Theorems for Exchangeable Binary Random Variables with Applications to Voting Theory

Serguei Kaniovski

Autumn 2018

WIFO

The Condorcet Jury Theorem (CJT)

Condorcet (1743-1794) argued for the superiority of collective over individual wisdom. His 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

CJT is a mathematical formalization of his ideas

- any jury comprising an odd number of jurors is more likely than any single juror to select the correct alternative,
- this likelihood becomes a certainty as the size of the jury tends to infinity

The theorem rests on four assumptions

- the jury decides between two alternatives by simple majority rule,
- each juror has a more than an even probability of being correct (competence),
- these probabilities are equal for all jurors (homogeneity),
- each juror votes independently (stochastic independence)

Overview of the presentation

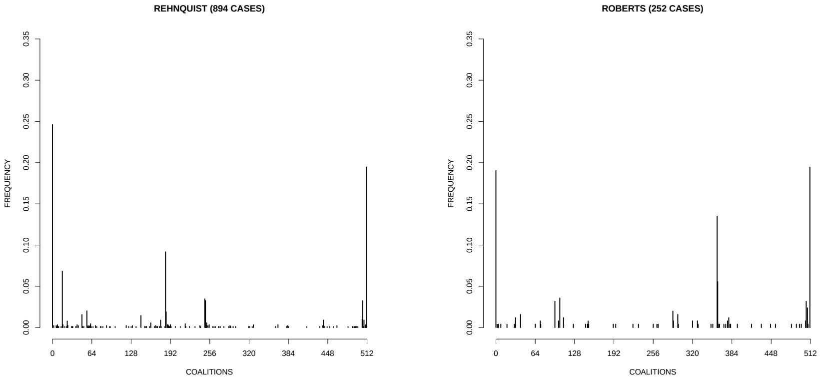
- Examples of expertise and power computations based on a non-trivial probability distribution over the set of all voting outcomes
- The property of exchangeability as a stochastic model of a representative agent (voter)
- Known parameterizations of the joint probability distribution of n correlated binary random variables
- The probability of at least k successes in n correlated binary trials. The generalized binomial distribution
- The bounds on this probability when the higher-order correlations are unknown
- Applications to the Condorcet Jury Theorem and voting power in the sense of Penrose-Banzhaf

Examples of voting under simple majority rule ($n = 3$)

$\mathbf{v} = (v_1, v_2, v_3)$			$p = 0.5$, $c = 0$	$p = 0.75$, $c = 0$	$p = 0.75$, $c = 0.2$	$p_1 = 0.75$, $p_{2,3} = 0.6$, $c = 0.2$
1	1	1	0.125	0.422	0.506	0.357
1	1	0	0.125	0.141	0.094	0.136
1	0	1	0.125	0.141	0.094	0.136
1	0	0	0.125	0.047	0.056	0.122
0	1	1	0.125	0.141	0.094	0.051
0	1	0	0.125	0.047	0.056	0.056
0	0	1	0.125	0.047	0.056	0.056
0	0	0	0.125	0.016	0.044	0.086
Condorcet probability			0.5	0.844	0.788	0.679
Banzhaf probability 1			0.5	0.376	0.3	0.384
Banzhaf probability 2			0.5	0.376	0.3	0.365
Banzhaf probability 3			0.5	0.376	0.3	0.365

Computing the probability of a correct verdict, or the voting power as the probability of casting a decisive vote, requires a joint probability distribution on the set of all voting profiles $\mathbf{v} \in \mathbb{R}^{2^n}$. The influence of voting weights and decision rule is separate from that of the distribution. Exchangeability leads to a representative agent model, in which the independence assumption is relaxed

Voting in the U.S. Supreme Court



Empirical evidence overwhelmingly refutes the assumption of independent votes required in the classic versions of the Condorcet Jury Theorem and the Banzhaf measure of voting power

The joint probability distribution of n binary r.v.

The Bahadur parametrization

$$
\begin{array}{lll}
Z_i = (V_i - p_i)/\sqrt{p_i(1-p_i)} & \text{for all } i = 1, 2, \ldots, n, & p_i = p \\
c_{i,j} = E(Z_i Z_j) & \text{for all } 1 \le i < j \le n, & c_{i,j} = c \\
c_{i,j,k} = E(Z_i Z_j Z_k) & \text{for all } 1 \le i < j < k \le n, & c_{i,j,k} = c_3 \\
\ldots & & \\
c_{1,2,\ldots,n} = E(Z_1 Z_2 \ldots Z_n), & & c_{1,2,\ldots,n} = c_n
\end{array}
$$

