Homo Oeconomicus 24(3/4): 357–380 • (2007)

www.accedoverlag.de

# On Penrose's Square-root Law and Beyond

Werner Kirsch
Fakultät für Mathematik und Informatik, FernUniversität Hagen, Germany
(eMail: werner.kirsch@fernuni-hagen.de)

**Abstract** In certain bodies, like the Council of the EU, the member states have a voting weight which depends on the population of the respective state. In this article we ask the question which voting weight guarantees a 'fair' representation of the citizens in the union. The traditional answer, the square-root law by Penrose, is that the weight of a state (more precisely: the voting power) should be proportional to the square-root of the population of this state. The square root law is based on the assumption that the voters in every state cast their vote independently of each other. In this paper we concentrate on cases where the independence assumption is not valid.

## 1. Introduction

All modern democracies rely on the idea of representation. A certain body of representatives, a parliament for example, makes decisions on behalf of the voters. In most parliaments each of its members represents roughly the same number of people, namely the voters in his or her constituency.

There are other bodies in which the members represent different numbers of voters. A prominent example is the Council of the European Union. Here ministers of the member states represent the population of their respective country. The number of people represented in the different states differs from about 400,000 for Malta to more than 82 million for Germany. Due to this fact the members of the Council have a certain number of votes depending on the size of the country they represent, e.g. 3 votes for Malta, 29 votes for Germany. The votes of a country cannot be split, but have to be cast as a block.[1]

Similar voting systems occur in various other systems, for example in the

[1] The current voting system in the Council is based on the treaty of Nice. It has additional components to the procedure described above, which are irrelevant in the present context. For a description of this voting system and further references see e.g. (Kirsch Preprint).



ISBN 978-3-89265-066-9 ISSN 0943-0180

Bundesrat, Germany's state chamber of parliament and in the electoral college in the USA.[2]

Let us call such a system in which the members represent subsystems (states) of different size a *heterogeneous voting system*. In the following we will call the assembly of representatives in a heterogeneous voting system the *council*, the sets of voters represented by the council members the *states*.

It is quite clear, that in a heterogeneous voting system a bigger state (by population) should have at least as many votes in the council as a smaller state. It may already be debatable whether the bigger states should have *strictly* more votes than the smaller states (cf. the Senate in the US constitution). And if yes, how much more votes the bigger state should get?

In this note we address the question: 'What is a fair distribution of power in a heterogeneous voting system?'

There exist various answers to this question, depending on the interpretation of the words 'fair' and 'power'.

The usual and quite reasonable way to formulate the question in an exact way is to use the concept of power indices. One calls a heterogeneous voting system fair if all voters in the member states have the same influence on decisions of the council. By 'same influence' we mean that the power index of each voter is the same regardless of her or his home state. If we choose then Banzhaf power index to measure the influence of a voter we obtain the celebrated Penrose's square-root law (see e.g. (Felsenthal and Machover 1998)).

The square-root law states that the distribution of power in a heterogeneous voting system is fair if the power (index) of each council member $i$ is proportional to $\sqrt{N_i}$, where $N_i$ is the population of the state which $i$ represents.

In their book (Felsenthal and Machover 1998) Felsenthal and Machover formulate a second square-root law. There they base the notion of 'fairness' on the concept of *majority deficit*.

The majority deficit is zero if the voters favoring the decision of the council are the *majority*. If the voters favoring the decision of the council are the *minority* then the majority deficit is the margin between the number of voters objecting to the decision and those agreeing with it (see Def. 3.3.16 in (Felsenthal and Machover 1998)).

The notion of fairness we propose in this paper is closely related to the concept of majority deficit. We will call a decision of the council *in agreement with the popular vote* if the percentage of voters agreeing with a proposal (popular vote) is as close as possible to the percentage of council votes in favor of the proposal. (We will make this notion precise in the next section.)

[2] The electoral college is not exactly a heterogeneous voting system in the sense defined below, but it is very close to it.

For both concepts we have to average over the possible voting configurations. This is usually done by assuming that voters vote independently of each other. The main purpose of this note is to investigate some (we believe reasonable) models where voters do not vote independently.

We will discuss two voting models with voting behavior which is *not* independent. The first model considers societies which have some kind of 'collective bias' (or 'common belief'). A typical situation of this kind is a strong religious group (or church) influencing the voting behavior of the voters. This model is discussed in detail in Section 3.

In the other model voters tend to vote the same way 'the majority does'. This is a situation where voters do not want to be different from others. We call this the *mean field model* referring to an analogous model from statistical physics. See Section 5 for this model.

In fact, both models can be interpreted in terms of statistical physics. Statistical physics considers (among many other things) magnetic systems. The elementary magnet, called a spin, has two possible states which are '+1' or '−1' (spin up, spin down). This models voting 'yes' or 'no' in a voting system. Physicists consider different kinds of interactions between the single spins, one given through an exterior magnetic-field - corresponding to a society with 'a collective bias' - or through the tendency of the spins to align - corresponding to the second voting model. We discuss the analogy of voting models with spin systems in Section 4.

Our investigations of voting models with statistical dependence is much inspired by the paper (Laruelle and Valenciano 2005). The first model is also based on the work by Straffin (Straffin 1982).

It does not come as a surprise that we obtain a square-root law for a model with independent voters, just as in the case considered by Felsenthal and Machover ((Felsenthal and Machover 1998)).

For the mean field model we still get a square-root law for the best possible representation in the council *as long as* the mutual interaction between voters is not too strong.

However as the coupling between voters exceeds a certain threshold, the fairest representation in the council is no longer given by votes proportional to $\sqrt{N_i}$ but rather by votes proportional to $N_i$. This is a typical example of a phase transition.

In the model of collective bias the fair representation weight depends on the strength of the collective bias for large populations. If this strength is independent of the population size fair representation is almost always given by voting weights proportional to $N_i$, the square-root law occurring only in marginal cases. However, if the collective bias decreases with increasing population one can get any power law behavior ${N_i}^{\alpha}$ for the optimal weight as long as $\frac{1}{2} \leq \alpha \leq 1$. In fact, statistical investigations on real life data suggest that this might happen (see (Gelman et al. 2004)). We leave the mathematical proofs of our results for the appendices.

## 2. The general model

We consider $N$ voters, denoted by $1, 2, \ldots, N$. Each of them may vote 'yes' or 'no'; abstentions are not allowed. The vote of the voter $i$ is denoted by $X_i$.

The possible voting results are $X_i = +1$ representing 'yes' and $X_i = -1$ for 'no'. We consider the quantity $X_i$ as random, more precisely there is a probability measure $\mathbb{P}$ on the space $\{-1, 1\}^N$ of possible voting results. We will call the measure $\mathbb{P}$ a *voting measure* in the following. $\mathbb{P}$ and its properties will be specified later. The conventional assumption on $\mathbb{P}$ is that the random quantities $X_i$ are independent from each other, but we are *not* making this assumption here.

Our interpretation of this model is as follows. The voters react on a proposal in a rational way, that is to say: A voter does *not* roll a dice to determine his or her voting behavior but he or she votes for or against a given proposal according to his/her personal belief, knowledge, experience etc. It is rather the proposal which is the source of randomness in this system. We imagine the voting system is fed with propositions in a completely random way. This could be either a real source of proposals or just a Gedankenexperiment to measure the behavior of the voting system.

