

RANGE OF ASYMPTOTIC BEHAVIOUR OF THE OPTIMALITY PROBABILITY OF THE EXPERT AND MAJORITY RULES

DANIEL BEREND* AND
LUBA SAPIR,** *Ben-Gurion University*

Abstract

We study the uncertain dichotomous choice model. In this model, a group of expert decision makers is required to select one of two alternatives. The applications of this model are relevant to a wide variety of areas. A decision rule translates the individual opinions of the members into a group decision, and is optimal if it maximizes the probability of the group making a correct choice. In this paper, we assume the correctness probabilities of the experts to be independent random variables selected from some given distribution. Moreover, the ranking of the members in the group is (at least partly) known. Thus, one can follow rules based on this ranking. The extremes are the expert rule and the majority rule. The probabilities of the two extreme rules being optimal were compared in a series of early papers, for a variety of distributions. In most cases, the asymptotic behaviours of the probabilities of the two extreme rules followed the same patterns. Do these patterns hold in general? If not, what are the ranges of possible asymptotic behaviours of the probabilities of the two extreme rules being optimal? In this paper, we provide satisfactory answers to these questions.

Keywords: Dichotomous choice; decision rule; expert rule; majority rule; optimality probability; partial information

2000 Mathematics Subject Classification: Primary 91B06
Secondary 91B12; 90B50

1. Introduction

1.1. Background

There are many situations where a group of expert decision makers is required to select one of two alternatives, of which exactly one is regarded as correct. A decision rule is a rule for translating the individual opinions of the members into a group decision. The decision skill of each expert is characterized by the individual's probability of making the right choice.

There are several aspects to the study of this so-called dichotomous choice model. One of them is dealing with the Condorcet jury theorem in various setups. Condorcet [14] believed that a group of individuals facing a binary choice and using a simple majority rule is likely to make the correct choice. Moreover, this likelihood tends to complete certainty as the number of members of the group tends to infinity (see [2]). A Condorcet jury theorem is a formulation of conditions substantiating this belief. The classical conditions of this theorem assume the

Received 22 February 2005; revision received 8 September 2005.

* Postal address: Departments of Mathematics and of Computer Science, Ben-Gurion University, Beer-Sheva, 84105, Israel. Email address: berend@cs.bgu.ac.il

** Postal address: Department of Industrial Engineering and Management, Ben-Gurion University, Beer-Sheva, 84105, Israel. Email address: lsapir@bgu.ac.il

independence of the decision makers and the same value $p > \frac{1}{2}$ of the individual correctness probabilities. Attempts to generalize the theorem in a variety of ways have been made (see [16], [19], [25], [12], [13], [10], [18], [1], [21], [5], [11], [3], [9], and [24]).

In some situations, the group is free to choose the decision rule, and the identification of the optimal rule is of primary importance (see [23], [20], [16], [15], [17], and [8]). If the probability of each member making the right choice is known and the alternatives are symmetric, then the optimal decision rule is a weighted majority rule (see [20]).

The starting point of this paper is the problem of identifying the optimal decision rule under partial information of the decision skills. Specifically, we assume the correctness probabilities of the group members to be independent random variables distributed according to some given distribution rule. Moreover, while the values these variables take are unknown, we assume that the ranking of the members in terms of their individual correctness probabilities is known. Thus, we can follow rules based on this ranking. The poles (or extremes) are the expert rule – following the advice of the most qualified individual while ignoring all the rest – and the majority rule – always taking the majority advice, even when advocated by mostly less qualified group members. The probabilities of these two ‘polar’ rules being optimal were compared in a series of early papers, for a variety of distributions. In most cases, the asymptotic behaviours of the probabilities of the two polar rules followed the same patterns. Do these patterns hold in general? If not, what are the ranges of possible asymptotic behaviours of the probabilities of the two extreme rules being optimal? In this paper we provide satisfactory answers to these questions.

Before turning to the main results, we present our model more precisely (in Section 1.2) and review previous results on our problem (in Section 1.3), which allows us to compare the optimality probabilities of the expert and majority rules for various types of distribution. Section 2, where we investigate the range of possible asymptotic behaviours of the optimality probabilities, contains the main results. In Section 3, we provide the proofs.

1.2. Setup and notation

A committee consisting of n members is required to select one of two alternatives, +1 and -1, of which exactly one is correct. The alternatives are symmetric, that is, a-priori equally likely to be correct, and the benefit (or loss) associated with a success (correct choice) or a failure (incorrect choice) is independent of the particular alternative chosen. We assume that the members are independent in their choices; p_i is the probability of the i th expert making the right choice. The vector $\mathbf{p} = (p_1, p_2, \dots, p_n)$ is the *vector of abilities* (or *skills*) or the *competence structure*.

A *decision rule* is a rule for translating the individual opinions into a group decision. The number of all possible decision rules in a group of size n is 2^{2^n} . A decision rule is *optimal* if it maximizes the probability of the group making the correct decision for all possible combinations of opinions. If the members indexed by some subset $A \subseteq \{1, 2, \dots, n\}$ of the group recommend the first alternative, while those indexed by $B = \{1, \dots, n\} \setminus A$ recommend the second, then (see [20]) the first alternative should be chosen if and only if

$$\prod_{i \in A} \frac{p_i}{1 - p_i} \geq \prod_{i \in B} \frac{p_i}{1 - p_i} \tag{1}$$

or, equivalently,

$$\sum_{i \in A} \log \frac{p_i}{1 - p_i} \geq \sum_{i \in B} \log \frac{p_i}{1 - p_i}. \tag{2}$$

In view of (1) and (2), it is natural to define the *odds* of an individual, whose probability of being correct is p , as $p/(1 - p)$, and the *logit transform* of p as $\log[p/(1 - p)]$. It will be convenient to consider the functions T and logit , defined by

$$T(p) = \frac{p}{1 - p},$$

$$\text{logit}(p) = \log \frac{p}{1 - p} = \log T(p).$$

In other words, the optimal decision rule for a committee with known competence structure \mathbf{p} is given by

$$\hat{R} = \text{sgn} \left(\sum_{i=1}^n w_i x_i \right), \tag{3}$$

where $w_i = \text{logit}(p_i)$ and x_i is the individual choice of the i th decision maker, defined by $x_i = +1$ for the first alternative and $x_i = -1$ for the second. With this notation, (3) reduces to a *weighted majority rule* with weights $w_i = \text{logit}(p_i)$.