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

where $\bar{\pi}_{\mathbf{v}} = \prod_{i=1}^{n} p_i^{v_i}(1-p_i)^{(1-v_i)}$ is the probability under the independence

The George - Bowman (1995) parametrization for exchangeable r.v.

$$\pi_i = \sum_{j=0}^{i} (-1)^j C_i^j \lambda_{n-i+j}, \text{ where } \lambda_i = P(X_1 = 1, X_2 = 1, \ldots, X_i = 1), \quad \lambda_0 = 1$$

The generalized binomial distribution

The probability of at least k successes in n correlated binary trials

$$P_n^k(\mathbf{p}, \mathbf{C}) = P_n^k(\mathbf{p}, \mathbf{I}) + \prod_{h=1}^{n} p_h \sum_{1 \le i < j \le n} c_{i,j} \alpha_i \alpha_j A_{i,j}^k(\boldsymbol{\alpha}), \quad \text{where}$$

$$A_{i,j}^k(\boldsymbol{\alpha}) = \begin{cases} 0, & k = 0 \\ \displaystyle\sum_{\substack{i_s \neq i,\, i_s \neq j \\ 1 \le i_1 < \cdots < i_{n-k} \le n}} \alpha_{i_1}^2 \ldots \alpha_{i_{n-k}}^2 - \sum_{\substack{j_s \neq i,\, j_s \neq j \\ 1 \le j_1 < \cdots < j_{n-k-1} \le n}} \alpha_{j_1}^2 \ldots \alpha_{j_{n-k-1}}^2, & k = 1, \ldots, n-1 \\ 1, & k = n \end{cases}$$

Here $\mathbf{p}$ is the vector of marginal probabilities, $\boldsymbol{\alpha}$ such that $\alpha_i = \sqrt{(1 - p_i)/p_i}$, $\mathbf{C} = (c_{ij})$ the $n \times n$ correlation matrix, $\mathbf{I}$ the $n \times n$ identity matrix

In the following we will use a simpler formula, in which the r.v. are exchangeable and all higher-order correlations vanish. In this case the distribution is completely defined by n, p and the second-order correlation coefficient c

The 3-parameter generalized binomial distribution

For an odd n, $c_3 = c_4 = \cdots = c_n = 0$ and $(p, c) \in \mathcal{B}_n$ (Bahadur set)

$$P_n^k(p, 0) = \sum_{t=k}^{n} C_n^t p^t (1-p)^{n-t} = I_p(k, n-k+1)$$

$$P_n^k(p, c) = I_p(k, n-k+1) + 0.5c(n-1)\left(\frac{k-1}{n-1} - p\right)\frac{\partial I_p(k, n-k+1)}{\partial p}$$

where $I_x(a, b) = B_x(a, b)/B(a, b)$ is the regularized incomplete beta function

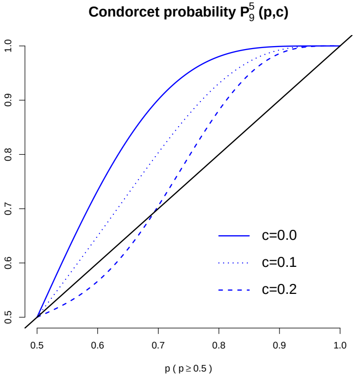
Bounds on $P_{n,p}^k$ for given n and p can be found by linear programming. Di Cecco (2011) provides bounds for given n, p and c such that $(p, c) \in \mathcal{B}_n$

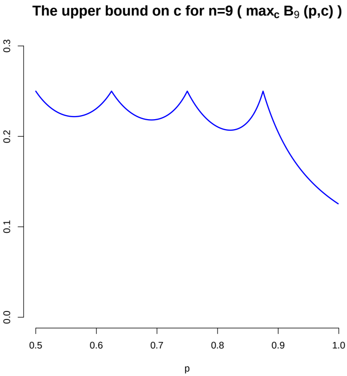
Bounds on $P_{n,p}^k$ when all correlation coefficients are unknown

$$\max\left\{\frac{np-k+1}{n-k+1}, 0\right\} \le P_{n,p}^k(c, c_3, \ldots, c_n) \le \min\left\{\frac{np}{k}, 1\right\}$$