The rationality of the voters implies that a voter who casts a 'yes' on a certain proposition will necessarily vote 'no' on the diametrically opposed proposition. Since we assume that the proposals are completely random any proposal and its antithetic proposal must have the same probability. This implies

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}\,. \tag{2.1}$$

More generally, we conclude that

$$\mathbb{P}(X_{i_1} = \xi_1, \ldots, X_{i_r} = \xi_r) = \mathbb{P}(X_{i_1} = -\xi_1, \ldots, X_{i_r} = -\xi_r) \tag{2.2}$$

for any set $i_1 \ldots, i_r$ of voters and any $\xi_1, \ldots \xi_r \in \{-1, 1\}$.

We call the property (2.2) the *symmetry* of the voting system. Any measure $\mathbb{P}$ satisfying (2.2) is called a *voting measure*.

The symmetry assumption (2.2) does *not* fix the probability measure $\mathbb{P}$. Only if we assume in addition that the $X_i$ are statistically independent we can conclude from (2.2) that

$$\mathbb{P}(X_{i_1} = \xi_1, \ldots, X_{i_r} = \xi_r) = \left(\frac{1}{2}\right)^r . \qquad (2.3)$$

So far, we have not specified any decision rule for the voting system. The above probabilistic setup is completely independent from the voting rule, a fact which was emphasized in the work (Laruelle and Valenciano 2005).

A *simple majority rule* for $X_1, \ldots, X_N$ is given by the decision rule: Accept a proposal if $\sum_{j=1}^{N} X_j > 0$ and reject it otherwise.

By a *qualified majority rule* we mean that at least a percentage $q$ (called the *quota*) of votes is required for the acceptance of a proposal. In term of the $X_j$ this means:

$$\sum_{j=1}^{N} X_j \geq (2q-1)N. \qquad (2.4)$$

Indeed, it is not hard to see that the number of affirmative votes is given by

$$\frac{1}{2}\left(\sum_{j=1}^{N} X_j + N\right).$$

»From this the assertion (2.4) follows.

In particular, the simple majority rule is obtained form (2.4) by choosing $q$ slightly bigger than $\frac{1}{2}$.

The sum $\sum_{j=1}^{N} X_j$ gives the difference between the number of 'yes'-votes and the number of 'no'-votes. We call the quantity

$$M(X) := \left|\sum_{j=1}^{N} X_j\right| \qquad (2.5)$$

the margin of the voting outcome $X = (X_1, \ldots, X_N)$. It measures the size of the majority with which the proposal is either accepted or rejected in simple majority voting.

In qualified majority voting with quota $q$ the corresponding quantity is the *q-margin* $M_q(X)$ given by:

$$M_q(X) := \left|\sum_{j=1}^{N} X_j - (2q-1)N\right| . \qquad (2.6)$$

Now, we turn to voting in the council. We consider $M$ states, the state number $\nu$ having $N_\nu$ voters. Consequently the total number of voters is $N = \sum N_\nu$. The vote of the voter $i$ in state $\nu$ is denoted by $X_{\nu i}$, $\nu = 1, \ldots, M$ and $i = 1, \ldots, N_\nu$.[3]

We suppose that each state government knows the opinion of (the majority of) the voters in that state and acts accordingly.[4] That is to say: If the majority of people in state $\nu$ supports a proposal, i.e. if

$$\sum_{i=1}^{N_\nu} X_{\nu i} > 0 \tag{2.7}$$

then the representative of state $\nu$ will vote 'yes' in the council otherwise he or she will vote 'no'. If we set $\chi(x) = 1$ for $x > 0$, $\chi(x) = -1$ for $x \le 0$ the representative of state $\nu$ will vote

$$\xi_\nu = \chi\left(\sum_{i=1}^{N_\nu} X_{\nu i}\right) \tag{2.8}$$

in the council. If the state $\nu$ has got a weight $w_\nu$ in the council the result of voting in the council is given by:

$$\sum_{\nu=1}^{M} w_\nu\, \xi_\nu = \sum_{\nu=1}^{M} w_\nu\, \chi\left(\sum_{i=1}^{N_\nu} X_{\nu i}\right). \tag{2.9}$$

Thus, the council's decision is affirmative if $\sum_{\nu=1}^{M} w_\nu\, \xi_\nu$ is positive, provided the council votes according to simple majority rule.

The result of a popular vote in all countries $\nu = 1, \ldots, N$ is

$$P = \sum_{\nu=1}^{M} \sum_{i=1}^{N_\nu} X_{\nu i}. \tag{2.10}$$

We will call voting weights $w_\nu$ for the council *fair* or *optimal*, if the council's vote is as close as possible to the public vote. To make this precise let us define

$$C = \sum_{\nu=1}^{M} w_\nu\, \chi\left(\sum_{i=1}^{N_\nu} X_{\nu i}\right) \tag{2.11}$$

[3] We label the states using Greek characters and the voters within a state by Roman characters.

[4] Although this is the central idea of representative democracy this idealization may be a little naive in practice.

the result of the voting in the council. Both $P$ and $C$ are random quantities which depend on the random variables $X_{\nu i}$. So, we may consider the mean square distance $\Delta$ between $P$ and $C$, i.e. denoting the expectation over the random quantities by $\mathbb{E}$, we have

$$\Delta = \mathbb{E}\Big((P-C)^2\Big) = \mathbb{E}\left(\left\{\sum_{\nu=1}^{M}\sum_{i=1}^{N_\nu} X_{\nu i} - \sum_{\nu=1}^{M} w_\nu \chi\Big(\sum_{i=1}^{N_\nu} X_{\nu i}\Big)\right\}^2\right). \tag{2.12}$$

In a democratic system the decision of the council should be as close as possible to the popular vote, hence we call a system of weights *fair* or *optimal* if $\Delta = \Delta(w_1, \ldots, w_M)$ is minimal among all possible values of $w_\nu$.

In the following we suppose that the random variables $X_{\nu i}$ and $X_{\mu j}$ are independent for $\nu \neq \mu$. This means that voters in different states are not correlated. We do not assume at the moment that two voters from the same state vote independently of each other.

We have the following result:

*Theorem 2.1* Fair voting in the council is obtained for the values

$$w_\nu = \mathbb{E}\left(\left|\sum_{i=1}^{N_\nu} X_{\nu i}\right|\right) = \mathbb{E}\Big(M(X_\nu)\Big).$$

This result can be viewed as an extension of Penrose's square-root law to the situation of correlated voters. We will see below that it gives $w_\nu \sim \sqrt{N_\nu}$ for independent voters.

Theorem 2.1 has a very easy - we hope convincing - interpretation: $w_\nu$ is the expected margin of the voting result in state $\nu$. In other words, it gives the expected number of people in state $\nu$ that agree with the voting of $\nu$ in their council minus those that disagree, i.e. the net number of voters which the council member of $\nu$ actually represents.