The assumption of full information regarding the decision maker’s competence is very restrictive and often far from being fulfilled. Therefore, a model incorporating incomplete information seems to be more realistic. Thus, we assume that the correctness probabilities p_i are distributed according to some known distribution law. This distribution is always supported on $[\frac{1}{2}, 1]$, and the p_i are independent. While we do not know the exact values the p_i take, we assume that the ranking of the members is known, so that we can follow rules based on this ranking. The poles are the expert rule and the simple majority rule. Clearly, there are numerous other decision rules between these two extremes. A comparison between the polar rules raises several questions. One concerns the nonasymptotic part of the Condorcet jury theorem: for which competence structures $\mathbf{p} = (p_1, p_2, \dots, p_n)$, of an odd-sized committee, is the probability of the committee making the correct choice when utilizing the simple majority rule larger than the correctness probability of any of its members, i.e. for which values of \mathbf{p} does

$$\sum_{y_1 + \dots + y_n \geq (n+1)/2} \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} \geq \max_{1 \leq i \leq n} p_i$$

hold, where $y_i = 1$ if the i th expert is correct and $y_i = 0$ otherwise? If the p_i are independent random variables distributed according to some known distribution, one might be interested in

$$\mathbb{P} \left(\sum_{y_1 + \dots + y_n \geq (n+1)/2} \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} \geq \max_{1 \leq i \leq n} p_i \right).$$

In this paper, we consider another question, concerned with the probability of the two polar rules being optimal. It is easily verified that the probability, $P_e(n)$, of the expert rule being optimal is the probability that, when the top expert disagrees with all other experts, the top expert is more likely to be correct than the others. The probability, $P_m(n)$, of the majority rule (defined, strictly speaking, only for odd $n = 2s + 1$) being optimal is equal to the probability that the $s + 1$ lowest ranked experts, when opposed by the s highest ranked experts, are more likely to be correct. Note that we are concentrating on the probability of a polar rule being optimal, rather than the probability of it being right. The probability of a decision rule being right may

TABLE 1: Optimality probabilities of all weighted majority rules for $n \leq 5$.

n	Rule and corresponding probability						
3	(1,0,0) 0.675	(1,1,1) 0.325	—	—	—	—	—
4	(1,0,0,0) 0.373	—	(1,1,1,0) 0.277	—	(2,1,1,1) 0.350	—	—
5	(1,0,0,0,0) 0.199	(1,1,1,1,1) 0.022	(1,1,1,0,0) 0.175	(3,1,1,1,1) 0.107	(2,1,1,1,0) 0.229	(3,2,2,1,1) 0.194	(2,2,1,1,1) 0.074

be expected to be quite high, and in fact to converge to 1 as the number of experts increases, as long as a ‘reasonable’ decision rule is employed. However, the probabilities correspond to the ‘average case’, whereas in this paper we deal with the ‘worst case’. That is, each decision rule has some *borderline* cases. While we may usually expect the simple majority rule to lead to the correct decision, we should hesitate to use it if, say, in a committee comprising eleven members, the six members known to be least qualified happen to favour one view while all five more qualified members hold the opposite view. Similarly, employing the expert rule would seem strange if we happen to be in the borderline situation in which the top expert is opposed by all the others. To claim that, in a specific case, the majority rule or the expert rule is optimal is tantamount to asserting that we should indeed favour the opinion of the six members over that of the five in the first example, or of the top expert in the second example. Consequently, by comparing the probabilities of the rules being optimal, as we do in this paper, we cannot conclude that one rule is better than the other. Rather, the comparison provides us with a view of the performance of the rules in question in some extreme cases, and hints to what extent we should modify them in those cases.

1.3. History and previous results

A general comprehensive study of weighted majority rules is a very complicated task, since the class of such rules becomes very large as the number of group members increases. For example, for a committee of size $n = 3$ it includes two weighted majority rules; for $n = 4$, three rules; for $n = 5$, seven rules; for $n = 6$, twenty-one rules; for $n = 7$, 135 rules; for $n = 8$, 2470 rules; and for $n = 9$, 172 958 rules (see [15] and [17]). Note that a weighted majority rule can be represented by many systems of weights. For instance, for $n = 3$, both systems of weights (1, 1, 1) and (4, 3, 2) define the simple majority rule, since in each of them any sum of two of the weights is larger than the remaining weight. Similarly, both (1, 0, 0) and (6, 3, 2) define the expert rule, since the weight of the most qualified expert is larger than the sum of weights of the other two.

Table 1 provides the probabilities of weighted majority rules to be optimal under the assumption of the uniform distribution on $[\frac{1}{2}, 1]$ of the correctness probabilities, as found by Nitzan and Paroush [20] (using standard Monte Carlo simulation). Each cell corresponds to the rule shown, and contains its probability of being optimal.

The comparison of these probabilities hints that the expert rule is more likely to be optimal than the majority rule even for quite small values of n .

The first explicit expression for $P_e(n)$ was due to Nitzan and Paroush [20], who studied the case of the log-normal distribution for the individual correctness probabilities. Below we list most of the results regarding the behaviour of $P_e(n)$ and $P_m(n)$ subsequently obtained. Recall that $n = 2s + 1$ in the majority rule.

1. $p_i \sim \text{uniform}[\frac{1}{2}, 1]$.

Expert rule. $nB_1(n) \leq P_e(n) \leq 2nI_1^{n-1}$, where

$$I_1^{n-1} = (2 \log 2 - 1)^{n-1} \quad \text{and} \quad B_1(n) = 2I_1^{n-1} - 2(3 - 4 \log 2)^{n-1}.$$

Majority rule. $P_m(n) \leq \frac{n}{3^s(s!)^2} \leq \frac{4^n}{n^n}$.

2. p_i is supported on $[\frac{1}{2}, 1]$ with density $\rho_p(x) = 2\alpha(2x - 1)^{\alpha-1}$, where $\alpha > 0$.