The Condorcet probability and the Bahadur set

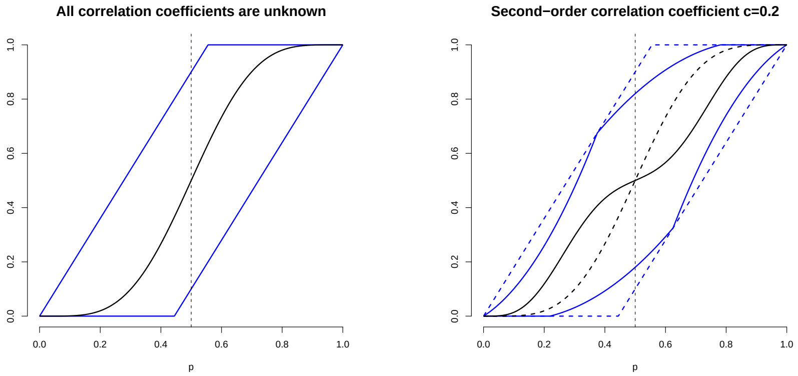
The set $\mathcal{B}_n$ contains all admissible values of c for given n and p, provided $c_3 = c_4 = \cdots = c_n = 0$





$\mathcal{B}_n$ is such that $0 < c < \frac{1}{n-1}$ for $p \approx 1$ and $0 < c < \frac{2}{n-1}$ for $p \approx 0.5$

Bounds on the probability of at least 5 successes in 9 trials



Generating a joint probability distribution (Regularization)

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:

$$\pi_{\mathbf{v}} \geq 0 \quad \text{for all} \quad \mathbf{v} \in \mathbf{V}, \tag{1}$$

$$\sum_{\mathbf{v} \in \mathbf{V}} \pi_{\mathbf{v}} = 1, \tag{2}$$

$$\sum_{\mathbf{v} \in \mathbf{V}(i)} \pi_{\mathbf{v}} = p_i \quad \text{for all} \quad i = 1, 2, \ldots, n, \tag{3}$$

$$\sum_{\mathbf{v} \in \mathbf{V}(i,j)} \pi_{\mathbf{v}} = p_i p_j + c_{i,j}\sqrt{p_i q_i p_j q_j} \quad \text{for all} \quad q_i = 1 - p_i, \quad 1 \leq i < j \leq n \tag{4}$$

- Constraints (1)-(4) define a convex polytope $\Delta \subset \mathbb{R}^{2^n}$
- Any point in Δ is a suitable distribution, but $\Delta = \emptyset$ can occur
- The system (2)-(4) of $1 + n + C_n^2$ equations for 2^n unknowns typically has infinitely many solutions for $n \geq 3$
- Regularization offers a practical method of finding a particular solution

Two suggestions on regularization

Quadratic metric

Given a benchmark distribution $d_{\mathbf{v}}$ and a distance measure f, choose a point in Δ that is closest to $d_{\mathbf{v}}$ in the sense of f. For example, take $d_{\mathbf{v}} = \bar{\pi}_{\mathbf{v}}$, the distribution in the case of independent votes, and f as the sum of squares of deviations

$$\min_{\pi_{\mathbf{v}}} \quad 0.5 \sum_{\mathbf{v}} (\pi_{\mathbf{v}} - \bar{\pi}_{\mathbf{v}})^2, \quad \text{subject to} \quad (1) - (4)$$

Solution vectors may have zero coordinates

Entropy function (Van Der Geest 2005)

Choose f so as to maximize the entropy function $H = -f$

$$\min_{\pi_{\mathbf{v}}} \quad \sum_{\mathbf{v}} \pi_{\mathbf{v}} \log(\pi_{\mathbf{v}}), \quad \text{subject to} \quad (2) - (4)$$

Solution vectors have strictly positive coordinates

A sufficient condition based on the generalized distribution

Assume an odd $n \geq 3$, $p_i > 0.5$ and $c_{i,j} = c > 0$. The effect of a common positive correlation on the collective expertise is negative under simple majority rule and positive under unanimity rule. This implies the existence of intermediate rules $\frac{n+1}{2} < k < n$, for which the effect changes its sign

Sufficient Condition 1

- If the sum of any $k-1$ of reciprocal probabilities $\{p_s^{-1}\}$ is not smaller than $n-1$, then $\forall i, j,\ i < j$, we have $A^m_{i,j}(\boldsymbol{\alpha}) \geq 0$, $\forall m \geq k$,
- If the sum of any $k-1$ of reciprocal probabilities $\{p_s^{-1}\}$ is not greater than $n-1$, then $\forall i, j,\ i < j$, we have $A^m_{i,j}(\boldsymbol{\alpha}) \leq 0$, $\forall m \leq k$

Further sufficient conditions

Sufficient Condition 1 has implications for an optimal choice of the voting rule in relationship to the harmonic mean $H = n(\sum_{i=1}^{n} p_i^{-1})^{-1}$ and the range of individual competences $\{p_i\}$