If we choose a multiple $c\,w_1, \ldots, c\,w_{N_\nu}$ $(c > 0)$ of the weights $w_1, \ldots .w_{N_\nu}$ we obtain the same voting system as the one defined by $w_1, \ldots, w_n$. In this sense the weights $w_\nu$ of Theorem 2.1 are not unique, but the voting system is.

We will prove Theorem 2.1 in section .2. We remark that the proof requires the symmetry assumption (2.2) and the independence of voters from *different* states.

The next step is to compute the expected margin $\mathbb{E}(M(X_\nu))$, at least asymptotically for large number of voters $N_\nu$. This quantity depends on the correlation structure between the voters in state $\nu$. As we will see, different correlations between voters give very different results for $\mathbb{E}(M(X_\nu))$ and hence for the optimal weight $w_\nu$.

We begin with the classical case of independent voters.

*Theorem 2.2* If the voters in state $v$ cast their votes independently of each other then

$$\mathbb{E}\left(\left|\sum_{i=1}^{N_v} X_{vi}\right|\right) \sim c\sqrt{N_v} \tag{2.13}$$

for large $N_v$.

Thus, we recover the square-root law as we expected. (For the square-root law see Felsenthal and Machover (Felsenthal and Machover 1998).) In terms of power indices the independence assumption is associated to the Banzhaf power index. Therefore, it is not surprising that also the Banzhaf index leads to a square-root rule.

It is questionable (as we know from the work of Gelman, Katz and Bafumi (Gelman et al. 2004)) whether the independent voters model is valid in many real-life voting systems. This is one of the reasons to extend the model as we do in the present paper.

## 3. The 'collective bias' model

In this section we define and investigate a model we dub the 'collective bias model'. In this model there exists a kind of common opinion in the society from which the individual voters may deviate but to which they agree in the mean. Such a 'common opinion' or 'collective bias' may have very different reasons: There may be a system of common values or common beliefs in the country under consideration, there may be an influential religious group or political ideology, there could be a strong tradition or simply a common interest based on economical needs. A 'collective bias' may also originate in a single person's influence on the media or in the pressure put onto voters by some powerful group. The obviously important differences in the origin of the common opinion are not reflected by the model as the purely technical outcome does not depend on it.

To model the collective bias we introduce a random variable $Z$, the collective bias variable, which takes values between −1 and +1. If $Z > 0$ the collective bias is in favor of the proposition under consideration. The closer the value of $Z$ to 1, the higher the expected percentage of voters in favor of the proposition. In particular, if $Z = 1$ *all* voters will vote 'Yes', while $Z = 0$ means the collective bias is neutral towards the proposal and the voters vote independent of each other and with probability one half for (or against) the proposal. In general, if the collective bias variable $Z$ has the value $\zeta$ the probability that the voter $i$ votes 'Yes' is $p_\zeta = \frac{1}{2}(1+\zeta)$, the probability for a 'No' is consequently $1 - p_\zeta = \frac{1}{2}(1-\zeta)$. The probability $p_\zeta$ is chosen such that the expectation value of $X_i$ is $\zeta$, the value of the 'collective bias' variable $Z$. Thus

$\zeta$ equals the expected fraction of voters supporting the proposal.

We remark that $Z$ is a random variable, which means it depends on the proposal under consideration. This models the fact that there may be a strong common belief on certain issues while there is no or merely a weak common opinion on others. For example, in a country with a strong influence of the catholic church there may be a strong common view about abortion among voters, but, perhaps, not about speed limits on highways or on the details of taxation.

Once the value $\zeta$ of $Z$ is chosen the voters vote independently of each other but with a probability for 'Yes' and 'No' which depends on $\zeta$. The voting results $(X_1, \ldots, X_N)$ are correlated through (and only through) the collective bias $Z$.

In the following we describe the 'collective bias model' in a formal way. We introduce a random variable $Z$ (the 'collective bias') with values in the interval $[-1,1]$ and with a probability distribution $\mu$, which we call the 'collective bias measure'. $\mu([a,b])$ is the probability that $Z$ takes a value in $[a,b]$. For a given $\zeta \in [-1,1]$ we denote by $P_\zeta$ the probability measure on $\{-1,1\}$ with:

$$P_\zeta(\, X_i = 1\,) = p_\zeta \; = \; \frac{1}{2}\,(1+\zeta) \tag{3.1}$$

and

$$P_\zeta(\, X_i = -1\,) = 1 - p_\zeta \; = \; \frac{1}{2}\,(1-\zeta)$$

We set $p_\zeta = \frac{1}{2}(1+\zeta)$. $p_\zeta$ is chosen such that we have $E_\zeta(X_i) = \zeta$ where $E_\zeta$ denotes the expectation value with respect to $P_\zeta$ .

Now, we define the voting measure $\mathbb{P}_\mu$ with respect to the 'collective bias measure' $\mu$. The *conditional probability* with respect to $\mathbb{P}_\mu$ *given* $Z = \zeta$ is obtained from:

$$\mathbb{P}_\mu(X_1 = \xi_1, \ldots, X_N = \xi_N \mid Z = \zeta) = \prod_{i=1}^{N} P_{p_\zeta}(X_i = \xi_i) \tag{3.2}$$

Thus, given the value $\zeta$ of the 'collective bias' variable $Z$, the voters vote independently of each other with expected outcome equal to $\zeta$. As a consequence of (3.2), the measure $\mathbb{P}_\mu$ is given by integrating over $\zeta$, hence:

$$\mathbb{P}_\mu(X_1 = \xi_1, \ldots, X_N = \xi_N) = \int \prod_{i=1}^{N} P_{p_\zeta}(X_i = \xi_i)\; d\mu(\zeta)\,. \tag{3.3}$$

To ensure that the probability $\mathbb{P}_\mu$ satisfies the symmetry condition (2.2) we have to require that

$$\mathbb{P}_\mu(Z \in [a,b]) = \mathbb{P}_\mu(Z \in [-b,-a])$$

i.e.

$$\mu([a,b]) = \mu([-b,-a]) \tag{3.4}$$

The probability measure $\mathbb{P}_\mu$ defines a whole class of examples, each (symmetric) probability measure $\mu$ on $[-1,1]$ defines its unique $\mathbb{P}_\mu$. For example, if we choose $\mu = \delta_0$, i.e. $\mu([a,b]) = 1$ if $a \leq 0 \leq b$ and $= 0$ otherwise, we obtain independent random variables $X_i$ as discussed in the final part of section 2. Indeed, $\mu = \delta_0$ means that $Z = 0$, consequently (3.3) defines independent random variables. Observe, that this is the only measure for which $Z$ assumes a fixed value, since the collective bias measure $\mu$ has to be symmetric (3.4).