Expert rule. (i) If $0 < \alpha < \frac{1}{2}$ then $nB_1(\alpha, n) \leq P_e(n) \leq nB_2(\alpha, n)$, where

$$B_1(\alpha, n) = 2\alpha \mathcal{I}_1^{n-1} - 2\alpha^2 \mathcal{I}_2^{n-1}, \quad B_2(\alpha, n) = B_1(\alpha, n) + \mathcal{I}_3^{n-1},$$

$$\mathcal{I}_k = 2\alpha \int_0^1 \frac{(1-t)^{\alpha-1} t^k}{(1+t)^{\alpha+1}} dt, \quad k = 1, 2, 3.$$

(ii) If $\alpha \geq \frac{1}{2}$ then $nB_1(\alpha, n) \leq P_e(n) \leq 2\alpha n \mathcal{I}_1^{n-1}$.

Majority rule. (i) If $0 < \alpha \leq 1$ then $P_m(n) \leq \alpha^s / (s!)^2$.

(ii) If $\alpha > 1$ then $P_m(n) \leq (n/(s!)^2)(\frac{1}{2}\alpha)^s$.

3. $\text{logit}(p_i) \sim \text{Exp}(\lambda)$, where $\lambda > 0$.

Expert rule. $P_e(n) = n/2^{n-1}$.

Majority rule. $P_m(n) = \frac{\binom{n-1}{s}}{(n^2-1)^s} \approx \sqrt{\frac{n}{2\pi}} \left(\frac{2}{n}\right)^n$.

4. $T_i = (1 - p_i)/p_i$ is supported on $[0, 1]$, with distribution function $G_T(x) = \sum_{k=1}^{\infty} c_k x^k$.

Expert rule. If $\sum_{k=1}^{\infty} c_k = 1$ or $\sum_{k=1}^{\infty} |c_k| E^{n-1}(T^k) < \infty$, where the powers of $E(T^k)$ are the moments of the distribution, then

$$P_e(n) = n \sum_{k=1}^{\infty} c_k E^{n-1}(T^k).$$

Thus,

$$P_e(n) = n c_{k_0} E^{n-1}(T^{k_0})(1 + o(1)),$$

where $c_{k_0} > 0$ is the first nonzero term in $(c_k)_{k=1}^{\infty}$.

Majority rule. None available.

5. $\text{logit}(p_i) \sim \text{uniform}[0, a]$, where $a > 0$.

Expert rule. $P_e(n) = 1/(n - 1)!$.

Majority rule. None available.

We consider these results part by part. Berend and Harmse [4] (part 1) obtained an explicit formula for $P_e(n)$ and an upper bound for $P_m(n)$ for the uniform distribution on $[\frac{1}{2}, 1]$. The combination of these results implies that the latter probability decays to 0 much faster than the former. The same conclusion holds for the generalized uniform distribution, considered by Berend and Sapir [7] (part 2). Sapir [22] explored the situation of logit transforms distributed exponentially (part 3). A comparison of $P_e(n)$ and $P_m(n)$ shows that, again, the expert rule has

a much better chance of being optimal than the majority rule. Berend and Sapir [6] calculated $P_e(n)$ for a general family of distributions, defined in part 4. This family contains as special cases a wide variety of distribution families, for example the family of distributions in which the density function of the correctness probability is a polynomial (on the interval $[\frac{1}{2}, 1]$), and, thus, as a very particular instance, the uniform distribution. Their main point was that $P_e(n)$, considered as a function of the group size, decreases to 0 exponentially fast for a very wide family of distributions. The only example with different behaviour is the distribution in part 5; there $P_e(n)$ decays faster than exponentially.

1.4. Main question

The results above suggest that $P_e(n)$ and $P_m(n)$ have typical asymptotic behaviours. Namely, $P_e(n)$ decreases as c^n for some $0 < c < 1$, while $P_m(n)$ does so (faster) as $(c'/n)^n$ for some $c' > 0$. This raises the following questions. Are the above asymptotic behaviours valid in general? If not, what are the ranges of possible asymptotic behaviours of the probabilities of the two extreme rules being optimal?

In fact, as mentioned earlier, the results of part 5 show a somewhat different pattern: $P_e(n)$ decreases faster than exponentially. However, this case is special in the sense that p_i assumes values only in some subinterval $[\frac{1}{2}, h]$ of $[\frac{1}{2}, 1]$. When the distribution of p_i is supported on the whole interval $[\frac{1}{2}, 1]$, one may ask whether or not the probability $P_e(n)$ decays exponentially.

2. Main results

2.1. Quick summary of main results

It turns out that, for $P_e(n)$, the answer to the first question is that essentially everything is possible; on the one hand, it may converge to 0 arbitrarily fast, and on the other hand it may converge to 1. For $P_m(n)$ the range of possibilities is narrower, as $P_m(n)$ cannot decay faster than $(c/n)^n$, for some $c > 0$. However, it may decay slower. More precisely, we always have $P_m(n) \leq c^n$ for some c , $0 < c < 1$; if, in addition, the $\text{logit}(p_i)$ are selected from a continuous distribution, then $P_m(n)$ decreases faster than ε^n for any $\varepsilon > 0$. The list below summarizes our results, showing the extremes of possible behaviour of the expert and majority rules. Note that, under each line containing one of our positive results (each part (i)), we formulate a converse (each corresponding part (ii)), showing that the result is essentially the best possible. For the majority rule, it is convenient to consider continuous and discontinuous (but not necessarily discrete) distributions separately. We emphasize that, throughout the paper, the distribution in question is assumed to be nontrivial (i.e. it does not take a single value with probability 1).

(I) Expert rule.

Lower bound. None: $P_e(n)$ may decrease arbitrarily fast.

Upper bound. None: we may have $P_e(n) \rightarrow 1$ as $n \rightarrow \infty$.

(II) Majority rule for a continuous distribution.

Lower bound. (i) $P_m(n) \geq (c/n)^n$ for some $c > 0$.

(ii) For any $\varepsilon > 0$, there exists a distribution such that $P_m(n) \leq (\varepsilon/n)^n$.

Upper bound. (i) $P_m(n) = o(1)^n$.

(ii) For any $(h_n)_{n=3}^\infty$ with $0 < h_n < 1$ and $h_n \rightarrow 0$ as $n \rightarrow \infty$, there exists a distribution such that $P_m(n) \geq h_n^n$.

(III) Majority rule for a discontinuous distribution.

Lower bound. (i) $P_m(n) \geq c^n$ for some $0 < c < 1$.

(ii) For any $\varepsilon > 0$, there exists a distribution such that $P_m(n) \leq \varepsilon^n$.

Upper bound. (i) $P_m(n) \leq c^n$ for some $0 < c < 1$.

