Reconstruction of the Probability Measure and the Coupling Parameters in a Curie-Weiss Model

Miguel Ballesteros*, Ramsés H. Mena*, Arno Siri-Jégousse*, and Gabor Toth*

Abstract

The Curie-Weiss model is used to study phase transitions in statistical mechanics and has been the object of rigorous analysis in mathematical physics. We analyse the problem of reconstructing the probability measure of a multi-group Curie-Weiss model from a sample of data by employing the maximum likelihood estimator for the coupling parameters of the model, under the assumption that there is interaction within each group but not across group boundaries. The estimator has a number of positive properties, such as consistency, asymptotic normality, and exponentially decaying probabilities of large deviations of the estimator with respect to the true parameter value. A shortcoming in practice is the necessity to calculate the partition function of the Curie-Weiss model, which scales exponentially with respect to the population size. There are a number of applications of the estimator in political science, sociology, and automated voting, centred on the idea of identifying the degree of social cohesion in a population. In these applications, the coupling parameter is a natural way to quantify social cohesion. We treat the estimation of the optimal weights in a two-tier voting system, which requires the estimation of the coupling parameter.

MSC 2020: 62F10, 82B20, 60F05, 91B12

Keywords: Curie-Weiss model, Mathematical physics, Statistical mechanics, Gibbs measures, Mathematical analysis, Maximum likelihood estimator, Consistency, Asymptotic normality, Large deviations, Two-tier voting systems, Optimal weights

1 Introduction

Models of ferromagnetism have long served as foundational tools in statistical mechanics, enabling the exploration of collective behaviour in systems with many interacting components. As with many other physics models, they have also attracted the attention of mathematical physicists who have contributed rigorous results which frequently confirm the physicists' intuition. Among the most celebrated is the Ising model, introduced in the early 20th century by Wilhelm Lenz and his student Ernst Ising [15]. In its classical form, the Ising model consists of a lattice of binary spins or magnets, where each spin interacts with its nearest neighbours and aligns in response to an external field and thermal fluctuations. Although the one-dimensional version exhibits no phase transition at finite temperature, higher-dimensional cases, such as the two-dimensional model solved by Onsager [29], exhibit rich phase behaviour and critical phenomena, making the Ising model a central object of study in both physics and applied mathematics.

The Curie-Weiss model, introduced as a mean-field approximation of the behaviour of ferromagnets [33], simplifies the spatial complexity by assuming that each spin interacts equally with all others. This global interaction structure permits a rigorous mathematical analysis of phase transitions and critical behaviour (see

*IIMAS-UNAM, Mexico City, Mexico

[10] for an in-depth treatment of the model). Beyond its physical origins, the Curie-Weiss model has been widely used in other domains, including economics [4], political science [17], and sociology [7]. An extension of the Curie-Weiss model is the specification that instead of having a homogeneous population of interacting agents, there are several identifiable groups which have distinct cultures or attitudes, manifesting in different voting decisions. The first version of a multi-group Curie-Weiss model was introduced in [7]; subsequently, similar models were studied in [3, 26, 27, 19, 20, 23], including the statistical problem of community detection [3, 26, 1].

One particularly compelling application of the Curie-Weiss model lies in the domain of collective decision-making, such as voting. In this context, each spin is interpreted as an individual voter casting a binary 'yes' or 'no' vote. The mutual influence between voters, akin to the ferromagnetic alignment in physical systems, can model peer pressure, shared information, or ideological affinity. This analogy becomes especially fruitful in the analysis of two-tier voting systems, where decisions are made in two stages: individuals vote to determine the outcome of a local group (e.g., a state or district), and these local outcomes are then aggregated at a higher level (e.g., in a council or federal assembly). The central question in this setting concerns the *assignment of optimal voting weights* to the representatives of each group, ensuring fair representation in the face of population imbalances or correlated voting behaviour (see [11] for a treatment of voting power).

This challenge is classically illustrated in institutions like the Council of the European Union or the United Nations Security Council. However, its significance extends far beyond geopolitics. In modern automated systems, such as recommendation engines, online platforms aggregating user preferences, and AI-based decision frameworks, users' preferences often exhibit correlated structures. Here too, the need arises to determine how individual or subgroup preferences should be weighted when computing global outcomes. As such, tools from statistical mechanics, and in particular *statistical estimation methods* applied to the Curie-Weiss model, provide both a conceptual and practical bridge between physical models and real-world decision-making systems.

The present article contributes to this line of inquiry by addressing the statistical estimation of the coupling parameters β_λ in a multi-group Curie-Weiss model, a quantity which modulates the strength of interaction between agents or voters within each of the $M \in \mathbb{N}$ groups indexed by $\lambda \in \mathbb{N}_M$[1]. Estimating β_λ from observed voting data yields critical insights into the degree of collective behaviour or ideological alignment, which can, in turn, inform the derivation of optimal weights in two-tier voting systems. Section 7 of this article explores this application in detail under the assumption that voters interact within each group but not across group boundaries, connecting theoretical results to policy-relevant and business-related settings. Imagine a population which is divided into M groups along cultural or national lines. An example is the European Union with its 27 member states. Each member state sends a representative to the Council of the European Union, where votes take place according to a set of fairly complicated rules. We consider the simpler case of a weighted voting system: each member state has a certain voting weight, and its representative's vote in the council is multiplied by the voting weight. Examples of weighted voting systems are 'one person, one vote' in popular votes or voting weights proportional to the population of each country in a council. By imposing a 'fairness criterion,' we can determine what the weights in the council ought to be. The theoretical optimal weights may depend on the underlying voting model (such as the Curie-Weiss model treated in this article) which describes the population's voting behaviour in probabilistic terms. If we reconstruct the underlying voting model, we also obtain estimators for the optimal voting weights, which can then be calculated from a sample of observations.