Another interesting example is the case when $\mu$ is the uniform distribution on $[-1,1]$, meaning that each value in the interval $[-1,1]$ is equally likely. This case was considered by Straffin (Straffin 1982). He observed that this model is intimately connected with the Shapley-Shubik power index. We will comment on this interesting connection and on Straffin's calculation in an appendix (section .1). To apply the 'collective bias' model to a given heterogeneous voting model we have to specify the measure $\mu$, of course. In fact, this measure may change from state to state. In particular, one may argue that larger states tend to have a less homogeneous population and hence, for example, the influence of a specific religious or political group will be smaller. As an example to this phenomenon, we will later discuss a model modifying Straffin's example where $\mu(dz) = \frac{1}{2}\chi_{[-1,1]}(z)\,dz$ (uniform distribution in $[-1,1]$) to a measure where $\mu_N$ depends on the population $N$, namely

$$\mu_N(dz) = \frac{1}{2a_N}\chi_{[-a_N,a_N]}(z)dz \tag{3.5}$$

with parameters $0 < a_N \leq 1$. In particular, if we have $a_N \to 0$ as $N \to \infty$, the parameter $a_N$ reflects the tendency of a common belief to decrease with a growing population.

Except for the trivial case $\mu = \delta_0$ the random variables $X_i$ are never independent under $\mathbb{P}_\mu$. This can be seen from the covariance

$$\langle X_i, X_j\rangle_\mu := \mathbb{E}_\mu(X_i X_j) - \mathbb{E}_\mu(X_i)\mathbb{E}_\mu(X_j)\ . \tag{3.6}$$

In (3.6) as well as in the following $\mathbb{E}_\mu$ denotes expectation with respect to $\mathbb{P}_\mu$. In fact, the random variables $X_i$ are always positively correlated:

*Theorem 3.1* For $i \neq j$ we have

$$\langle X_i, X_j\rangle_\mu = \int \zeta^2\, d\mu(\zeta)\ . \tag{3.7}$$

The quantity $\int \zeta^2\, d\mu(\zeta)$ is called the second moment of the measure $\mu$. Since the first moment $\int \zeta\, d\mu(\zeta)$ vanishes due to (3.4) the second moment equals the variance of $\mu$. Observe that $\int \zeta^2 d\mu(\zeta) = 0$ implies $\mu = \delta_0$. For independent random variables $\langle X_i, X_j\rangle_\mu = 0$, so (3.7) implies that $X_i, X_j$ depend on each other unless $\mu = \delta_0$.

To investigate the impact of the collective bias measure $\mu$ on the ideal weight in a heterogeneous voting model we have to compute the quantity

$$\mathbb{E}_\mu(|\sum_{i=1}^{N} X_i|) \tag{3.8}$$

for a measure $\mu$ and population $N$ (at least for large $N$). This is done with the help of the following theorem:

*Theorem 3.2* We have:

$$\Big|\, \mathbb{E}_\mu(\frac{1}{N}\,|\sum_{i=1}^{N} X_i\,|) - \int\, |\zeta|\, d\mu(\zeta)\Big| \leq \frac{1}{\sqrt{N}} \tag{3.9}$$

Let us define $\overline{\mu} = \int |\zeta|\, d\mu(\zeta)$. If we choose $\mu \neq \delta_0$ independent of the (population of the) state Theorem 3.2 implies that the optimal weight in the council is *proportional* to $N$ (rather than $\sqrt{N}$). This is true in particular for the original Straffin model (Straffin 1982) where $\mu_n \equiv \frac{1}{2}\chi_{[-1,1]}(z)\, dz$ which corresponds to the Shapley-Shubik power index (see section .1). We have:

*Theorem 3.3* If the collective bias measure $\mu \neq \delta_0$ is independent of $N$ then the optimal weight in the council is given by:

$$w_N \;=\; \mathbb{E}_\mu(\sum_{i=1}^{N}\,|X_i|) \;\sim\; \overline{\mu}\, N \tag{3.10}$$

If $\mu = \mu_N$ depends on the population then

$$\mathbb{E}_{\mu_N}(|\sum_{i=1}^{N} X_i|) \sim\; N\,\overline{\mu}_N$$

as long as $\overline{\mu}_N \geq \frac{1}{N^{1/2-\varepsilon}}$ for some $\varepsilon > 0$. However, if $\overline{\mu}_N \leq \frac{1}{N^{1/2+\varepsilon}}$, then

$$E_{\mu_N}(|\sum_{i=1}^{N} X_i|) \sim \sqrt{N}\,.$$

Hence, in this case we rediscover a square-root law.

We summarize:

*Theorem 3.4* Let us suppose that a state with a population of size $N$ is characterized by a collective bias measure $\mu_N$, then:

1. If

$$\overline{\mu}_N = \int |\zeta| \, d\mu_N(\zeta) \geq C \frac{1}{N^{1/2-\varepsilon}} \tag{3.11}$$

for some $\varepsilon > 0$ and for all large $N$ then the optimal weight $w_N$ is given by:

$$w_N = \mathbb{E}_\mu\left( \Big| \sum_{i=1}^{N} X_i \Big| \right) \sim N \, \overline{\mu}_N. \tag{3.12}$$

2. If

$$\overline{\mu}_N = \int |\zeta| \, d\mu_N(\zeta) \leq C \frac{1}{N^{1/2+\varepsilon}} \tag{3.13}$$

then for large $N$ the optimal weight $w_N$ is given by:

$$w_N = \mathbb{E}_\mu\left( \Big| \sum_{i=1}^{N} X_i \Big| \right) \sim \sqrt{N}. \tag{3.14}$$

*Example* In our Straffin-type example (3.5) we choose:

$$\mu_N(dz) = \frac{1}{2a_N} \chi_{[-a_N, a_N]}(z) dz, \tag{3.15}$$

then:

$$\overline{\mu}_N = \frac{1}{2} a_N. \tag{3.16}$$

Let us assume $a_N \sim C N^{-\alpha}$ for $0 \leq \alpha \leq 1$. Then, if $\alpha > \frac{1}{2}$ we have $w_N \sim \sqrt{N}$ and if $\alpha < \frac{1}{2}$ we obtain $w_N \sim C N^{1-\alpha}$.

*Remarks 3.5*

1. Our result shows that in all cases the optimal weight $w_N$ satisfies $C\sqrt{N} \leq w_N \leq N$. It is a matter of empirical studies to determine which measure $\mu_N$ is appropriate to the given voting system. Any of the empirical results of (Gelman et al. 2004) can be modeled by an appropriate choice of $\mu_N$.

2. It is only $\overline{\mu}_N$ that enters the formulae (3.12) and (3.14), no other information about $\mu_N$ is relevant. The quantities $\overline{\mu}_N$ can be estimated using Theorem 3.2. In fact, more is true by the following result.

*Theorem 3.6* Let $P_N$ be the distribution of $\frac{1}{N}\sum_{i=1}^{N} X_i$ under the measure $\mathbb{P}_{\mu_N}$ then the sequence of measures $P_N - \mu_N$ converges weakly to 0.

Theorem 3.6 tells us that the collective bias measure can be recovered from voting results. Let us denote by $S$ a voting result, i. e. $S = \frac{1}{N}\sum_{i=1}^{N} X_i$. In other words, $\frac{1}{2}(S+1)$ is the fraction of affirmative votes. Theorem 3.6 tells us that the probability distribution of $S$ approximates the measure $\mu_N$ for large $N$. On the other hand the distribution of $S$ can be estimated from independent voting samples (defining the empirical distribution).