(ii) For any $\varepsilon < 1$, there exists a distribution such that $P_m(n) \geq \varepsilon^n$.

2.2. Expert rule

The nonexistence of a nontrivial lower bound for $P_e(n)$ is the result of the following theorem.

Theorem 1. *For any positive sequence $(h_n)_{n=3}^\infty$, there exists a continuous distribution function $G_{\text{logit}(p)}(x)$, strictly increasing on $[0, \infty)$, such that*

$$P_e(n) \leq h_n, \quad n \geq 3.$$

The requirement (here and in some of the following theorems) that $G_{\text{logit}(p)}(x)$ be strictly increasing on $[0, \infty)$ strengthens the theorem. In fact, it explains that the one example encountered earlier in which $P_e(n)$ decreases faster than exponentially (part 5 of the first list of results) is not to be attributed to the fact that $G_{\text{logit}(p)}(x)$ is there supported on a finite interval.

The next theorem shows that there exists no nontrivial upper bound either.

Theorem 2. *There exists a continuous distribution function $G_{\text{logit}(p)}(x)$, strictly increasing on $[0, \infty)$, for which*

$$P_e(n) \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

Remark 1. The constructions we provide in the proofs of Theorem 1 and 2 are of continuous distribution functions. Minor modifications provide discrete distributions with the same properties.

2.3. Simple majority rule

In view of previous results, we might guess that $P_m(n)$ is always asymptotically smaller than $P_e(n)$. Thus, taking into account Theorem 1, the following theorem may come as something of a surprise. Throughout this section n is odd, with $n = 2s + 1$.

Theorem 3. *For any distribution function $G_{\text{logit}(p)}(x)$ and any positive constant*

$$c < \sup_{x \geq 0} \left(x \overline{\lim}_{\delta \rightarrow 0} \frac{G_{\text{logit}(p)}(x + \delta) - G_{\text{logit}(p)}(x)}{\delta} \right),$$

we have

$$P_m(n) \geq \left(\frac{c}{s} \right)^n \tag{4}$$

for all sufficiently large n .

For certain distributions, the constant c in (4) may be arbitrarily large.

Remark 2. We may have

$$\sup_{x \geq 0} \left(x \overline{\lim}_{\delta \rightarrow 0} \frac{G_{\text{logit}(p)}(x + \delta) - G_{\text{logit}(p)}(x)}{\delta} \right) = \infty.$$

In this case, for all sufficiently large n , the theorem gives

$$P_m(n) \geq \left(\frac{c}{s}\right)^n$$

for any positive constant c .

Corollary 1. For any differentiable distribution function $G_{\text{logit}(p)}(x)$ with density

$$g_{\text{logit}(p)}(x) = G'_{\text{logit}(p)}(x),$$

and any positive constant $c < \sup_{x \geq 0} x g_{\text{logit}(p)}(x)$, we have

$$P_m(n) \geq \left(\frac{c}{s}\right)^n$$

for all sufficiently large n .

Example 1. For $p_i \sim \text{uniform}[\frac{1}{2}, 1]$, Corollary 1 combined with the upper bound found by Berend and Harmse [4] yields

$$\left(\frac{0.89}{n}\right)^n \leq P_m(n) \leq \left(\frac{4}{n}\right)^n, \quad n \geq n_0,$$

where n_0 is some positive odd number. Indeed,

$$G_{\text{logit}(p)}(x) = \frac{e^x - 1}{e^x + 1} \quad \text{and} \quad g_{\text{logit}(p)}(x) = \frac{2e^x}{(e^x + 1)^2} \quad \text{for } x \geq 0.$$

Hence, $\sup_{x \geq 0} x g_{\text{logit}(p)}(x) \approx 1.5435 \cdot g_{\text{logit}(p)}(1.5435) > 0.447$.

Example 2. For $\text{logit}(p_i) \sim \text{Exp}(\lambda)$, discussed in [22], we have

$$\sup_{x \geq 0} x g_{\text{logit}(p)}(x) = \sup_{x \geq 0} x e^{-x} = g_{\text{logit}(p)}(1) > 0.367$$

and, therefore,

$$P_m(n) \geq \left(\frac{0.734}{n}\right)^n, \quad n \geq n_0,$$

where n_0 is some positive odd number. Note that, in this case (see part 3 of our first list) we actually know the exact value of $P_m(n)$:

$$P_m(n) = \binom{n-1}{s} (n^2 - 1)^{-s} = \sqrt{\frac{n}{2\pi}} \left(\frac{2}{n}\right)^n \left(1 + O\left(\frac{1}{n}\right)\right).$$

The following theorem shows that the lower bound obtained in Theorem 3 cannot be qualitatively improved when we have a continuous distribution function $G_{\text{logit}(p)}$.

Theorem 4. For any $\varepsilon > 0$, there exists a continuous distribution function $G_{\text{logit}(p)}(x)$, strictly increasing on $[0, \infty)$, such that $P_m(n) \leq (\varepsilon/s)^n$ for all n .

The next theorem gives an upper bound for $P_m(n)$ for any continuous distribution function.

Theorem 5. For any continuous distribution function $G_{\logit(p)}(x)$ and any $\varepsilon > 0$, $P_m(n) \leq \varepsilon^n$ for sufficiently large n .

The next theorem shows that the upper bound obtained in Theorem 5 cannot be qualitatively improved.

Theorem 6. For any sequence $(h_n)_{n=3}^\infty$ with $0 < h_n < 1$ and $h_n \rightarrow 0$ as $n \rightarrow \infty$, there exists a continuous distribution function $G_{\logit(p)}(x)$, strictly increasing on $[0, \infty)$, such that $P_m(n) \geq h_n^n$ for every n .

We now turn to discontinuous distributions. It turns out that the extreme possible asymptotic behaviour of $P_m(n)$ is quite different in the discontinuous case than in the continuous one. Both bounds in the discontinuous case are larger than their continuous counterparts.

Theorem 7. Let $G_{\logit(p)}(x)$ be any (discontinuous) distribution function, and let

$$c = \max_{x \geq 0} P(\logit(p) = x).$$

Then $P_m(n) \geq c^n$ for each n .

Again, Theorem 7 cannot be qualitatively improved.

