# Aggregation of Correlated Votes and Condorcet's Jury Theorem

Serguei Kaniovski*

September 11, 2008

**Abstract**

This paper proves two theorems for homogeneous juries that arise from different solutions to the problem of aggregation of dichotomous choice. In the first theorem, negative correlation increases the competence of the jury, while positive correlation has the opposite effect. An enlargement of the jury with positive correlation can be detrimental up to a certain size, beyond which it becomes beneficial. The second theorem finds a family of distributions for which correlation has no effect on a jury's competence. The approach allows us to compute the bounds on a jury's competence as the maximum and minimum probability of it being correct for a given competence and dependence structure.

*JEL-Codes:* C63, D72
*Key Words:* dichotomous choice, Condorcet's Jury Theorem, correlated votes

*Acknowledgments: I would like to thank Krishna Ladha, Shmuel Nitzan, Donald Nolte, Peyton Young and Alexander Zaigraev for comments on an earlier draft, and Paul van der Geest for providing computer programs for the entropy solution.
Austrian Institute of Economic Research (WIFO), P.O. Box 91; A-1103 Vienna, Austria.
Email address: serguei.kaniovski@wifo.ac.at.

# 1 Introduction

Condorcet argued for the superiority of collective wisdom over individual wisdom. His famous 1785 essay rationalized entrusting important decisions to a group rather than to an individual, and advocated simple majority as the best collective decision rule. The larger the group, the more likely it is to discover the truth. The larger the jury, the more likely it is to convict the guilty or acquit the innocent. The significance of Condorcet's analysis lies in offering a positive argument in favor of the democratic ideas widely appreciated on normative grounds.

Condorcet's Jury Theorem (CJT) is a mathematical formalization of his ideas. In its classic version the theorem rests on the following four assumptions: a) the jury decides between two alternatives by voting under simple majority rule, b) each juror has a more than an even probability of being correct (competence), c) these probabilities are equal for all jurors (homogeneity), and d) each juror votes independently. Then,

1. any jury comprising an odd number of jurors is more likely than any single juror to select the correct alternative;

2. this likelihood becomes a certainty as the size of the jury tends to infinity.

It is customary to refer to 1 as the non-asymptotic and 2 as the asymptotic part of the theorem.

Much research is concerned with relaxing the homogeneity and independence assumptions.[1] Both assumptions are unrealistic, as jurors are subject to many individual and contextual factors, which are likely to influence their decisions in different ways. A realistic model should allow varying and correlated competencies that translate into varying probabilities and correlations between the votes. The aim of such a model is to forecast the probabilities of different voting

[1] This literature assumes sincere voting. Sincere voting may not be an equilibrium behavior in the presence of informational asymmetries among jurors (Austen-Smith and Banks 1996, Feddersen and Pesendorfer 1997).

outcomes.

I assume that the *competence* and *dependence* of a jury is specified by the probability of a juror's vote being correct and the Pearson product-moment correlation coefficient between two correct votes as the simplest and most widely used measure of probabilistic dependence. The probabilities and correlation coefficients can be assumed (voting scenario) or estimated from ballot data.

Computing the probability of the jury being correct requires a *joint probability distribution* on the set of voting profiles, but individual probabilities and correlation coefficients do not uniquely define a distribution. For a jury of size $n$, distributions satisfying given probabilities and correlation coefficients form a convex set in a $2^n$ dimensional Euclidean space, where $2^n$ is the number of all conceivable voting profiles in a jury of size $n$.

Solving the aggregation problem amounts to finding a joint probability distribution that has given individual probabilities and correlation coefficients. Bahadur (1961) characterizes the solution using higher-order correlation coefficients. The difficulty in applying his result lies in the large number of parameters it therefore requires. Estimating higher-order correlations to a reasonable degree of accuracy requires extensive voting data.

Alternatively, one can use *regularization* to find a particular solution (Van Der Geest 2005, Kaniovski and Pflug 2007). Regularization is a widely used technique that imposes additional criteria on the solution in order for it to have certain properties, in our case uniqueness. It offers a practical solution to the aggregation problem.

Since in its full generality the aggregation problem admits varying probabilities and correlation coefficients, it can be used to model both correlated as well as heterogeneous votes numerically. Workable explicit solutions are available for homogeneous juries only. In a *homogeneous* jury each vote has an equal probability of being correct, and each pair of votes correlates with the same coefficient of correlation.

I prove two non-asymptotic theorems for homogeneous juries that arise from different solutions to the aggregation problem. Both theorems corroborate existing results, but one theorem differs substantially with respect to tolerance of correlation. It shows that the probability of a homogeneous jury collectively making the correct decision under simple majority rule can be *independent* of the coefficient of correlation.

Next I review the extensions of CJT to heterogeneous and correlated votes. Section 3 states the aggregation problem and its solutions. Section 4 presents two jury theorems based on explicit solutions. Section 5 provides examples illustrating the effect of the probabilities and correlations on a jury's competence. Section 6 discusses the bounds to a jury's competence as the minimum and maximum probability of the jury being correct. The final section provides a summary.

## 2 Extensions of Condorcet's Jury Theorem

Proofs of the classic version of CJT can be found in Young (1988) and Boland (1989). A proof is obtained using a recurrence relation for the collective probability of deciding correctly when the size of the jury increases by two jurors. Monotonicity of this probability with respect to size establishes both parts of the theorem.

### 2.1 Heterogeneous but independent jurors

The non-asymptotic part of the theorem does not generally hold for heterogeneous juries. In a heterogeneous jury all jurors are competent, but differently so. Ben-Yashar and Paroush (2000) propose a refinement of the CJT that remains valid regardless of the specific distribution of competence within a jury. They show the non-asymptotic part of CJT to hold if the average competence of a randomly chosen group of more than three jurors is higher than that of any single juror. Berend and Sapir (2005) extended this result to show that a random augmentation of a randomly chosen group always improves the competence of the group.

As for the asymptotic part of the theorem, Boland (1989) already proposed a refinement in terms of the average competence for which it holds for a heterogeneous jury. Berend and Paroush (1998) provide the most general result.