Note that the empirical distribution of voting results $\frac{1}{N}\sum_{i=1}^{N} X_i$ is the quantity considered in (Gelman et al. 2004). Theorem 3.6 tells us that the distribution of the voting results for large number $N$ of voters is approximately equal to the distribution $\mu_N$. In particular, in the case of independent voters the voting result is always extremely tight while for Straffin's example any voting result has the same probability, i.e. it is equally likely that a proposal gets 99% or 53% of the votes. The general 'collective bias model' defined above is an extension both of the independent voting model and of Straffin's model. This general model can be fit to any distribution of voting results.

## 4. Voting models as spin systems

Spin systems are a central topic in statistical physics. They model magnetic phenomena. The spin variables, usually denoted by $\sigma_i$, may take values in the set $\{-1, +1\}$ with $+1$ and $-1$ meaning 'spin up' and 'spin down' respectively. The spin variables model the elementary magnets of a material (say the electrons or nuclei in a solid). The index $i$ runs over an index set $I$ which represents the set of elementary magnets. We may (and will) take $I = \{1, 2, \dots, N\}$ in the following.

A *spin configuration* is a sequence $\{\sigma_i\}_{i\in\{1,\dots,N\}} \in \{-1, +1\}^N$. A configuration of spins $\{\sigma_i\}_{i\in I}$ has a certain *energy* which depends on the way the spins interact with each other and (possibly) an exterior magnetic field. The energy, a function of the spin configuration, is usually denoted by $\mathcal{E} = \mathcal{E}(\{\sigma_i\}_{i\in I})$. Spin systems prefer configurations with small energy. For example in so called ferromagnetic systems, magnetic materials we encounter in every day's life, the energy of spins pointing in the same direction is smaller than the one for antiparallel spins, hence there is a tendency that spins line up, a fact that leads to the existence of magnetic materials.

The temperature $T$ measures the strength of fluctuations in a spin system. If the temperature $T$ of a system is zero, there are no fluctuations and the spins will stay in the configuration(s) with the smallest energy. However, if the temperature is positive, there is a certain probability that the spin configuration deviates from the one with the smallest energy. The probability to find a spin system at temperature $T > 0$ in a a configuration $\{\sigma_i\}$ is given by:

$$p\left(\{\sigma_i\}\right) \;=\; \mathcal{Z}^{-1}\, e^{-\frac{1}{T}\,\mathcal{E}(\{\sigma_i\})} \tag{4.1}$$

The quantity $\mathcal{Z}$ is merely a normalization constant, to ensure that the right hand side of (4.1) defines a probability (i. e. gives total probability equal to one). Consequently:

$$\mathcal{Z} \;=\; \sum_{\{\sigma_i\}\in\{-1,1\}^N} e^{-\frac{1}{T}\,\mathcal{E}(\{\sigma_i\})} \tag{4.2}$$

A probability distribution as in (4.1) is called a *Gibbs measure*. It is customary to introduce the inverse temperature $\beta = \frac{1}{T}$ and to write (4.1) as:

$$p\left(\{\sigma_i\}\right) \;=\; \mathcal{Z}^{-1}\, e^{-\beta\,\mathcal{E}(\{\sigma_i\})} \tag{4.3}$$

There is a reason to introduce spin systems here: Obviously, any spin system can be interpreted as a voting system, we just interchange the words spin configuration and voting result as well as the symbols $\sigma_i$ and $X_i$. In fact, a Gibbs measure $p$ (as in (4.3)) defines a voting system with voting measure $p$, as long as $\mathcal{E}(\{\sigma_i\}) = \mathcal{E}(\{-\sigma_i\})$, so that $p$ satisfies the symmetry condition (2.2). Moreover, any voting measure can be obtained from a Gibbs measure.

In particular, independent voting corresponds to the energy functional $\mathcal{E}(\{\sigma_i\}) \equiv 1$. In this case any configuration has the same energy, so that no configuration is more likely than any other.

The 'collective bias' model is given by an energy function:

$$\mathcal{E}(\{\sigma_i\}) \;=\; -\,h\sum_i \sigma_i \tag{4.4}$$

where $h$ is a random variable connected to the collective bias variable $Z$ by:

$$\frac{1}{2}\,(1+Z) = \frac{e^h}{e^h + e^{-h}}. \tag{4.5}$$

Note, when $h$ runs from $-\infty$ to $\infty$ in (4.5) the value of $Z$ runs monotonously from $-1$ to $+1$.

In term of statistical physics in this model the spins do not interact with each other, but they do interact with a random but constant exterior field. The inverse temperature $\beta$ is superfluous in this model as it can be absorbed in the magnetic field strength $h$.

## 5. The voters' interaction model

In the collective bias model the voting behavior of each voter is influenced by a preassigned, a priori given collective bias variable $Z$ (by an exterior magnetic field in the spin picture). The correlation between the voters results from the general voting tendency described by the value of $Z$.

In this section we investigate a model with a *direct* interaction between the voters, namely a tendency of the voters to vote in agreement with each other. In the view of statistical physics this corresponds to the tendency of magnets to align. There are various models in statistical physics to prescribe such a situation. Presumably the best known one is the Ising model where neighboring spins interact in the prescribed ways. The neighborhood structure is most of the time given by a lattice (e.g. $\mathbb{Z}^d$). The results on the system depend strongly on that neighborhood structure, in the case of the lattice $\mathbb{Z}^d$ on the dimension $d$.

In the following we consider another, in fact easier model where no such assumption on the local 'neighborhood' structure has to be made. We consider it an advantage of the model that very little of the microscopic correlation structure of a specific voting system enters into the model.

The model we are going to consider is known in statistical mechanics as the *Curie-Weiss model* or the *mean field model* (see e.g. (Thompson 1972), (Bolthausen and Sznitman 2002) or (Dorlas 1999)). In this model a given voter (spin) interacts with all the other voters (resp. spins) in a way which makes it more likely for the voters (spins) to agree than to disagree. This is expressed through an energy function $\mathcal{E}$ which is smaller if voters agree. Note that a *small* energy for a given voting configuration (relative to the other configurations) leads to a *high* probability of that configuration relative to the others through formula (4.3).

The energy $\mathcal{E}$ for a given voting outcome $\{X_i\}_{i=1\ldots N}$ is given in the mean field model by:

$$\mathcal{E}\big(\{X_i\}\big) = -\frac{J}{N-1}\sum_{\substack{i,j\\ i\neq j}} X_i X_j. \tag{5.1}$$

Here $J$ is a non negative number called the coupling constant. According to (5.1) the energy contribution of a single voter $X_i$ is expressed through the averaged voting result of all other voters $\frac{1}{N-1}\sum_{j\neq i} X_j$. If $X_i$ agrees in sign with this average the voter $i$ makes a negative contribution to the total energy, otherwise $X_i$ will increase the total energy. The strength of this negative or positive contribution is governed by the coupling constant $J$. In other words: Situations for which voter $i$ agrees with the other voters in average are more likely than others. This can be seen from the formula for the probability of a given voting outcome, namely:

$$p\left(\{X_i\}\right) = \mathcal{Z}^{-1}\, e^{-\beta\,\mathcal{E}(\{X_i\})} = \mathcal{Z}^{-1}\, e^{\,\beta J \frac{1}{N-1}\sum_{i\neq j} X_i X_j} \tag{5.2}$$

where as before

$$\mathcal{Z} = \sum_{\{X_i\}\in\{\pm 1\}^N} e^{-\beta\,\mathcal{E}(\{X_i\})}. \tag{5.3}$$

Since the probability density $p$ depends only on the product of $\beta$ and $J$ we may absorb the parameter $J$ into the inverse temperature $\beta$. So without loss of generality we can set $J = 1$. We denote the probability density (5.2) by $p_{\beta,N}$ and the corresponding expectation by $\mathbb{E}_{\beta,N}$.