Sufficient Condition 2

- If the sum of any $k-1$ of $\{p_s^{-1}\}$ is not smaller than $n-1$, then $H \leq \frac{k-1}{n-1}$,
- If the sum of any $k-1$ of $\{p_s^{-1}\}$ is not greater than $n-1$, then $H \geq \frac{k-1}{n-1}$,
- If $p_i \in (0.5, \frac{k-1}{n-1}]$, $\forall i$, then $A_{i,j}^m(\boldsymbol{\alpha}) \geq 0$, $\forall m \geq k$, $\forall i, j$, $i < j$,
- If $p_i \in [\frac{k-1}{n-1}, 1)$, $\forall i$, then $A_{i,j}^m(\boldsymbol{\alpha}) \leq 0$, $\forall m \leq k$, $\forall i, j$, $i < j$

and

Sufficient Condition 3

If there exist $\frac{n+1}{2} \leq m < l \leq n$ such that all $\{p_i\}$ lie within the interval $[\frac{m-1}{n-1}, \frac{l-1}{n-1}]$, then $A_{i,j}^{(n+1)/2}(\boldsymbol{\alpha}) < 0, \ldots, A_{i,j}^m(\boldsymbol{\alpha}) \leq 0, A_{i,j}^l(\boldsymbol{\alpha}) \geq 0, \ldots, A_{i,j}^n(\boldsymbol{\alpha}) > 0$, $\forall i, j$, $i < j$

Conclusions: Condorcet Jury Theorem

- The effect of correlation on the jury's competence is negative for voting rules close to simple majority and positive for voting rules close to unanimity
- If the individual competence is low, it may be better to hire one expert rather than several. In all other cases simple majority rule is the optimal decision rule. A jury operating under simple majority rule will not necessarily benefit from an enlargement, unless the enlargement is substantial. The higher the individual competence, the sooner an enlargement will be beneficial
- Correlation-robust voting rules minimizes the effect of correlation on collective competence by making it as close as possible to that of a jury of independent jurors. The optimal correlation-robust voting rule should be preferred to simple majority rule if mitigating the effect of correlation is more important than maximizing the accuracy of the jury decision
- For a given competence, compute the bounds to a jury's competence as the minimum and maximum probability of a jury being correct
- Using the generalized binomial distribution, we generalize the Condorcet Jury Theorem by allowing heterogeneity of experts, positive correlation between the votes, and qualified majority rules. For analytical tractability, we assume that any two votes correlate with the same correlation coefficient. The conventional wisdom holds that the groupthink or bandwagon effects diminish the collective competence. We show that this effect can be positive or negative, and provide sufficient conditions for it to have a certain sign

Probability of casting a decisive vote (Banzhaf)

In a ‘one person, one vote’ election with two alternatives a vote is decisive if it breaks a tie. With $n+1$ voters, the probability of a tie equals $C_n^{\frac{n}{2}}\pi_{\frac{n}{2}}$

For an odd n, $c_3 = c_4 = \cdots = c_n = 0$ and $(p, c) \in \mathcal{B}_n$

$$V_n^k(p,0) = C_n^{\frac{n}{2}} p^{\frac{n}{2}}(1-p)^{\frac{n}{2}}$$

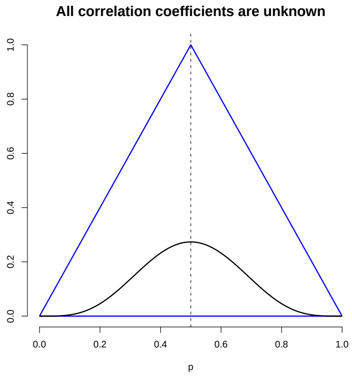
$$V_n^k(p,c) = V_n^k(p,0) + \frac{nc}{4} C_n^{\frac{n}{2}} p^{\frac{n-2}{2}}(1-p)^{\frac{n-2}{2}} \left(\frac{n(2p-1)^2}{2} + 2p(1-p) - 1\right)$$

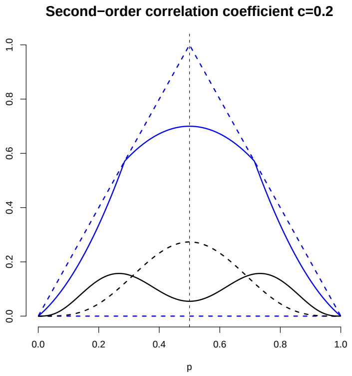
Bounds on $V_n^k(c, c_3, \ldots, c_n)$ when all coefficients are unknown