### 2.2 Correlated votes

Existing models of correlated votes are of two kinds: one that endorses an explicitly sequential view of voting which induces a joint probability distribution on the set of voting profiles, and one that studies the conditions under which the collective competence increases with the size, given a joint probability distribution.

In Boland (1989) and Boland, Proschan and Tong (1989) the course of voting is shaped by an opinion leader's vote or an external factor, such as the quality of court evidence. Berg (1993) models voting as an urn process, in which a juror's competence depends on all preceding votes. There is a state dependence in the process of reaching a decision, with the possibility of a lock-in on the incorrect alternative (Page 2006). There is both heterogeneity, as a juror's probability of being correct may depend on her position in the voting sequence, and dependence, although pairwise correlations may decay with the time lag between the votes.

Ladha (1993) formulates a juror's competence as a probability conditional on the juror's information. While the votes are conditionally independent, unconditionally they are positively correlated. This way of looking at competence is useful when there is an explicit need to model the role of common information. Ladha (1992) obtains sufficient conditions for both parts of the theorem as an upper bound on the average of positive correlation coefficients. The probability of the jury's decision being correct increases with the competency of the jurors and decreases with positive correlation between their competencies, subject to an inverse relationship between the size of the jury and the admissible dependence. Ladha (1993) obtains stronger results for exchangeable binary random variables; Ladha (1995) extends them to several particular distributional assumptions.

Berend and Sapir (2007) ask whether adding a randomly chosen juror to a randomly chosen group of jurors improves the group's competence, and derive a necessary and sufficient condition in terms of the joint probability distribution on the set of voting profiles. We will presently see that, regardless of how correlations arise, any such distribution has Bahadur's representation. His analysis offers a unifying theoretical framework for dichotomous choice models with correlation.

## 3 The aggregation problem

In a jury comprising an odd number of jurors $n$, let $p_i$, $i = 1, 2, \ldots, n$, denote the probability of the $i$-th juror voting for the correct alternative and $c_{i,j}$, $1 \leq i < j \leq n$, the correlation coefficient between any two such votes. There will be $C_n^2$ such correlation coefficients, one for each order-independent pair of votes.[2] The correlation coefficient $c_{i,j}$ is a pairwise measure of dependence.

A voting profile is a binary vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$, whose $i$-th coordinate $v_i = 1$ if juror $i$ votes for the correct alternative, and $v_i = 0$ otherwise. Let $\mathbf{V}$ be the set of all voting profiles, $\mathbf{V}(i)$ the set of voting profiles in which juror $i$ votes for the correct alternative, i.e. the set of all binary vectors $\mathbf{v}$ such that $v_i = 1$, and $\mathbf{V}(i, j) = \mathbf{V}(i) \cap \mathbf{V}(j)$ the set of voting profiles in which jurors $i$ and $j$ both vote for the correct alternative, i.e. the set of all binary vectors $\mathbf{v}$ such that

[2] Here and below $C_n^x$ denotes the binomial coefficient $C_n^x = n!/[x!(n-x)!]$ for $n, x \in \mathbb{N}$, where $C_n^x = 0$ for $n < x$.

$v_i = v_j = 1$.

The sets $\mathbf{V}$, $\mathbf{V}(i)$ and $\mathbf{V}(i,j)$ respectively contain $2^n$, $2^{n-1}$ and $2^{n-2}$ elements. For example, for $n = 3$ there will be eight voting profiles A:(1,1,1), B:(1,1,0), C:(1,0,1), D:(1,0,0), E:(0,1,1), F:(0,1,0), G:(0,0,1), and H:(0,0,0). The set $\mathbf{V}$ contains all eight vectors. The set $\mathbf{V}(2)$ contains vectors A, B, E, and F, as only they have 1 in the second coordinate. The set $\mathbf{V}(2,3)$ contains vectors A and E, as only they have 1 in the second and third coordinates.

A joint probability distribution $\pi_{\mathbf{v}}$ on the set of voting profiles $\mathbf{V}$ that satisfies given marginal probabilities and correlation coefficients must satisfy the following constraints:

$$\begin{aligned}
\pi_{\mathbf{v}} &\geq 0 \quad \text{for all} \quad \mathbf{v} \in \mathbf{V}; && (1)\\
\sum_{\mathbf{v}\in\mathbf{V}} \pi_{\mathbf{v}} &= 1; && (2)\\
\sum_{\mathbf{v}\in\mathbf{V}(i)} \pi_{\mathbf{v}} &= p_i \quad \text{for all} \quad i = 1, 2, \ldots, n; && (3)\\
\sum_{\mathbf{v}\in\mathbf{V}(i,j)} \pi_{\mathbf{v}} &= p_i p_j + c_{i,j}\sqrt{p_i q_i p_j q_j} \quad \text{for all} \quad q_i = 1 - p_i, \quad 1 \leq i < j \leq n. && (4)
\end{aligned}$$

Let $V_i$ be the Bernoulli random variable signifying the value of the vote and $v_i$ be its realization. Equality (4) follows since for two Bernoulli random variables $V_i$, $V_j$ with $E(V_i) = p_i$, $E(V_j) = p_j$:

$$c_{i,j} = \frac{P\{V_i = 1, V_j = 1\} - p_i p_j}{\sqrt{p_i q_i p_j q_j}}.$$

Constraints (1)-(4) define a convex polytope $\Delta \subset \mathbb{R}^{2^n}$. Any point in $\Delta$ is a suitable distribution. Such a distribution does not exist if $\Delta = \emptyset$, i.e. when the constraints are inconsistent. The system (2)-(4) comprising $1 + n + C_n^2$ equations for $2^n$ unknowns typically has infinitely many solutions for $n \geq 3$. Consequently, if a distribution exists, it will typically not be unique. Two methods of finding a particular solution include adding sufficiently many equations to uniquely characterize the distribution, or regularization.

### 3.1 Bahadur's solution (B-solution)

Bahadur's (1961) method uses higher-order correlations to uniquely characterize a distribution. This introduces further linear constraints to (2)-(4) until the feasible set $\Delta$ becomes a singleton.