Our goal is to compute the average:

$$w_N \;=\; \mathbb{E}_{\beta,N}\Big(\,\big|\sum_{i=1}^{N} X_i\,\big|\Big). \tag{5.4}$$

The quantity $w_N$ gives the optimal weight in the council for a population of $N$ voters with a correlation structure given by a mean-field model with inverse temperature $\beta$. We will see that the value of $w_N$ changes dramatically when $\beta$ changes from a value below one to a value above one. This has to do with the fact that the mean-field model undergoes a phase transition at the inverse temperature $\beta = 1$ (see (Bolthausen and Sznitman 2002; Dorlas 1999; Thompson 1972)).

*Theorem 5.1*

1. If $\beta < 1$ then

$$w_N \;=\; \mathbb{E}_{\beta,N}\Big(\,\big|\sum_{i=1}^{N} X_i\,\big|\Big) \;\sim\; \frac{\sqrt{2}}{\sqrt{\pi}}\,\frac{1}{\sqrt{1-\beta}}\,\sqrt{N} \qquad \text{as } N \to \infty. \tag{5.5}$$

2. If $\beta > 1$ then

$$w_N \;=\; \mathbb{E}_{\beta,N}\Big(\,\big|\sum_{i=1}^{N} X_i\,\big|\Big) \;\sim\; C(\beta)\; N \qquad \text{as } N \to \infty. \tag{5.6}$$

*Remarks 5.2*

1. By $x_N \sim y_N$ as $N \to \infty$ we mean that $\lim_{n\to\infty} \frac{x_N}{y_N} = 1$.

2. The constant $C(\beta)$ in (5.6) can be computed: If $\beta > 1$ then $C(\beta)$ is the (unique) positive solution $C$ of

$$\tanh(\beta\, C) = C. \tag{5.7}$$

Note that for $\beta \leq 1$ there is no positive solution of equation (5.7).

Theorem 5.1 can be understood quite easily on an intuitive level. We recall that the temperature $T$ measures the strength of fluctuations, in other words: Low temperature (= large $\beta = \frac{1}{T}$) means high order in the system, high temperature (= small $\beta$) means disorder. Hence, the theorem says, that for strong order the expected voting result is well above (resp. well below) 50 % and the ideal weight is proportional to the size of the population, while for highly fluctuating societies polls are as a rule very tight and one obtains a square root law for the ideal representation.

The proof of Theorem 5.1 will be given in section .4.

## 6. Conclusions

The above calculations show that one can reproduce the square-root law as well as the results of (Gelman et al. 2004) and other laws by assuming particular correlation structures among the voters of a certain country. To find the right model is a question of adjusting the parameters of the models to empirical data of the country under consideration. Moreover, the models allow us to investigate questions about voting systems on a theoretical level. We believe that the models described above can help to understand voting behavior in many situations.

To *design* a nonhomogeneous voting system for a *constitution* in the light of our results is a question of different nature. Even knowing the correlation structure of the countries in question exactly would be of limited value to design a constitution. Constitutions are meant for a long term period, correlation structures of countries on the other hand are changing even on the scale of a few years.

One might argue that modern societies have a tendency to decrease the correlation between their members. In all modern states, at least in the West, the influence of churches, parties, and unions is constantly declining.

In addition to this it seems more important to protect small countries against a domination of the big ones than the other way round. This motivates us to choose a square-root law in these long term cases.

## Appendix

### A *Power indices and Straffin's model*

Here we investigate some connection of our models with power indices. Power indices are usually defined through the ability of voters to change the voting result by their vote. To define power indices so we have to introduce a general setup for voting systems. This extends the considerations of the rest of this paper where we considered only weighted voting. Our presentation below is inspired by (Laruelle and Valenciano 2005) and (Straffin 1982).

Let $\mathcal{V} = \{1, \ldots, N\}$ be the set of voters. The (microscopic) *voting outcome* is a vector $X = (X_1, \ldots, X_N) \in \{-1, +1\}^N$. Of course, $X_i = 1$ means that the voter $i$ approves the proposal under consideration, while $X_i = -1$ means $i$ rejects the proposal. We call $\Omega = \{-1, +1\}^N$ together with a probability measure $\mathbb{P}$ a *voting space*, if $\mathbb{P}$ is invariant under the transformation $T : \omega \mapsto -\omega$, thus $\mathbb{P}(\{X\}) = \mathbb{P}(\{-X\})$ (see section 2 for a discussion of this property).

A *voting rule* is a function $\phi : \{-1,+1\}^N \longrightarrow \{-1,+1\}$. The voting rule associates to a microscopic voting outcome $X = (X_1,\ldots,X_N)$ a macroscopic voting result, i. e. the decision of the assembly. Thus, $\phi(X) = 1$ (resp. $\phi(X) = -1$) means that the proposal is approved (resp. rejected) by the assembly $\mathcal{V}$ if the microscopic voting outcome is $X = (X_1,\ldots,X_N)$. We always assume that the voting rule $\phi$ is *monotone*: If $X_i \le Y_i$ for all $i$ then $\phi(X) \le \phi(Y)$. We also suppose that $\phi(-1,\ldots,-1) = -1$ and $\phi(+1,\ldots,+1) = +1$.

Following (Laruelle and Valenciano 2005) we say that a voter $i$ is *successful* for a voting outcome $X$ if $\phi(X) = X_i$. Let us set $(X_1,\ldots,X_N)^{i,-} = (X_1,\ldots,X_{i-1},-X_i,X_{i+1},\ldots,X_N)$. We call a voter $i$ *decisive* for $X$ if $\phi(X) \neq \phi(X^{i,-})$, i. e. if the voting result changes if $i$ changes his/her mind.

Given a voting space $(\{-1,1\}^N,\mathbb{P})$ and a voting rule $\phi$ we define the ($\mathbb{P}$-)power index $\beta$ by:

$$\beta(i) = \beta_{\mathbb{P}}(i) = \mathbb{P}\{X \in \{-1,+1\}^N \mid i \text{ is decisive for } X\} \tag{A.1}$$

Laruelle and Valenciano (Laruelle and Valenciano 2005) show that many known power indices are examples of the general concept (A.1). For example it is not difficult to show that one obtains the total Banzhaf index (see (Banzhaf 1965) or (Taylor 1995)) if $\mathbb{P}$ is the probability measure of independent voting.

If we take $\mathbb{P} = \mathbb{P}_\mu$ to be the voting measure corresponding to the collective bias measure $\mu$ we get a whole family of power indices from (A.1). Straffin (Straffin 1982) (see also (Paterson Preprint)) demonstrates that if $\mu$ is the uniform distribution on $[-1,1]$ then $\beta_{P_\mu}$ is just the Shapley-Shubik index (see (Shapley and Shubik 1954) or (Taylor 1995)).