Theorem 8. For any $\varepsilon > 0$, there exists a discrete distribution function $G_{\logit(p)}(x)$ such that

$$P_m(n) \leq \varepsilon^n, \quad n \geq 3.$$

Remark 3. Theorem 8 follows easily from the next theorem if we consider n to be sufficiently large. We must provide a separate proof to demonstrate that it holds for all $n \geq 3$.

Theorem 9 gives an upper bound for $P_m(n)$.

Theorem 9. Let $G_{\logit(p)}(x)$ be any distribution function, and let

$$c = \max_{x \geq 0} P(\logit(p) = x).$$

Then, for any $\varepsilon > 0$,

$$P_m(n) \leq (c + \varepsilon)^n$$

for all sufficiently large n .

Remark 4. If $G_{\logit(p)}(x)$ is continuous then Theorem 9 reduces to Theorem 5.

As in the case of continuous distributions, the next theorem shows that the upper bound obtained in Theorem 9 cannot be qualitatively improved for a discontinuous distribution function $G_{\logit(p)}$.

Theorem 10. There exists a discontinuous distribution function $G_{\logit(p)}(x)$ such that, for any positive $\varepsilon < 1$, $P_m(n) \geq \varepsilon^n$ for all n .

3. Proofs

Throughout this section, let Y_i , $i = 1, 2, \dots, n$, be the order statistics of $\logit(p_i)$, such that $Y_1 \leq Y_2 \leq \dots \leq Y_n$.

3.1. Expert rule

Proof of Theorem 1. If the theorem is valid for a certain sequence $(h_n)_{n=3}^\infty$, then it is certainly valid for any sequence $(h'_n)_{n=3}^\infty$ satisfying $h'_n \geq h_n$ for each n . Hence, we may assume that $(h_n)_{n=3}^\infty$ is strictly decreasing to 0 and that $h_3 \leq 1$. Also, it will be convenient to let the sequence (h_n) start at $n = 1$, setting h_1 to be any value larger than 2 and $h_2 = 2$.

We construct a positive function $l(x)$ on $[1, \infty)$, as follows. For any positive integer $x = n$, let $l(n) = h_n$ and extend l linearly to each interval $[n, n + 1]$, $n = 1, 2, \dots$. Define a function $G_{\text{logit}(p)}$ by

$$G_{\text{logit}(p)}(x) = \begin{cases} 0, & x \leq 0, \\ \frac{l(1 + 1/x^2)x^2}{2(x^2 + 1)}, & 0 < x \leq 1, \\ 1 - \frac{l(x^2 + 1)}{2(x^2 + 1)}, & 1 < x < \infty. \end{cases}$$

Since $l(x)$ is continuous, positive, and strictly decreases to 0 as $x \rightarrow \infty$, with $l(2) = 2$, it is easy to see that $G_{\text{logit}(p)}(x)$ is a continuous distribution function supported on $[0, \infty)$. Suppose that $\text{logit}(p_i)$, $i = 1, 2, \dots, n$, are random variables distributed according to $G_{\text{logit}(p)}$, and recall that Y_i are the order statistics of $\text{logit}(p_i)$. Consider the events

$$A = \left\{ Y_n \geq \sum_{i=1}^{n-1} Y_i \right\}, \quad B = \{ Y_n \geq (n - 1)Y_1 \},$$

$$C = \left\{ Y_1 \leq \frac{1}{\sqrt{n-1}} \right\} \cup \{ Y_n \geq \sqrt{n-1} \}.$$

Clearly $A \subseteq B \subseteq C$. Thus,

$$\begin{aligned} P_e(n) &= P(A) \leq P(C) \leq P(Y_n \geq \sqrt{n-1}) + P\left(Y_1 \leq \frac{1}{\sqrt{n-1}}\right) \\ &\leq n P(\text{logit}(p) \geq \sqrt{n-1}) + n P\left(\text{logit}(p) \leq \frac{1}{\sqrt{n-1}}\right) \\ &= n \left(1 - G_{\text{logit}(p)}(\sqrt{n-1}) + G_{\text{logit}(p)}\left(\frac{1}{\sqrt{n-1}}\right) \right) \\ &= n \left(\frac{l(n)}{2n} + \frac{l(n)}{2n} \right) = l(n) = h_n, \quad n \geq 3. \end{aligned}$$

Proof of Theorem 2. The function $G_{\text{logit}(p)}(x)$ defined by

$$G_{\text{logit}(p)}(x) = \begin{cases} 0, & x < 0, \\ 1 - \frac{1}{\log(x + e)}, & 0 \leq x < \infty, \end{cases}$$

is clearly continuous and supported on $[0, \infty)$. Let $\text{logit}(p_i)$ be independent random variables distributed according to $G_{\text{logit}(p)}(x)$, and let

$$Y_n = \max(\text{logit}(p_1), \dots, \text{logit}(p_n)).$$

Let X_1, X_2, \dots, X_{n-1} respectively be the variables $\text{logit}(p_1), \text{logit}(p_2), \dots, \text{logit}(p_n)$, with the maximal $\text{logit}(p_i)$ excluded. Clearly

$$P_e(n) = \int_0^\infty \rho_{Y_n}(t) P\left(\sum_{i=1}^{n-1} X_i \leq t \mid Y_n = t\right) dt, \tag{5}$$

where ρ_{Y_n} is the density function of Y_n . We need to estimate $P(\sum_{i=1}^{n-1} X_i \leq t \mid Y_n = t)$. The conditional distribution function of each X_i under the condition $Y_n = t$ is

$$G_{X_i|Y_n}(x \mid t) = P(\text{logit}(p) \leq x \mid \text{logit}(p) \leq t) = \begin{cases} 0, & x < 0, \\ \frac{1 - 1/\log(x + e)}{1 - 1/\log(t + e)}, & 0 \leq x < t, \\ 1, & x \geq t. \end{cases}$$

Therefore,

$$\begin{aligned} E(X_i \mid Y_n = t) &= \frac{1}{1 - 1/\log(t + e)} \int_0^t \frac{x \, dx}{(x + e) \log^2(x + e)} \\ &= \frac{1}{1 - 1/\log(t + e)} \int_e^{t+e} \frac{1}{\log^2 x} \, dx - e. \end{aligned}$$