Higher order correlation coefficients measure dependence between $k$-tuples of votes. Let $Z_i = (V_i - p_i)/\sqrt{p_i q_i}$ for all $i = 1, 2, \ldots, n$. Then,

$$\begin{aligned} c_{i,j} &= E(Z_i Z_j) \quad \text{for all} \quad 1 \leq i < j \leq n; \\ c_{i,j,k} &= E(Z_i Z_j Z_k) \quad \text{for all} \quad 1 \leq i < j < k \leq n; \\ &\ldots \\ c_{1,2,\ldots,n} &= E(Z_1 Z_2 \ldots Z_n). \end{aligned}$$

In total, there are $\sum_{i=2}^{n} C_n^i = 2^n - n - 1$ correlation coefficients, which together with $n$ marginal probabilities uniquely define a distribution on $2^n$ voting profiles.

Bahadur's solution is given by

$$\pi_{\mathbf{v}} = \bar{\pi}_{\mathbf{v}} \left( 1 + \sum_{i<j} c_{i,j} z_i z_j + \sum_{i<j<k} c_{i,j,k} z_i z_j z_k + \cdots + c_{1,2,\ldots,n} z_1 z_2 \ldots z_n \right),$$

where $\bar{\pi}_{\mathbf{v}} = \prod_{i=1}^{n} p_i^{v_i} q_i^{1-v_i}$ denotes the distribution in the case of independent votes.[3]

Bahadur's method requires an excessive number of parameters to characterize the distribution. To reduce complexity, he proposes a truncated solution (B-solution), which accounts for

[3] Bahadur (1961) ascribes the discovery of this representation to Lazarsfeld (1956).

pairwise correlation coefficients only

$$\pi_{\mathbf{v}} = \bar{\pi}_{\mathbf{v}}\left(1 + \sum_{i<j} c_{i,j} z_i z_j\right).$$

It is accurate if all higher-order correlation coefficients are zero. If some higher-order coefficients are not zero, then additional constraints on $c_{i,j}$'s for given $n$ and $p_i$'s are required for the B-solution to have non-negative coordinates.

In the special case of a homogeneous voting body in which $p_i = p$ and $c_{i,j} = c$, $\pi_{\mathbf{v}}$ can be expressed in terms of the number of votes in favor of the correct alternative $t(\mathbf{v}) = \sum_{i=1}^{n} v_i$.[4] In this case the B-solution simplifies to

$$\pi_{\mathbf{v}} = \bar{\pi}_{\mathbf{v}}\left\{1 + \frac{c}{2pq}\left[t^2 - t + p(n-1)(np - 2t)\right]\right\}, \quad \text{where} \quad \bar{\pi}_{\mathbf{v}} = p^t q^{n-t}. \tag{5}$$

Bahadur provides exact bounds on the common correlation coefficient $c$ required for (5) to define a distribution. Denote $\gamma = \min_t\{[t - (n-1)p - 0.5]^2\} \leq 0.25$. Then,

$$-\frac{2q}{pn(n-1)} \leq c \leq \frac{2pq}{(n-1)pq + 0.25 - \gamma}. \tag{6}$$

The bound on $n$ is tighter for negative $c$ than for positive $c$. For $c > 0$, $c \sim O(n^{-1})$, whereas $|c| \sim O(n^{-2})$ for $c < 0$. The nature of the constraints is such that for the given $p$'s, $|c| \to 0$ as $n \to \infty$. That is, in a large homogeneous jury each pair of jurors may be only weakly dependent. It can be easily verified that $c \leq 2/(n-1)$ for $c > 0$.

The following numerical methods impose the non-negativity constraint on the solution.

[4]Instead of $t(\mathbf{v})$, I will use the simpler notation $t$, keeping in mind that the sum of the coordinates depends on the vector, i.e. the voting profile.

### 3.2 Regularization using an entropy function (E-solution)

Regularization ensures uniqueness by choosing a particular solution as a minimum point in $\Delta$ of a strictly convex function $f : \mathbb{R}^{2^n} \mapsto \mathbb{R}$.

Van Der Geest (2005) proposes choosing $f = \sum_{\mathbf{v}} \pi_{\mathbf{v}} \log(\pi_{\mathbf{v}})$ so as to maximize the entropy function $H = -f$.[5] This distribution can be obtained by solving the following non-linear optimization problem:

$$\min_{\pi_{\mathbf{v}}} \quad \sum_{\mathbf{v}} \pi_{\mathbf{v}} \log(\pi_{\mathbf{v}}),$$

subject to constraints (2)-(4). Constraint (1) is superfluous, as the objective function is not defined for $\pi_{\mathbf{v}} = 0$. Solution vectors generated using this method have strictly positive coordinates, so that all voting profiles occur with a positive probability. The complexity of the objective function makes an explicit solution unfeasible.

### 3.3 Regularization using a quadratic metric (Q-solution)

The gist of this method is as follows: Given a benchmark distribution $d_{\mathbf{v}}$ and a distance measure $f$, choose a point in $\Delta$ that is closest to $d_{\mathbf{v}}$ in the sense of $f$. Kaniovski and Pflug (2007) propose $d_{\mathbf{v}} = \bar{\pi}_{\mathbf{v}}$ and $f$ as the sum of squares of deviations. This choice leads to the following quadratic optimization problem:

$$\min_{\pi_{\mathbf{v}}} \quad 0.5 \sum_{\mathbf{v}} (\pi_{\mathbf{v}} - \bar{\pi}_{\mathbf{v}})^2,$$

subject to constraints (1)-(4). Solution vectors generated by this method may have zero coordinates, so that some voting profiles may occur with probability zero.

With the full set of constraints imposed, the above quadratic optimization problem can only

[5] Maximum entropy distributions have a long tradition in statistical physics and information theory (Levine and Tribus 1978).

be solved numerically. For $p_i = p$, Kaniovski (2008a) obtains an explicit solution by the method of Lagrangian multipliers. In the case of a homogeneous jury, when $p_i = p$ and $c_{i,j} = c$, the Q-solution reads

$$\pi_{\mathbf{v}} = \bar{\pi}_{\mathbf{v}} + 2^{2-n} pqc \Big[ 0.5n(n-1) + 2t(t-n) \Big], \quad \text{where} \quad \bar{\pi}_{\mathbf{v}} = p^t q^{n-t}. \tag{7}$$

Similarly to the B-solution, it ignores the complementary slackness condition imposed by (1) and thus is valid only if $\pi_{\mathbf{v}} \geq 0$ for all $\mathbf{v}$.