In a subsequent publication on general power indices we will give a derivation of this fact in the current framework.

## B *Proofs for section 2*

We start with a short Lemma:

*Lemma B.1* Suppose $X_1,\ldots,X_N$ are $\{-1,1\}$–valued random variables with the symmetry property (2.2) then

$$\mathbb{E}(\sum_{i=1}^{N} X_i) = 0 \tag{B.1}$$

and

$$\mathbb{E}\Big( \sum_{i=1}^{N} X_i \ \chi(\sum_{i=1}^{N} X_i) \Big) = \mathbb{E}\big( | \sum_{i=1}^{N} X_i | \big) \,. \tag{B.2}$$

*Remark B.2* As defined above $\chi(x) = 1$ if $x > 0$, $\chi(x) = -1$ if $x \leq 0$.

*Proof* (2.2) implies

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$$

hence $\mathbb{E}(X_i) = 0$ and (B.1) follows.

To prove (B.2) we observe that

$$\mathbb{E}\Bigg( | \sum_{1}^{N} X_i | \Bigg) \ = \ \mathbb{E}\Bigg( \sum_{i=1}^{N} X_i ; \sum_{i=1}^{N} X_i > 0 \Bigg) - \mathbb{E}\Bigg( \sum_{i=1}^{N} X_i ; \sum_{i=1}^{N} X_i < 0 \Bigg)$$

$$= \mathbb{E}\Bigg( \sum_{i=1}^{N} X_i \ \chi(\sum_{i=1}^{N} X_i) \Bigg) . \qquad \square$$

We turn to the proof of Theorem 2.1.

*Proof (Theorem 2.1)* Let us abbreviate: $S_\nu := \sum_{i=1}^{N_\nu} X_{\nu i}$.

Observe that the $S_\nu$ are independent by assumption and satisfy $\mathbb{E}(S_\nu) = 0$, moreover

$$\mathbb{E}\Bigg( S_\nu \ \chi(S_\mu) \Bigg) = 0 \text{ if } \nu \neq \mu \tag{B.3}$$

and

$$\mathbb{E}\Bigg( S_\nu \ \chi(S_\nu) \Bigg) = \mathbb{E}\big( |S_\nu| \big) \tag{B.4}$$

by Lemma B.1. To find the minimum of the function

$$\Delta(w_1, \ldots, w_M) = \mathbb{E}\big( (\sum_{1}^{M} S_\nu - \sum_{1}^{M} w_\nu \ \chi(S_\nu))^2 \big)$$

we look at the zeros of $\frac{\partial \Delta}{\partial w_\mu}$.

$$0 = \frac{\partial \Delta}{\partial w_\mu} = -2 \, \mathbb{E}\big( (\sum_{1}^{M} S_\nu - \sum_{1}^{M} w_\nu \ \chi(S_\nu)) \chi(S_\mu) \big)$$

$$= -2\,\mathbb{E}(S_\mu\;\chi(S_\mu) - w_\mu\;\chi(S_\mu)\;\chi(S_\mu))\;. \tag{B.6}$$

So

$$w_\mu\;\mathbb{E}((\chi(S_\mu))^2) = \mathbb{E}(S_\mu\;\chi(S_\mu)) = \mathbb{E}(|S_\mu|)\;.$$

Since $\chi(S_\mu)^2 = 1$ we obtain

$$w_\mu = \mathbb{E}(|S_\mu|)\;.$$

□

We turn to the proof of Theorem 2.2.

*Proof* Let $X_1, \ldots, X_N$ be $\{-1, 1\}$–valued random variables with $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$. Then

$$\mathbb{E}(|\sum_1^N X_i|) = \sqrt{N}\;\mathbb{E}(|\frac{1}{\sqrt{N}}\sum_1^N X_i|)\;.$$

By the central limit theorem (see e.g. (Lamperti 1996)) $\frac{1}{\sqrt{N}}\sum_1^N X_i$ has asymptotically a normal distribution with mean zero and variance 1, hence $\mathbb{E}(|\frac{1}{\sqrt{N}}\sum_1^N X_i|) \to \frac{\sqrt{2}}{\sqrt{\pi}}$. □

## C *Proofs for section 3*

*Proof (Theorem 3.1)* Since $\mathbb{E}_\mu(X_i) = 0$,

$$\langle X_i, X_j\rangle_\mu = \mathbb{E}_\mu(X_i X_j)$$

$$= \mathbb{P}_\mu(X_i = X_j = 1) + \mathbb{P}_\mu(X_i = X_j = -1) - 2\mathbb{P}_\mu(X_i = 1, X_j = -1)$$

$$= \int d\mu(\zeta) \quad \{P_{\frac{1}{2}(1+\zeta)}(X_i = X_j = 1) + P_{\frac{1}{2}(1+\zeta)}(X_i = X_j = -1)$$
$$-2P_{\frac{1}{2}(1+\zeta)}(X_i = 1, X_j = -1)\}$$

$$= \int d\mu(\zeta) \quad \{\frac{1}{4}(1+\zeta)^2 + \frac{1}{4}(1-\zeta)^2 - \frac{1}{2}(1-\zeta^2)\}$$

$$= \int \zeta^2 d\mu(\zeta)\;. \tag{C.1}$$

□

To prove Theorem 3.2 we need the following Lemma:

*Lemma C.1* $\mathbb{E}_\mu(\frac{1}{N}|\sum(X_i - Z)|) \le \frac{1}{\sqrt{N}}$.

*Proof*

$$\mathbb{E}_\mu\Big(\frac{1}{N}\big|\sum(X_i - Z)\big|\Big) = \frac{1}{N}\,\mathbb{E}_\mu\Big(\big|\sum(X_i - Z)\big|\Big)$$

$$\leq \frac{1}{N}\Big\{\,\mathbb{E}_\mu\Big(\big(\sum(X_i - Z)\big)^2\Big)\Big\}^{1/2}$$

$$= \frac{1}{N}\Big\{\int d\mu(\zeta)\; E_{p_\zeta}\Big(\big(\sum_1^N(X_i - \zeta)\big)^2\Big)\Big\}^{1/2}. \tag{C.2}$$

Given $Z = \zeta$ the random variables $X_i - \zeta$ have mean zero and are independent with respect to the measure $P_{p_\zeta}$, thus

$$E_{p_\zeta}\Big(\big(\sum_1^N(X_i - \zeta)\big)^2\Big) = N E_{p_\zeta}(X_i - \zeta)^2 = N(1-\zeta^2) \leq N\,,$$

hence

$$(C.2) \leq \frac{1}{\sqrt{N}}\Big(\int d\mu(\zeta)(1-\zeta^2)\Big)^{1/2} \leq \frac{1}{\sqrt{N}}\,. \qquad \square$$

Using Lemma C.1 we are in a position to prove Theorem 3.2:

*Proof* (1) Suppose that:

$$\overline{\mu}_N = \int |\zeta|\; d\mu_N(\zeta) \geq C\,\frac{1}{N^{1/2-\varepsilon}} \tag{C.3}$$

then we estimate:

$$\mathbb{E}_{\mu_N}\Big(\frac{1}{N}\big|\sum_1^N X_i\big|\Big) = \mathbb{E}_{\mu_N}\Big(\big|\frac{1}{N}\sum_1^N(X_i - Z) + Z\big|\Big) \tag{C.4}$$

$$\leq \mathbb{E}_{\mu_N}(|Z|) + \mathbb{E}_{\mu_N}\Big(\big|\frac{1}{N}\sum_1^N(X_i - Z)\big|\Big) \leq \overline{\mu}_N + \frac{1}{\sqrt{N}} \tag{C.5}$$

by Lemma C.1. Moreover

$$\mathbb{E}_{\mu_N}\Big(\frac{1}{N}\big|\sum_1^N X_i\big|\Big) \geq \mathbb{E}_{\mu_N}(|Z|) - \mathbb{E}_{\mu_N}\Big(\big|\frac{1}{N}\sum X_i - Z\big|\Big) \quad \geq \overline{\mu}_N - \frac{1}{\sqrt{N}}\,.$$

Hence

$$\Big|\,\mathbb{E}_{\mu_N}\Big(\frac{1}{N}\big|\sum_1^N X_i\big|\Big) - \overline{\mu}_N\,\Big| \leq \frac{1}{\sqrt{N}} \tag{C.7}$$

which proves (3.12).

(2) To prove (3.14) we obtain by the same reasoning as above:

$$|\,\mathbb{E}_{\mu_N}(\frac{1}{N}|\sum_{1}^{N}X_i|) \;-\; \overline{\mu}_N\,| \;\leq\; \frac{1}{\sqrt{N}} \tag{C.8}$$

□

We end this section with the proof of Theorem 3.6:

*Proof* We have to prove that for bounded continuous functions $f$:

$$\int \Big( f(\frac{1}{N}\sum_{i=1}^{N} X_i) - f(Z) \Big)\, d\,\mathbb{P}_{\mu_N} \;\rightarrow\; 0. \tag{C.9}$$

The convergence (C.9) is clear for continuously differentiable $f$ from Lemma C.1. It follows for arbitrary bounded continuous $f$ by a density argument. □

## D *Proofs for section 5*

In this section we prove Theorem 5.1.

*Proof (Theorem 5.1 (1))* We denote by $E_0^{(N)}$ the expectation of the coin tossing model for $N$ independent symmetric $\{+1,-1\}$-valued random variables, i.e.:

$$E_0^{(N)}\big(F(X_1,\ldots,X_N)\big) \;=\; \frac{1}{2^N}\sum_{\{x_i\}\in\{+1,-1\}^N} f(x_1,\ldots,x_N). \tag{D.1}$$

We set:

$$\mathcal{Z}_{\beta N} = E_0^{(N)}\Big( e^{\frac{\beta}{2}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i\right)^2} \Big) \tag{D.2}$$

and:

$$\mathcal{X}_{\beta N} \;=\; E_0^{(N)}\Big(\,|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i|\; e^{\frac{\beta}{2}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i\right)^2}\Big). \tag{D.3}$$

Then:

$$\mathbb{E}_{\beta N}(|\sum_{i=1}^{N}X_i|) \;=\; \sqrt{N}\;\frac{\mathcal{X}_{\beta,N}}{\mathcal{Z}_{\beta,N}}. \tag{D.4}$$

Under the probability law $E_0^{(N)}$ the random variables $X_i$ are centered and independent, thus the central limit theorem (see e.g. (Lamperti 1996)) tells us that $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i$ converges in distribution to a standard normal distribution. Consequently, for $\beta < 1$ and $N \to \infty$ :

$$\mathcal{Z}_{\beta N} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(1-\beta)\,x^2}{2}}\, dx \;=\; \frac{1}{\sqrt{1-\beta}} \tag{D.5}$$

and:

$$\mathcal{X}_{\beta N} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |x|\, e^{-\frac{(1-\beta)\,x^2}{2}}\, dx \;=\; \frac{\sqrt{2}}{\sqrt{\pi}}\, \frac{1}{1-\beta}. \tag{D.6}$$

Consequently:

$$\mathbb{E}_{\beta N}(|\sum_{i=1}^{N} X_i|) \;=\; \sqrt{N}\, \frac{\mathcal{X}_{\beta,N}}{\mathcal{Z}_{\beta,N}} \;\sim\; \frac{\sqrt{2}}{\sqrt{\pi}}\, \frac{1}{\sqrt{1-\beta}}\, \sqrt{N}. \tag{D.7}$$

□

*Proof (Theorem 5.1 (2))* By Theorem 6.3 in (Bolthausen and Sznitman 2002) the distribution $\nu_N$ of $S_N = \frac{1}{N} \sum_{i=1}^{N} X_i$ converges weakly to the measure $\nu = \delta_{-C(\beta)} + \delta_{C(\beta)}$ where $C(\beta)$ was defined in (5.7).

Hence,

$$\mathbb{E}_{\beta}(|\sum_{i=1}^{N} X_i|) = N\, \mathbb{E}_{\beta}(|S_N|) \;=\; N \int |\lambda|\, d\nu_N(\lambda)$$

$$\approx N \int |\lambda|\, d\nu(\lambda) \;=\; N\, C(\beta). \tag{D.8}$$

□

## Acknowledgements

It is a pleasure to thank Hans-Jürgen Sommers, Duisburg-Essen, and Wojciech Słomczyński and Karol Życzkowski, Krakow, for valuable discussions. The author would also like to thank Helmut Hölzler, Göttingen, for pointing out a notational inconsistency in a previous version of the paper.

## References

Banzhaf, J. (1965). Weighted Voting Doesn't Work: A Mathematical Analysis. *Rutgers Law Review* 19: 317–343.

Bolthausen, E. and Sznitman, A. (2002). *Ten lectures on random media*.

Dorlas, T. (1999). *Statistical Mechanics*. Institute of Physics Publishing.

Felsenthal, D. and Machover, M. (1998). *The measurement of power: theory and practice, problems and paradoxes*. Edward Elgar.

Gelman, A., Katz, J. and Bafumi, J. (2004). Standard voting power indexes don't work: an empirical analysis. *British Journal of Political Science* 34: 657–674.

Kirsch, W. (Preprint). On the distribution of power in the Council of ministers of the EU. Available from http://www.ruhr-uni-bochum.de/mathphys/.

Lamperti, J. (1996). *Probability*. Wiley.

Laruelle, A. and Valenciano, F. (2005). Assessing success and decisiveness in voting situations. *Social Choice and Welfare* 24: 171–197.

Paterson, I. (Preprint). A Lesser Known Probabilistic Approach to the Schapley-Shubik Index and Useful Related Voting Measures.

Shapley, L. and Shubik, M. (1954). A method for evaluating the distribution of power in a commitee system. *American Political Science Review* 48: 787–792.

Straffin, P. (1982). Power indices in politics. In Brams, S. J., Lucas, W. F. and Straffin, P. D. (eds), *Political and related models*. Springer.

Taylor, A. (1995). *Mathematics and Politics*. Springer.

Thompson, C. (1972). *Mathematical Statistical Mechanics*. Princeton University Press.