For $t \geq 31$, we have $\log^2(t + e) - t/e \leq 1$ and

$$\begin{aligned} E(X_i \mid Y_n = t) &\leq \frac{1}{1 - 1/\log(t + e)} \left(\int_e^{(t+e)/\log^2(t+e)} \frac{dx}{\log^2 x} + \int_{(t+e)/\log^2(t+e)}^{t+e} \frac{dx}{\log^2 x} \right) \\ &\leq \frac{1}{1 - 1/\log(t + e)} \left(\frac{t + e}{\log^2(t + e)} + \frac{t + e}{\log^2[(t + e)/\log(t + e)]} \right) \\ &\leq \frac{1}{1 - 1/\log(t + e)} \frac{(2 + c_1)(t + e)}{\log^2(t + e)}, \end{aligned}$$

for some positive constant c_1 . Thus,

$$E\left(\sum_{i=1}^{n-1} X_i \mid Y_n = t\right) \leq \frac{n - 1}{1 - 1/\log(t + e)} \frac{(2 + c_1)(t + e)}{\log^2(t + e)}.$$

By Markov's inequality, we obtain

$$P\left(\sum_{i=1}^{n-1} X_i > t \mid Y_n = t\right) \leq n \left(1 + \frac{e}{t}\right) \frac{c_1 + 2}{\log(t + e)(\log(t + e) - 1)}. \tag{6}$$

Now we claim that

$$\lim_{n \rightarrow \infty} P(e^{n/\log n} - e \leq Y_n \leq e^{n \log n} - e) = 1. \tag{7}$$

Indeed, since $G_{Y_n}(x) = (1 - 1/\log(x + e))^n$ for $x > 0$, we have

$$\begin{aligned} &\lim_{n \rightarrow \infty} P(e^{n/\log n} - e \leq Y_n \leq e^{n \log n} - e) \\ &= \lim_{n \rightarrow \infty} \left(\left(1 - \frac{1}{n \log n}\right)^{n \log n} \right)^{1/\log n} - \lim_{n \rightarrow \infty} \left(\left(1 - \frac{\log n}{n}\right)^{n/\log n} \right)^{\log n} \\ &= 1. \end{aligned}$$

From (6) and (5), it follows that

$$\begin{aligned}
 P_e(n) &\geq \int_{e^{n/\log n - e}}^{e^{n \log n - e}} \rho_{Y_n}(t) \left(1 - n \left(1 + \frac{e}{t} \right) \frac{c_1 + 2}{\log(t + e)(\log(t + e) - 1)} \right) dt \\
 &\geq \left(1 - \frac{(c_1 + 2)n(1 + e/[e^{n/\log n} - e])}{(n/\log n)(n/\log n - 1)} \right) \int_{e^{n/\log n - e}}^{e^{n \log n - e}} \rho_{Y_n}(t) dt \\
 &\geq \left(1 - \frac{c_2}{(1/\log n)(n/\log n - 1)} \right) \int_{e^{n/\log n - e}}^{e^{n \log n - e}} \rho_{Y_n}(t) dt
 \end{aligned}$$

for some constant c_2 . Hence, (7) implies that $P_e(n) \rightarrow 1$ as $n \rightarrow \infty$.

3.2. Simple majority rule

Proof of Theorem 3. Consider an arbitrary interval $I(x, \delta) = [x, x + \delta)$. For any s , let $l \equiv l(s) \geq 1$ be the integer determined by

$$x \left(1 + \frac{1}{s} \right)^{l-1} \leq x + \delta < x \left(1 + \frac{1}{s} \right)^l. \tag{8}$$

Thus, $I(x, \delta) \subseteq \bigcup_{j=1}^l I_j(x, \delta)$, where $I_j(x, \delta) = [x(1 + 1/s)^{j-1}, x(1 + 1/s)^j)$. Clearly, for $1 \leq j \leq l$ we have

$$P_m(n) \geq P(\text{logit}(p_i) \in I_j(x, \delta), i = 1, \dots, n) = P^n(\text{logit}(p) \in I_j(x, \delta)).$$

Thus,

$$P_m(n) \geq \left(\max_{1 \leq j \leq l} P(\text{logit}(p) \in I_j(x, \delta)) \right)^n \geq \left(\frac{G_{\text{logit}(p)}(x + \delta) - G_{\text{logit}(p)}(x)}{l} \right)^n. \tag{9}$$

Now let

$$c < \sup_x \left(x \overline{\lim}_{\delta \rightarrow 0^+} \frac{G_{\text{logit}(p)}(x + \delta) - G_{\text{logit}(p)}(x)}{\delta} \right). \tag{10}$$

Take an interval $I(x_0, \delta_0)$ such that $x_0(G_{\text{logit}(p)}(x_0 + \delta_0) - G_{\text{logit}(p)}(x_0))/\delta_0 > c$. Using (9) with the chosen interval, we obtain

$$P_m(n) \geq \left(\frac{G_{\text{logit}(p)}(x_0 + \delta) - G_{\text{logit}(p)}(x_0)}{l} \right)^n > \left(\frac{c\delta_0}{lx_0} \right)^n. \tag{11}$$

By (8), we obtain $(1 + 1/s)^{l-1} \leq 1 + \delta_0/x_0 \leq (1 + 1/s)^l$. This implies that, as $s \rightarrow \infty$,

$$(1 + 1/s)^l \rightarrow 1 + \delta_0/x_0$$

and, therefore, $l/s \rightarrow \log(1 + \delta_0/x_0)$. It follows that

$$e^{l/s} \rightarrow 1 + \delta_0/x_0 < e^{\delta_0/x_0}$$

as $s \rightarrow \infty$, so that for sufficiently large s we have $l/s < \delta_0/x_0$. Consequently, (11) in Theorem 3 yields the inequality required for values of c satisfying (10). The inequality can be proved similarly for

$$c < \sup_x \left(x \overline{\lim}_{\delta \rightarrow 0^-} \frac{G_{\text{logit}(p)}(x + \delta) - G_{\text{logit}(p)}(x)}{\delta} \right).$$

Proof of Theorem 4. Consider the generalized uniform distribution

$$G_p(x) = (2x - 1)^\alpha, \quad 0.5 \leq x \leq 1, 0 < \alpha \leq 1,$$

discussed in [7]. From part 2 of our first list, we have $P_m(n) \leq \alpha^s / (s!)^2$. By a version of Stirling’s formula, $s! \geq \sqrt{2\pi}(s + 1)^{s+1/2}e^{-s-1}$ and, consequently, $P_m(n) \leq (e\sqrt{\alpha}/s)^n / 2\sqrt{\alpha}$. For each $n \geq 2$, we have

$$\frac{(\sqrt{\alpha})^n}{2\sqrt{\alpha}} \leq (\sqrt{\alpha})^{n-1} = (\alpha^{1/4})^n \alpha^{(n-2)/4} \leq (\alpha^{1/4})^n.$$

Given an $\varepsilon > 0$, taking $\alpha < (\varepsilon/e)^4$ yields the required result.