### 3.4 Remarks on applications

Unlike sequential voting models that capture group dynamics, the above framework applies in the baseline case of simultaneous and anonymous voting, or when the expertise of several experts whose opinions have been expressed individually is pooled into a collective judgment. This is the classic setting of CJT.

Bahadur's is the only explicit and general solution to the aggregation problem. However, there are two practical reasons to prefer the numerical methods based on regularization in applied work. First, they require significantly fewer parameters. Second, they can be applied even if not all pairwise correlation coefficients are known. In this case one should use those constraints which contain the known correlation coefficients, leaving the others unspecified. Once a probability distribution satisfying the marginal probabilities and known correlation coefficients is found, the unknown correlation coefficients can be inferred from it. This allows us to first specify and then refine an incomplete model, offering considerable flexibility in applied work.

Models of actual juries which impose competence and correlation structures should use the numerical solutions in which the full set of constraints is imposed. If distributions satisfying the probabilities and correlation coefficients exist, a numerical solution will find one such distribution. Examples in Section 5 show that numerical solutions differ from the B-solution, implying that some higher-order correlations differ from zero.

Once a distribution is found, the probabilities any event of interest can be computed. Kaniovski (2008a) computes the probability of casting a decisive vote in simple-majority games with equal voting weights. He then modifies Penrose's (1946) Square Root Rule for the apportionment of voting power in two-tier voting systems, in which votes in the bottom tiers correlate. Kaniovski (2008b) provides several examples of Boland et al.'s (1989) opinion-leader theme.

## 4 Two jury theorems

The above solutions can be used to compute the probabilities of any voting outcome and therefore also the probability of a jury collectively making the correct decision under any given voting rule. Under simple majority rule this probability equals

$$M_n(\mathbf{p}, \mathbf{c}) = \sum_{t=\frac{n+1}{2}}^{n} \pi_{\mathbf{v}}. \tag{8}$$

The notation highlights the dependence of this probability on the vector of $n$ marginal probabilities $\mathbf{p}$ and the vector of $C_n^2$ pairwise correlation coefficients $\mathbf{c}$. I will refer to $M_n$ under B, E, and Q solutions as $M_n^B$, $M_n^E$ and $M_n^Q$. The distribution is unique in the case of independent votes, so $M_n^B(\mathbf{p}, 0) = M_n^E(\mathbf{p}, 0) = M_n^Q(\mathbf{p}, 0) = M_n(\mathbf{p}, 0)$. For a homogeneous jury, $M_n(\mathbf{p}, \mathbf{c}) = M_n(p, c)$ as $p_i = p$ and $c_{i,j} = c$.

Next I prove two jury theorems based on the B and Q solutions.

**Theorem 1.** *Given the B-solution, the probability of a homogeneous jury collectively making* *the correct decision under simple majority rule is given by*

$$M_n^B(p,c) = M_n(p,0) + cn(n-1)(0.5-p)C_{n-2}^{\frac{n-1}{2}} p^{\frac{n-1}{2}} q^{\frac{n-1}{2}}.$$

*Proof.* Substitute the solution (5) in probability (8). Since $t(\mathbf{v}) = \sum_{i=1}^n v_i$, any sums involving $t$ must be understood as double-sums. We have,

$$\sum_{t=\frac{n+1}{2}}^{n} t = \sum_{t=\frac{n+1}{2}}^{n} C_n^t t = n \sum_{t=\frac{n+1}{2}}^{n} C_{n-1}^{t-1}; \tag{9}$$

$$\sum_{t=\frac{n+1}{2}}^{n} t^2 = \sum_{t=\frac{n+1}{2}}^{n} C_n^t t^2 = n(n-1) \sum_{t=\frac{n+1}{2}}^{n} C_{n-2}^{t-2} + n \sum_{t=\frac{n+1}{2}}^{n} C_{n-1}^{t-1}, \tag{10}$$

and

$$M_n^B(p,c) = M_n(p,0) + \frac{cn(n-1)}{2pq}\Bigg[ \underbrace{\sum_{t=\frac{n+1}{2}}^{n} C_{n-2}^{t-2} p^t q^{n-t}}_{S_1} - 2p \underbrace{\sum_{t=\frac{n+1}{2}}^{n} C_{n-1}^{t-1} p^t q^{n-t}}_{S_2} + p^2 M_n(p,0) \Bigg]. \tag{11}$$

Simplify $S_1$ and $S_2$ using the following identities $C_{n-2}^{\frac{n-3}{2}} = C_{n-2}^{\frac{n-1}{2}} = 0.5 C_{n-1}^{\frac{n-1}{2}}$ and the recursive equations:

$$M_n(p,0) = \mathcal{M}_{n-1} + 2pC_{n-2}^{\frac{n-1}{2}} p^{\frac{n-1}{2}} q^{\frac{n-1}{2}}, \quad \text{where} \quad \mathcal{M}_{n-1} = \sum_{t=\frac{n+1}{2}}^{n-1} C_{n-1}^t p^t q^{n-t-1};$$

$$M_n(p,0) = \mathcal{M}_{n-2} + \frac{p}{q}(3-2p) C_{n-2}^{\frac{n-1}{2}} p^{\frac{n-1}{2}} q^{\frac{n-1}{2}}, \quad \text{where} \quad \mathcal{M}_{n-2} = \sum_{t=\frac{n+1}{2}}^{n-2} C_{n-2}^t p^t q^{n-t-2}.$$

