public Choice* 132(3): 401-20.

Condorcet, N.C. de. 1785. *Essai sur l'application de l'analyse a la probabilite des decisions rendues a la pluralite des voix*, Paris, 20: 27-32.

Coughlan, P.J. 2000. "In Defense of Unanimous Verdicts: Mistrials, Communication and Strategic Voting." *American Political Science Review* 94: 375-93.

Dempster, A. P., N. M. Laird and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society. Series B (Methodological)* 39(1): 1-38.

Feddersen, T., and W. Pesendorfer. 1998. "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts Under Strategic Voting." *American Political Science Review* 92: 23-35.

Grofman, B., and S.L. Feld. 1983. "Determining Optimal Weights for Expert Judgment." in *Information Pooling and Group Decision Making: Proceedings of the Second University of California, Irvine, Conference on Political Economy*, Edited by Grofman, B. and G. Owen, JAI Press Inc.

Grofman, B., G. Owen and S. Feld. 1983. "Thirteen Theorems in Search of the Truth." *Theory and Decision* 15: 261-78.

Karotkin, D. 1994. "Effect of the Size of the Bench on the Correctness of Court Judgments: The Case of Israel." *International Review of Law and Economics* 14: 371-75.

List, C., and P. Petit. 2002. "Aggregating Sets of Judgments: An Impossibility Result." *Economics and Philosophy* 18: 89-110.

McLennan, A. 1998. "Consequences of the Condorcet Jury Theorem for Beneficial Information Aggregation by Rational Agents." *American Political Science Review* 92: 413-18.

Miller, N. 1996. "Information, Individual Errors, and Collective Performance: Empirical Evidence on the Condorcet Jury Theorem." *Group Decision and Negotiation* 5: 211-28.

Nitzan, S., and J. Paroush. 1985. *Collective Decision Making: An Economic Outlook*, Cambridge University Press, Cambridge, England.

Nitzan, S., and J. Paroush. 1982. “Optimal Decision Rules in Uncertain Dichotomous Choice Situations.” *International Economic Review* 23(2): 289-97.

Nurmi, H. 2002. *Voting Procedures Under Uncertainty*. Springer, Berlin-Heidelberg.

Owen, G., B. Grofman and S. Feld. 1989. "Proving a Distribution Free Generalization of the Condorcet Jury Theorem." *Mathematical Social Sciences* 17: 1-16.

Paroush, J. 1998. "Stay Away from Fair Coins: A Condorcet Jury Theorem." *Social Choice and Welfare* 15: 15-20.

Shapley, L. and B. Grofman. 1984. "Optimizing Group Judgmental Accuracy in the Presence of Interdependencies." *Public Choice* 43: 329-343.

Warfield, S.K., Zou, K.H., Wells, W.M. 2004. "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation." *IEEE Transactions on Medical Imaging* 23(7): 903-921.

Young, P. 1996. "Group Choice and Individual Judgment", in *Perspectives on Public Choice*, Dennis Mueller ed. New York: Cambridge University Press.

Young, P. 1995. "Optimal Voting Rules." *Journal of Economic Perspectives* 9: 51-64.

Young, P. 1988. "Condorcet's Theory of Voting." *American Political Science Review* 82(4): 1231-44.