Proof of Theorem 5. As hinted at in Remark 4, Theorem 5 corresponds to a special case of Theorem 9 with $c = 0$.

Proof of Theorem 6. Without loss of generality, we may assume the sequence (h_n) to be nonincreasing. Construct $G_{\text{logit}(p)}$ as follows. First, we define it at some ‘special’ points:

$$\begin{aligned} G_{\text{logit}(p)}(0) &= 0, & G_{\text{logit}(p)}(1) &= \frac{1 - h_3}{2}, \\ G_{\text{logit}(p)}\left(1 + \frac{1}{s}\right) &= \frac{1 - h_3}{2} + h_{2s+1}, & s &= 1, 2, 3, \dots \end{aligned}$$

(Note that $1 + 1/s$ increases as s decreases.) Complete $G_{\text{logit}(p)}$ to an increasing function in some way. Clearly, if all the $\text{logit}(p_i)$ are in the range $[1, 1 + 1/s]$, then the sum of any $s + 1$ of them does not exceed the sum of the others. Hence,

$$\begin{aligned} P_m(n) &\geq P(1 \leq \text{logit}(p_i) \leq 1 + 1/s, i = 1, 2, \dots, n) \\ &= (G_{\text{logit}(p)}(1 + 1/s) - G_{\text{logit}(p)}(1))^n \\ &= h_{2s+1}^n = h_n^n, \end{aligned}$$

which gives the desired estimate.

Proof of Theorem 7. Let $\text{logit}(p_i)$, $i = 1, 2, \dots, n$, be independent random variables distributed according to $G_{\text{logit}(p)}(x)$. Let x_0 be a point such that $P(f(x) = x_0) = c$, and again let Y_i , $i = 1, 2, \dots, n$, be the order statistics of $\text{logit}(p_i)$. Obviously,

$$P_m(n) = P\left(\sum_{i=1}^{s+1} Y_i \geq \sum_{i=s+2}^n Y_i\right) \geq P(\text{logit}(p_i) = x_0, i = 1, 2, \dots, n) = c^n.$$

Proof of Theorem 8. Let $\varepsilon > 0$ and choose an integer $m > (1/\varepsilon)^3$. Define the distribution $G_{\text{logit}(p)}$ so as to assign equal probabilities $1/m$ to the points $1, 3, 9, \dots, 3^{m-1}$. We claim that this distribution satisfies the inequality required in the theorem. Let $\text{logit}(p_i)$ be selected randomly and independently of $G_{\text{logit}(p)}$. Consider the events

$$\begin{aligned} A &= \left\{ \sum_{i=1}^{s+1} Y_i \geq \sum_{i=s+2}^n Y_i \right\}, & D &= \bigcup_{j=0}^{m-1} \{ \text{logit}(p_i) = 3^j, i = 1, \dots, n \}, \\ E &= \bigcup_{j=0}^{m-2} \{ Y_1 = 3^j, Y_2 = \dots = Y_n > 3^j \}. \end{aligned}$$

We can easily check that, for this distribution, $A = D \cup E$. Thus,

$$P_m(n) = m \frac{1}{m^n} + \sum_{j=0}^{m-1} \frac{1}{m} \frac{m-1-j}{m^{n-1}} = \left(\frac{1}{m} \left(\frac{m(m+1)}{2} \right)^{1/n} \right)^n$$

and, for each $n \geq 3$,

$$P_m(n) \leq \left(\frac{1}{m} \left(\frac{m(m+1)}{2} \right)^{1/3} \right)^n = \left(\left(\frac{(m+1)}{2m^2} \right)^{1/3} \right)^n \leq \left(\left(\frac{1}{m} \right)^{1/3} \right)^n < \varepsilon^n.$$

Proof of Theorem 9. Let $\delta > 0$. Choose points $a_0, a_1, \dots, a_m, 0 = a_0 < a_1 < \dots < a_m = \infty$, such that $G_{\text{logit}(p)}(a_{j+1}) - G_{\text{logit}(p)}(a_j) < c + \delta, j = 0, 1, \dots, m - 1$. By adding more points between consecutive a_j s, we may further assume that

$$G_{\text{logit}(p)}(a_{j+3}) - G_{\text{logit}(p)}(a_j) < c + \delta, \quad j = 0, 1, \dots, m - 3.$$

Let $\text{logit}(p_i), i = 1, 2, \dots, n$, be independent random variables distributed according to some distribution function and let Y_i be the order statistics of $\text{logit}(p_i)$. Determine the integer k satisfying

$$k \geq \max_{0 \leq j \leq m-1} \frac{a_j}{a_{j+1} - a_j} > k - 1.$$

Consider the events

$$A = \left\{ \sum_{i=1}^{s+1} Y_i \geq \sum_{i=s+2}^n Y_i \right\}, \quad D_j = \{\text{logit}(p_i) \leq a_j \text{ for at most } k \text{ indices } i\},$$

$$E_j = \{\text{logit}(p_i) \geq a_{j+1} \text{ for at most } k - 1 \text{ indices } i\}, \quad B_j = D_j \cup E_j.$$