Let $i = t-2$ and $j = t-1$. Then,

$$\begin{aligned}
S_1 &= \sum_{i=\frac{n-3}{2}}^{n-2} C_{n-2}^{i} p^{i+2} q^{n-i-2} = p^2\Big[\mathcal{M}_{n-2} + C_{n-2}^{\frac{n-3}{2}} p^{\frac{n-3}{2}} q^{\frac{n-1}{2}} + C_{n-2}^{\frac{n-1}{2}} p^{\frac{n-1}{2}} q^{\frac{n-3}{2}}\Big] \\
&= p^2\Big[\mathcal{M}_{n-2} + \frac{1}{pq} C_{n-2}^{\frac{n-1}{2}} p^{\frac{n-1}{2}} q^{\frac{n-1}{2}}\Big] = p^2 M_n(p,0) + pq(1+2p) C_{n-2}^{\frac{n-1}{2}} p^{\frac{n-1}{2}} q^{\frac{n-1}{2}}; \\
S_2 &= \sum_{j=\frac{n-1}{2}}^{n-1} C_{n-1}^{i} p^{j+1} q^{n-j-1} = p\Big[\mathcal{M}_{n-1} + C_{n-1}^{\frac{n-1}{2}} p^{\frac{n-1}{2}} q^{\frac{n-1}{2}}\Big] \\
&= pM_n(p,0) + 2pq C_{n-2}^{\frac{n-1}{2}} p^{\frac{n-1}{2}} q^{\frac{n-1}{2}}.
\end{aligned}$$

A substitution of $S_1$ and $S_2$ in (11) completes the proof. □

Several *ceteris paribus* results follow, provided condition (6) on the existence of a distribution holds, and the jury is competent ($p > 0.5$). Two corollaries follow immediately:

**Corollary 1.1.** *$M_n^B(p, c_1) > M_n^B(p, c_2)$ for $c_1 < c_2 < 0$ and $c_2 > c_1 > 0$. Negative correlation increases the competence of a homogeneous and competent jury; positive correlation decreases the competence of such a jury.*

**Corollary 1.2.** *$M_n^B(0.5, c_1) = M_n^B(0.5, c_2)$ for $c_1 \neq c_2$. Correlation has no effect on the competence of a randomizing jury.*

In the following, we assume $c > 0$. Consider $dM_n^B(p,c)/dp$. Following Boland (1989),

$$M_n(p,0) = nC_{n-1}^{\frac{n-1}{2}} \int_0^p x^{\frac{n-1}{2}} (1-x)^{\frac{n-1}{2}} dx \quad \text{whence} \quad \frac{dM_n(p,0)}{dp} = nC_{n-1}^{\frac{n-1}{2}} p^{\frac{n-1}{2}} q^{\frac{n-1}{2}}.$$

Substituting the above expression and using the identity $C_{n-2}^{\frac{n-1}{2}} = 0.5 C_{n-1}^{\frac{n-1}{2}}$ we obtain

$$\frac{dM_n^B(p,c)}{dp} = nC_{n-1}^{\frac{n-1}{2}} p^{\frac{n-3}{2}} q^{\frac{n-3}{2}} [(1 - 0.5c(n-1))pq + 0.5c(n-1)^2(0.5-p)^2].$$

This derivative is positive for $c > 0$ in view of the right hand side of inequality (6), which implies $c \leq 2/(n-1)$. Consequently,

**Corollary 1.3.** *When $c > 0$, $M_n^B(p_1, c) > M_n^B(p_2, c)$ for $p_1 > p_2$. For positive correlation, the higher the individual competence, the higher collective competence is.*

Next, consider $M_n^B(p, c) - M_{n-2}^B(p, c)$. Following Boland (1989),

$$M_n(p, 0) - M_{n-2}(p, 0) = (p - q) C_{n-2}^{\frac{n-1}{2}} p^{\frac{n-1}{2}} q^{\frac{n-1}{2}}.$$

Substituting the above expression and using the identity $(n-1) C_{n-2}^{\frac{n-1}{2}} = 4(n-2) C_{n-4}^{\frac{n-3}{2}}$ we obtain

$$M_n^B(p, c) - M_{n-2}^B(p, c) = C_{n-2}^{\frac{n-1}{2}} p^{\frac{n-1}{2}} q^{\frac{n-1}{2}} \left[ (p - q) + c(n-1)(0.5 - p) \underbrace{\left( n - \frac{n-3}{4pq} \right)}_{F(n)} \right].$$

For $c > 0$, the term involving $c$ is negative if $F(n) > 0$, or when $n < \frac{3}{1-4pq}$.[6] This suggests the final corollary to Theorem 1:

**Corollary 1.4.** *For positive correlation, the effect of larger size on competence is ambiguous, but becomes positive for a sufficiently large size.*

Theorem 1 established that the probability of the jury being correct increases with the juror's competency and decreases with positive correlation between their competencies; and negative correlation is beneficial. However, the second theorem produces a remarkably different result.

**Theorem 2.** *Given the Q-solution, the probability of a homogeneous jury collectively making the correct decision under simple majority rule is independent of the correlation coefficient, i.e. $M_n^Q(p, c) = M_n(p, 0)$.*

[6] The fraction is nonnegative because $\max[pq] = 0.25$ (at $p = 0.5$).

*Proof.* Substitute the solution (7) in the probability (8). To prove the theorem, show

$$\sum_{t=\frac{n+1}{2}}^{n} \left[0.5n(n-1) + 2t(t-n)\right] = 0.$$

In view of $\sum_{t=\frac{n+1}{2}}^{n} n(n-1) = 2^{n-1}n(n-1)$ and the identities (9)-(10), the above equality is equivalent to

$$\sum_{t=\frac{n+1}{2}}^{n} C_{n-1}^{t-1} - \sum_{t=\frac{n+1}{2}}^{n} C_{n-2}^{t-2} = 2^{n-3}.$$

Let us prove the above identity. Subtracting

$$\left(C_{n-1}^{\frac{n-1}{2}} - C_{n-2}^{\frac{n-3}{2}}\right) + \left(C_{n-1}^{\frac{n+1}{2}} - C_{n-2}^{\frac{n-1}{2}}\right) + \cdots + \left(C_{n-1}^{n-1} - C_{n-2}^{n-2}\right) = \left(C_{n-2}^{\frac{n-1}{2}} + C_{n-2}^{\frac{n+1}{2}}\right) + \cdots + 0.$$

This sum equals one half of the sum $\sum_{t=0}^{n} C_{n-2}^{t} = 2^{n-2}$, which is symmetric about $C_{n-2}^{\frac{n-3}{2}} = C_{n-2}^{\frac{n-1}{2}}$, and which follows from the basic identity $\sum_{t=0}^{n} C_{n}^{t} = 2^{n}$. □

For the Q-solution the probability of a jury being correct is independent of the correlation coefficient, provided the jury is homogeneous and the decision is made by voting under simple majority rule. In this case the expertise of a homogeneous jury cannot be impaired or improved by the independence of individual competencies. In order for a correlation to make a difference, either the jury must be heterogeneous or a supermajority must be the decision rule.