$$0 \leq V_n^k(c, c_3, \ldots, c_n) \leq 2\min\{p, 1-p\}$$

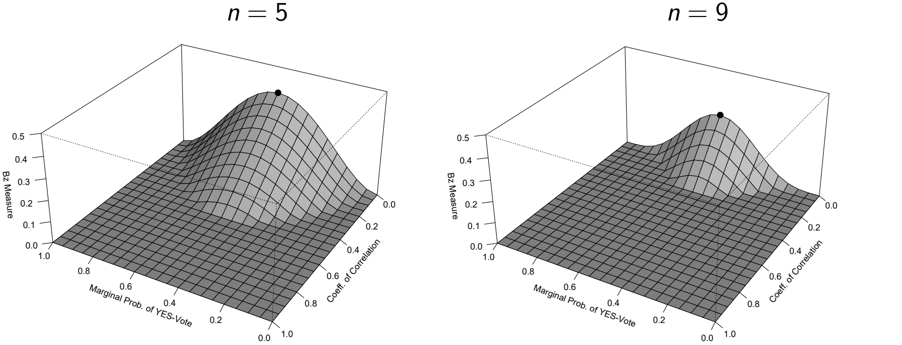
For given n, p and c, bounds can be found by linear programming

Bounds the probability of casting a decisive vote





The bias in Banzhaf measure of voting power



The conditional probability of a decisive YES vote, when $p_i = p$ and $c_{i,j} = c \geq 0$ for all i, j. The Bz measure is unbiased when $p = 0.5$ and $c = 0$ (filled points). The probability bias incurred by p deviating from 0.5 is polynomial, whereas the correlation bias incurred by c deviating from 0 is linear

Conclusions: Voting power

- We can assess the magnitude of numerical error or bias in the Banzhaf measure that occurs when equiprobability and independence assumptions are not met. The probability bias is more severe than the correlation bias. Common positive correlation biases the measure upwards, common negative correlation downwards
- Despite the Banzhaf measure being a valid measure of *a priori* voting power and thus useful for evaluating the rules at the constitutional stage of a voting body, it is a poor measure of the actual probability of being decisive at any time past that stage
- Derive a modified square-root rule for the representation in two-tier voting systems that takes into account the sizes of the constituencies and the heterogeneity of their electorates. Since in a homogeneous electorate the votes are positively correlated, the larger and the more homogeneous the electorate, the less power a vote has
- Develop realistic voting scenarios that reflect the preferences of the voters via a correlation matrix. Then generate a consistent joint probability distribution and compute the probabilities of interest
- Compute the bounds to voting power as the minimum and maximum probability of the voter being decisive

Literature 1

Parameterizations

- Bahadur, R.R., "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
- George, E. O., Bowman, D., "A full likelihood procedure for analyzing exchangeable binary data", *Biometrics*, 1995, 51, pp. 512–523
- Zaigraev, A., Kaniovski, S., "A note on the probability of at least K successes in N correlated binary trials", *Operations Research Letters*, 2013, 41(1), pp. 116–120

Bounds

- Zaigraev, A., Kaniovski, S., "Exact bounds on the probability of at least k successes in n exchangeable Bernoulli trials as a function of correlation coefficients", *Statistics & Probability Letters*, 2010, 80(13–14), pp. 1079–1084
- Di Cecco, D., "A geometric approach to a class of optimization problems concerning exchangeable binary variables", *Statistics & Probability Letters*, 2011, 81(3), pp. 411–416

Literature 2

Condorcet Jury Theorem

- Kaniovski, S., Zaigraev, A., “Optimal jury design for homogeneous juries with correlated votes”, *Theory and Decision*, 2011, 71(4), pp. 439–459
- Kaniovski, S., “Aggregation of correlated votes and Condorcet’s Jury Theorem”, *Theory and Decision*, 2010, 69(3), pp. 453–468 (**Generating Joint Probability Distributions**)
- Kaniovski, S., “An invariance result for homogeneous juries with correlated votes”, *Mathematical Social Sciences*, 2009, 57(2), pp. 213–222 (**Mute the Effect of Correlation on Collective Competence**)

Voting Power

- Straffin, P. D., “Homogeneity, independence and power indices”, *Public Choice*, 1977, 30, pp. 107–118 (**Stochastic Models for Voting Power Indices**)
- Kaniovski, S., “The exact bias of the Banzhaf measure of power when votes are neither equiprobable nor independent”, *Social Choice and Welfare*, 2008, 31(2), pp. 281–300 (**Generating Joint Probability Distributions**)
- Kurz, S. Napel. S., “The roll call interpretation of the Shapley value”, *Economics Letters*, 2018, 173(2), pp. 108–112 (**Exchangeability and the Shapley-Shubik Index**)