Clearly, $A \subseteq B_j$ for each j and all $n > 2k + 1$. Let $I = [a_{j_0}, a_{j_0+1})$ be the interval containing the maximal number of $\text{logit}(p_i)$. For a j_0 with $1 \leq j_0 \leq m - 2$, we have

$$A \subseteq B_{j_0-1} \cap B_{j_0+1}. \tag{12}$$

Let $H_{j_0} = D_{j_0-1} \cap E_{j_0+1}$. We claim that

$$B_{j_0-1} \cap B_{j_0+1} \subseteq H_{j_0} \tag{13}$$

for $n > mk$. Indeed, since the interval I contains at least n/m of the $\text{logit}(p_i)$, the event E_{j_0-1} is impossible for $n > m(k - 1)$, and, hence, $B_{j_0-1} = D_{j_0-1}$. Similarly, D_{j_0+1} is impossible for $n > mk$, and, hence, $B_{j_0+1} = E_{j_0+1}$. Thus, we obtain (13). If $j_0 = 0$ we have

$$A \subseteq H_0, \tag{14}$$

where $H_0 = E_1$. If $j_0 = m - 1$ we have

$$A \subseteq H_{m-1}, \tag{15}$$

where $H_{m-1} = D_{m-2}$. By (12), (13), (14), and (15), we obtain

$$\begin{aligned}
 P_m(n) &\leq \sum_{j=0}^{m-1} P(H_j) \\
 &= \sum_{i=0}^{k-1} \binom{n}{i} (1 - G(a_2))^i G^{n-i}(a_2) \\
 &\quad + \sum_{i=0}^k \binom{n}{i} G^i(a_{m-2}) (1 - G(a_{m-2}))^{n-i} \\
 &\quad + \sum_{j=1}^{m-2} \sum_{i=0}^k \sum_{t=0}^{k-1} \binom{n}{i} \binom{n-i}{t} G^i(a_j) (1 - G(a_{j+3}))^t (G(a_{j+3}) - G(a_j))^{n-i-t} \\
 &\leq 2 \sum_{i=0}^k \binom{n}{i} (m-2)^i (c + \delta)^i (c + \delta)^{n-i} \\
 &\quad + \sum_{j=1}^{m-2} \sum_{i=0}^k \sum_{t=0}^{k-1} \binom{n}{i} \binom{n-i}{t} j^i (m-j-3)^t (c + \delta)^n \\
 &\leq Q(m, k) n^{2k-1} (c + \delta)^n
 \end{aligned}$$

for all sufficiently large n and some constant $Q(m, k)$. Given an $\varepsilon > 0$, by taking $\delta < \varepsilon$ we obtain the required result.

Proof of Theorem 10. For any ε , $0 < \varepsilon < 1$, construct a distribution function $G_{\text{logit}(p)}(x)$ that assigns mass ε to some point $x_0 > 0$. Obviously,

$$P_m(n) \geq P(\text{logit}(p_i) = x_0, i = 1, \dots, n) = \varepsilon^n,$$

which implies the result.

Acknowledgements

The authors express their gratitude to J. Harmse for his helpful comments and suggestions, and would like to thank the referee for his comments.

References

- [1] AUSTEN-SMITH, D. AND BANKS, J. S. (1996). Information aggregation, rationality, and the Condorcet jury theorem. *Amer. Political Sci. Rev.* **90**, 35–45.
- [2] BAKER, K. M. (ed.) (1976). *Condorcet: Selected Writings*. Bobbs-Merrill, Indianapolis, IN.
- [3] BEN-YASHAR, R. AND PAROUSH, J. (2000). A nonasymptotic Condorcet jury theorem. *Soc. Choice Welf.* **17**, 189–199.
- [4] BEREND, D. AND HARMSE, J. (1993). Expert rule versus majority rule under partial information. *Theory Decision* **35**, 179–197.
- [5] BEREND, D. AND PAROUSH, J. (1998). When is Condorcet’s jury theorem valid? *Soc. Choice Welf.* **15**, 481–488.
- [6] BEREND, D. AND SAPIR, L. (2001). Optimality of the expert rule under partial information. *Acta Appl. Math.* **69**, 141–162.
- [7] BEREND, D. AND SAPIR, L. (2002). Expert rule versus majority rule under partial information. II. *J. Appl. Math. Decision Sci.* **6**, 79–99.
- [8] BEREND, D. AND SAPIR, L. (2003). Between the expert and majority rules. *Adv. Appl. Prob.* **35**, 941–960.

- [9] BEREND, D. AND SAPIR, L. (2005). Monotonicity in Condorcet jury theorem. *Soc. Choice Welf.* **24**, 83–92.
- [10] BERG, S. (1993). Condorcet's jury theorem revisited. *Europ. J. Political Econ.* **9**, 437–446.
- [11] BERG, S. AND PAROUSH, J. (1998). Collective decision making in hierarchies. *Math. Soc. Sci.* **35**, 233–244.
- [12] BOLAND, P. J. (1989). Majority systems and the Condorcet jury theorem. *The Statistician* **38**, 181–189.
- [13] BOLAND, P. J., PROSCHAN, F. AND TONG, Y. L. (1989). Modelling dependence in simple and indirect majority systems. *J. Appl. Prob.* **26**, 81–88.
- [14] CARITAT, M. J. A. N., Marquis de Condorcet. (1785). *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*. Paris.
- [15] GRADSTEIN, M. AND NITZAN, S. (1986). Performance evaluation of some special classes of weighted majority rules. *Math. Soc. Sci.* **12**, 31–46.
- [16] GROFMAN, B., OWEN, G. AND FELD, S. (1983). Thirteen theorems in search of the truth. *Theory Decision* **15**, 261–278.
- [17] KAROTKIN, D. (1998). The network of weighted majority rules and weighted majority games. *Games Econom. Behav.* **22**, 299–315.
- [18] LADHA, K. K. (1995). Information polling through majority-rule voting: Condorcet's jury theorem with correlated votes. *J. Econ. Behav. Organizat.* **26**, 353–372.
- [19] MILLER, N. R. (1986). Information, electorates, and democracy: some extensions and interpretations of Condorcet jury theorem. In *Information Pooling and Group Decision Making*, eds B. N. Grofman and G. Owen, JAI Press, Greenwich, CT, pp. 173–192.
- [20] NITZAN, S. AND PAROUSH, J. (1985). *Collective Decision Making*. Cambridge University Press.
- [21] PAROUSH, J. (1998). Stay away from fair coins: a correct Condorcet jury theorem. *Soc. Choice Welf.* **15**, 15–20.
- [22] SAPIR, L. (1998). The optimality of the expert and majority rules under exponentially distributed competence. *Theory Decision* **45**, 19–35.
- [23] SAPIR, L. (2004). Comparison of the polar decision rules for various types of distributions. *Theory Decision* **56**, 325–343.
- [24] SAPIR, L. (2005). Generalized means of jurors' competencies and marginal changes of jury's size. *Math. Soc. Sci.* **50**, 83–101.
- [25] YOUNG, H. P. (1989). Condorcet's theory of voting. *Amer. Political Sci. Rev.* **82**, 1231–1244.