The significance of Theorem 2 lies in showing that an empirically relevant second-order dependence structure such as a correlation matrix between the votes conveys little information about the sensitivity of the Condorcet probability to correlation. The possibilities range from completely insensitive, to very sensitive, as the examples in Section 5 show.

### 4.1 The special case $c = 1$

Discussion of Section 3 shows that joint probability distributions typically do not exist for very high correlation. However, in the special case of perfectly positively correlated votes the joint probability distribution exists and is given by:

$$P(v_1 = v_2 = \cdots = v_n = 1) = p;$$

$$P(v_1 = v_2 = \cdots = v_n = 0) = 1 - p.$$

The collective competence of a perfectly correlated jury equals that of a single juror, or $M_n(p, 1) = p$.

Note that a joint probability distribution does not exist when the votes are perfectly negatively correlated. If one juror is correct whenever another is incorrect, and vice versa, then by virtue of binary choice a third juror cannot be simultaneously discordant with the former two. Three jurors cannot be mutually contrarian. Positive correlations reflect common rather than contrarian tendencies. They make high consensus voting outcomes more probable than they would be if the jurors were independent.[7]

## 5 Examples

Table 1 illustrates the effect of correlation for $n = 3$. In the case of equally probable and uncorrelated votes, all eight voting profiles are equally probable (I). Positive correlation makes broad coalitions more probable and tight coalitions less probable, whereas negative correlation has the opposite effect (II, III). All three solutions produce identical distributions. The competence of a randomizing jury does not depend on $c$.

[7]For a further discussion of correlations as a model of preferences, see Kaniovski (2008b).

Example IV of competent and independent jurors corresponds with the assumptions of CJT. The jury's competence is reflected in higher probabilities of voting profiles in which the correct decisions are frequent, and a higher probability of collectively making the correct decision. The distribution in the case of correlated votes is unique.

Introducing positive correlation increases the probability of occurrence of all broad coalitions, including those with a high proportion of incorrect decisions (V). Yet for the Q-solution, these two effects offset exactly in the subset of voting profiles in which the jury collectively decides correctly, leaving $M_n^Q(\mathbf{p}, \mathbf{c})$ unchanged. Example VI shows that this does not hold for heterogeneous juries.

All examples suggest that simple majority rule is essential for invariance of $M_n^Q(\mathbf{p}, \mathbf{c})$. In all examples positive correlation improves the likelihood of the correct decision being reached when a supermajority is required to reach a verdict, but the higher the supermajority, the lower the likelihood that the jury will arrive at the correct decision.

TABLE 1 ABOUT HERE

## 6 The bounds approach

The fact that the marginal probabilities and correlation coefficients do not uniquely define a distribution necessitates choosing a particular distribution with all the arbitrariness this entails. The concept of the bounds approach is to find the distributions that minimize or maximize $M_n(\mathbf{p}, \mathbf{c})$ for given $\mathbf{p}$ and $\mathbf{c}$ by solving linear programming problems:

$$\underline{M_n}(\mathbf{p}, \mathbf{c}) = \min_{\pi_\mathbf{v}} M_n(\mathbf{p}, \mathbf{c}) \quad \text{and} \quad \overline{M_n}(\mathbf{p}, \mathbf{c}) = \max_{\pi_\mathbf{v}} M_n(\mathbf{p}, \mathbf{c}),$$

subject to constraints (1)-(4). Probabilities $\underline{M_n}(\mathbf{p},\mathbf{c})$ and $\overline{M_n}(\mathbf{p},\mathbf{c})$ provide the bounds on the jury's competence consistently with a given competence and dependence structure.

Table 2 presents the bounds for the six examples of the previous section. The probabilities and correlation coefficients do not uniquely define a distribution. In all cases, solutions $\underline{\pi}_{\mathbf{v}}$ and $\overline{\pi}_{\mathbf{v}}$ differ due to different third-order correlation coefficients, leading to the bounds on $M_n(\mathbf{p},\mathbf{c})$.

The solutions offer a simple means of generating distinct distributions, as the convexity of $\Delta$ ensures that an infinite number of them can be obtained as $\pi_{\mathbf{v}} = \lambda\underline{\pi}_{\mathbf{v}} + (1-\lambda)\overline{\pi}_{\mathbf{v}}$, $\lambda \in [0,1]$. Note that $\underline{\pi}_{\mathbf{v}} = \overline{\pi}_{\mathbf{v}}$ does not imply the uniqueness of a distribution satisfying constraints (2)-(4), as they do not uniquely define a distribution.

TABLE 2 ABOUT HERE

# 7 Summary

This paper proves two jury theorems for homogeneous juries with correlated votes. The theorems arise from different solutions to the problem of aggregation of dichotomous choice. The solution is a joint probability distribution on the set of voting profiles that satisfies given probabilities and correlation coefficients. Bahadur (1961) offers a general solution to the aggregation problem in terms of higher-order correlation coefficients. However, his solution requires an excessive number of parameters and extensive data in order to arrive at reliable estimates.

In the empirically relevant case of a dependence structure specified only up to the pairwise correlation coefficients, the distribution is not unique. Different distributions produce different probabilities of a jury being correct. I obtain two non-asymptotic theorems for homogeneous juries with correlated votes, as explicit solutions exist for such juries only.

The first theorem uses Bahadur's solution truncated to the pairwise correlations. It shows that negative correlation increases the competence of the jury, while positive correlation decreases the competence of the jury. It also shows that for positive correlation, an enlargement of the jury can be detrimental up to a certain size beyond which it becomes beneficial. The second theorem finds a family of distributions for which this probability under simple majority rule is independent of the correlation coefficient. The two theorems are obtained using two different solutions to the aggregation problem. Bounds on the admissible correlation coefficient for each solution are provided.

The multiplicity of solutions when only pairwise correlations are known allows us to compute the bounds on the competence of a jury as the maximum and minimum probability of it being correct for a given competence and dependence structure.

I conclude the paper by emphasizing the approach I have discussed is sufficiently general to be useful in empirical applications. First, it can be used to construct an accurate model of the jury based on prior beliefs about the competence of individual jurors and the degree of commonality or rivalry among their competencies. Once expressed in terms of the probabilities and correlation coefficients, these beliefs could be used to forecast the probabilities of different voting outcomes, including the probability of a jury making the correct decision under any given voting rule. Second, it can be used for the empirical modeling of actual juries based on past voting behavior. This line of research includes the empirical testing of the independence assumption.

# References


Austen-Smith, D. and Banks, J. S.: 1996, Information aggregation, rationality and the Condorcet jury theorem, *American Political Science Review* **90**, 34–45.

Bahadur, R. R.: 1961, A representation of the joint distribution of responses to $n$ dichotomous items, *in* H. Solomon (ed.), *Studies in item analysis and prediction*, Stanford University Press, pp. 158–168.

Ben-Yashar, R. and Paroush, J.: 2000, A nonasymptotic Condorcet jury theorem, *Social Choice and Welfare* **17**, 189–199.

Berend, D. and Paroush, J.: 1998, When is Condorcet's jury theorem valid?, *Social Choice and Welfare* **15**, 481–488.

Berend, D. and Sapir, L.: 2005, Monotonicity in Condorcet jury theorem, *Social Choice and Welfare* **24**, 83–92.

Berend, D. and Sapir, L.: 2007, Monotonicity in Condorcet's jury theorem with dependent voters, *Social Choice and Welfare* **28**, 507–528.

Berg, S.: 1993, Condorcet's jury theorem, dependency among jurors, *Social Choice and Welfare* **10**, 87–95.

Boland, P. J.: 1989, Majority systems and the Condorcet jury theorem, *The Statistician* **38**, 181–189.

Boland, P. J., Proschan, F. and Tong, Y. L.: 1989, Modelling dependence in simple and indirect majority systems, *Journal of Applied Probability* **26**, 81–88.

Feddersen, T. and Pesendorfer, W.: 1997, Voting behavior and information aggregation in elections with private information, *Econometrica* **65**, 1029–1058.

Kaniovski, S.: 2008a, The exact bias of the Banzhaf measure of power when votes are neither equiprobable nor independent, *Social Choice and Welfare* **31**, 281–300.

Kaniovski, S.: 2008b, Straffin meets Condorcet. What can a voting power theorist learn from a jury theorist?, *Homo Oeconomicus* **forthcoming**.

Kaniovski, Y. M. and Pflug, G. C.: 2007, Risk assessment for credit portfolios: a coupled Markov chain model, *Journal of Banking & Finance* **31**, 2303–2323.

Ladha, K. K.: 1992, The Condorcet's jury theorem, free speech and correlated votes, *American Journal of Political Science* **36**, 617–634.

Ladha, K. K.: 1993, Condorcet's jury theorem in light of de Finetti's theorem: majority-rule voting with correlated votes, *Social Choice and Welfare* **10**, 69–85.

Ladha, K. K.: 1995, Information polling through majority rule voting: Condorcet's jury theorem with correlated votes, *Journal of Economic Behavior and Organization* **26**, 353–372.

Lazarsfeld, P. F.: 1956, Some observations on dichotomous systems. Sociology Department, Columbia University, New York.

Levine, R. D. and Tribus, M. (eds): 1978, *Maximum entropy formalism*, The MIT Press.

Page, S. E.: 2006, Path dependence, *Quarterly Journal of Political Science* **1**, 87–115.

Penrose, L. S.: 1946, The elementary statistics of majority voting, *Journal of Royal Statistical Society* **109**, 53–57.

Van Der Geest, P. A. G.: 2005, The binomial distribution with dependent Bernoulli trials, *Journal of Statistical Computation and Simulation* **75**, 141–154.

Young, H. P.: 1988, Condorcet's theory of voting, *American Political Science Review* **82**, 1231–1244.

Table 1: The probability of a correct collective decision $M_3(\mathbf{p}, \mathbf{c})$

| | | | | I | II | III | IV | V | VI |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $p = 0.5$ $c = 0$ | $p = 0.5$ $c = 0.2$ | $p = 0.5$ $c = -0.2$ | $p = 0.75$ $c = 0$ | $p = 0.75$ $c = 0.2$ | $p_1 = 0.75$ $p_{2,3} = 0.6$ $c = 0.2$ |
| $v_1$ | $v_2$ | $v_3$ | | | | B-solution | | | |
| 1 | 1 | 1 | $\surd$ | 0.125 | 0.2 | 0.05 | 0.422 | 0.506 | 0.357 |
| 1 | 1 | 0 | $\surd$ | 0.125 | 0.1 | 0.15 | 0.141 | 0.094 | 0.136 |
| 1 | 0 | 1 | $\surd$ | 0.125 | 0.1 | 0.15 | 0.141 | 0.094 | 0.136 |
| 1 | 0 | 0 | | 0.125 | 0.1 | 0.15 | 0.047 | 0.056 | 0.122 |
| 0 | 1 | 1 | $\surd$ | 0.125 | 0.1 | 0.15 | 0.141 | 0.094 | 0.051 |
| 0 | 1 | 0 | | 0.125 | 0.1 | 0.15 | 0.047 | 0.056 | 0.056 |
| 0 | 0 | 1 | | 0.125 | 0.1 | 0.15 | 0.047 | 0.056 | 0.056 |
| 0 | 0 | 0 | | 0.125 | 0.2 | 0.05 | 0.016 | 0.044 | 0.086 |
| $M_3^B(\mathbf{p}, \mathbf{c})$ | | | | 0.5 | 0.5 | 0.5 | 0.844 | 0.788 | 0.679 |
| $v_1$ | $v_2$ | $v_3$ | | | | E-solution | | | |
| 1 | 1 | 1 | $\surd$ | 0.125 | 0.2 | 0.05 | 0.422 | 0.498 | 0.351 |
| 1 | 1 | 0 | $\surd$ | 0.125 | 0.1 | 0.15 | 0.141 | 0.102 | 0.141 |
| 1 | 0 | 1 | $\surd$ | 0.125 | 0.1 | 0.15 | 0.141 | 0.102 | 0.141 |
| 1 | 0 | 0 | | 0.125 | 0.1 | 0.15 | 0.047 | 0.048 | 0.116 |
| 0 | 1 | 1 | $\surd$ | 0.125 | 0.1 | 0.15 | 0.141 | 0.102 | 0.057 |
| 0 | 1 | 0 | | 0.125 | 0.1 | 0.15 | 0.047 | 0.048 | 0.051 |
| 0 | 0 | 1 | | 0.125 | 0.1 | 0.15 | 0.047 | 0.048 | 0.051 |
| 0 | 0 | 0 | | 0.125 | 0.2 | 0.05 | 0.016 | 0.052 | 0.092 |
| $M_3^E(\mathbf{p}, \mathbf{c})$ | | | | 0.5 | 0.5 | 0.5 | 0.844 | 0.804 | 0.691 |
| $v_1$ | $v_2$ | $v_3$ | | | | Q-solution | | | |
| 1 | 1 | 1 | $\surd$ | 0.125 | 0.2 | 0.05 | 0.422 | 0.478 | 0.336 |
| 1 | 1 | 0 | $\surd$ | 0.125 | 0.1 | 0.15 | 0.141 | 0.122 | 0.156 |
| 1 | 0 | 1 | $\surd$ | 0.125 | 0.1 | 0.15 | 0.141 | 0.122 | 0.156 |
| 1 | 0 | 0 | | 0.125 | 0.1 | 0.15 | 0.047 | 0.028 | 0.102 |
| 0 | 1 | 1 | $\surd$ | 0.125 | 0.1 | 0.15 | 0.141 | 0.122 | 0.072 |
| 0 | 1 | 0 | | 0.125 | 0.1 | 0.15 | 0.047 | 0.028 | 0.036 |
| 0 | 0 | 1 | | 0.125 | 0.1 | 0.15 | 0.047 | 0.028 | 0.036 |
| 0 | 0 | 0 | | 0.125 | 0.2 | 0.05 | 0.016 | 0.072 | 0.106 |
| $M_3^Q(\mathbf{p}, \mathbf{c})$ | | | | 0.5 | 0.5 | 0.5 | 0.844 | 0.844 | 0.72 |

$\surd$ indicates correct collective decisions

Table 2: Bounds on a jury's competence $\underline{M_3}(\mathbf{p},\mathbf{c})$, $\overline{M_3}(\mathbf{p},\mathbf{c})$

| | | | | I | II | III | IV | V | VI |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $p=0.5$ $c=0$ | $p=0.5$ $c=0.2$ | $p=0.5$ $c=-0.2$ | $p=0.75$ $c=0$ | $p=0.75$ $c=0.2$ | $p_1=0.75$ $p_{2,3}=0.6$ $c=0.2$ |
| $v_1$ | $v_2$ | $v_3$ | | | | | $\underline{\pi}_{\mathbf{v}}$ | | |
| 1 | 1 | 1 | $\surd$ | 0.25 | 0.3 | 0.1 | 0.438 | 0.55 | 0.408 |
| 1 | 1 | 0 | $\surd$ | 0 | 0 | 0.1 | 0.125 | 0.05 | 0.084 |
| 1 | 0 | 1 | $\surd$ | 0 | 0 | 0.1 | 0.125 | 0.05 | 0.084 |
| 1 | 0 | 0 | | 0.25 | 0.2 | 0.2 | 0.063 | 0.1 | 0.173 |
| 0 | 1 | 1 | $\surd$ | 0 | 0 | 0.1 | 0.125 | 0.05 | 0 |
| 0 | 1 | 0 | | 0.25 | 0.2 | 0.2 | 0.063 | 0.1 | 0.108 |
| 0 | 0 | 1 | | 0.25 | 0.2 | 0.2 | 0.063 | 0.1 | 0.108 |
| 0 | 0 | 0 | | 0 | 0.1 | 0 | 0 | 0 | 0.035 |
| $\underline{M_3}(\mathbf{p},\mathbf{c})$ | | | | 0.25 | 0.3 | 0.4 | 0.813 | 0.7 | 0.577 |
| $v_1$ | $v_2$ | $v_3$ | | | | | $\overline{\pi}_{\mathbf{v}}$ | | |
| 1 | 1 | 1 | $\surd$ | 0 | 0.1 | 0 | 0.375 | 0.45 | 0.300 |
| 1 | 1 | 0 | $\surd$ | 0.25 | 0.2 | 0.2 | 0.188 | 0.15 | 0.192 |
| 1 | 0 | 1 | $\surd$ | 0.25 | 0.2 | 0.2 | 0.188 | 0.15 | 0.192 |
| 1 | 0 | 0 | | 0 | 0 | 0.1 | 0 | 0 | 0.066 |
| 0 | 1 | 1 | $\surd$ | 0.25 | 0.2 | 0.2 | 0.188 | 0.15 | 0.108 |
| 0 | 1 | 0 | | 0 | 0 | 0.1 | 0 | 0 | 0 |
| 0 | 0 | 1 | | 0 | 0 | 0.1 | 0 | 0 | 0 |
| 0 | 0 | 0 | | 0.25 | 0.3 | 0.1 | 0.063 | 0.1 | 0.142 |
| $\overline{M_3}(\mathbf{p},\mathbf{c})$ | | | | 0.75 | 0.7 | 0.6 | 0.938 | 0.9 | 0.792 |

$\surd$ indicates correct collective decisions