Prior work has been done on the estimation problem of parameters of spin models such as those we mentioned. A classical reference is the book [16]. The estimation of the interaction parameter from data belonging the

[1] We will write $\mathbb{N}_m := \{1, \dots, m\}$ for any $m \in \mathbb{N}$

realm of the social sciences was studied in [13]. Work has also been done on the estimation of parameters in more complicated models referred to as spin glass models with random interaction structure (see [5, 6]).

Our methodology is grounded in *maximum likelihood estimation*, which is a natural and widely used approach in parametric inference, provided we have a good reason to assume we know the underlying statistical model which generated the data. The maximum likelihood estimator for the coupling parameter is particularly attractive because it is *well-defined for all values of* β_λ, including for parameter values close to the critical regime (and typical samples from this distribution). Furthermore, the maximum likelihood estimator enjoys desirable statistical properties: it is *consistent*, meaning it converges in probability to the true parameter value as the number of observations increases. It is *asymptotically normal*, allowing for the construction of confidence intervals and hypothesis tests. The estimator satisfies *large deviation principles*, providing robust upper bounds on the probability of estimation errors in finite samples.

The main computational drawback of this approach is the necessity to compute the normalisation constant (called partition function in statistical physics) of the Curie-Weiss model, which scales exponentially with the number of voters. This challenge is well known in the literature and has inspired a range of approximation techniques. The path of further research will lead us to consider maximum likelihood estimators based on an approximation to the maximum likelihood optimality condition (8), which sidesteps the costly calculation of the partition function at the cost of sacrificing the ability to calculate estimators for any possible sample. Another avenue leads to alternate estimators based on observing a sample of the votes from a subset of the population. A third direction we are planning to explore is the generalisation of the Curie-Weiss model to interacting groups of voters. This research programme will hopefully provide a clear picture of how to estimate interaction parameters in voting models, with applicability beyond the Curie-Weiss model.

In summary, this article situates the Curie-Weiss model at the intersection of statistical physics, statistics, and political decision-making. By developing and exhaustively analysing the maximum likelihood estimator for the coupling parameters, we aim to provide a rigorous and interpretable method for quantifying interaction strength in voting populations, with applications ranging from international councils to algorithmic aggregation in digital platforms. Given this applicability to different academic disciplines, we have tried to provide complete and comprehensible proofs that appeal to a wide audience composed of mathematicians, statisticians, physicists, economists, and political scientists. The main results of this article are Propositions 8 and 9, and Theorem 10 (to be found in Section 3). Since the estimator we study (see Definition 7) is defined by an implicit condition (8), we require Proposition 8 to be certain the estimator is uniquely determined for any sample of observations. As we will show, there are realisations of the sample which lead to estimates which do not correspond to our assumption that the true parameters are non-negative real numbers. Proposition 9 assures us that the probability of these realisations decays exponentially to 0 as the sample size goes to infinity. Finally, Theorem 10 contains the aforementioned statistical properties of the estimator: consistency, asymptotic normality, and a large deviation principle.

We present the Curie-Weiss model in the next section. The maximum likelihood estimator is defined in Section 3, where we also state the main results of this article. Sections 4 and 5 contain the proofs of the main results. Section 6 is about the standard error of the statistics in this article. The topic of optimal weights in two-tier voting systems is treated in Section 7. Finally, an Appendix to the article contains some auxiliary results we employ.

2 The Curie-Weiss Model

Let the sets $\mathbb{N}_{N_\lambda}$, $\lambda \in \mathbb{N}_M$, represent M groups of voters. We will denote the space of voting configurations for this population by

$$\Omega_{N_1+\cdots+N_M} := \{-1,1\}^{N_1+\cdots+N_M}.$$

Each individual vote cast $x_{\lambda i} \in \Omega_1$ will be indexed by $\lambda \in \mathbb{N}_M$ denoting the group and $i \in \mathbb{N}_{N_\lambda}$ the identity of the voter in question. We will refer to each $(x_{11}, \ldots, x_{1N_1}, \ldots, x_{M1}, \ldots, x_{MN_M}) \in \Omega_{N_1+\cdots+N_M}$ as a voting configuration, which consists of a complete record of the votes cast by the entire population on a certain issue. We model their behaviour in binary voting situations with the following voting model:

Definition 1. Let $N_\lambda \in \mathbb{N}$ and $\beta_\lambda \in \mathbb{R}$, $\lambda \in \mathbb{N}_M$. We set $\boldsymbol{N} := (N_1, \ldots, N_M)$ and $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_M)$. The *Curie-Weiss model* (CWM) is defined for all voting configurations $(x_{11}, \ldots, x_{1N_1}, \ldots, x_{M1}, \ldots, x_{MN_M}) \in \Omega_{N_1+\cdots+N_M}$ by

$$\mathbb{P}_{\boldsymbol{\beta},\boldsymbol{N}}\left(X_{11} = x_{11}, \ldots, X_{MN_M} = x_{MN_M}\right) := Z_{\boldsymbol{\beta},\boldsymbol{N}}^{-1} \exp\left(\frac{1}{2}\sum_{\lambda=1}^{M} \frac{\beta_\lambda}{N_\lambda}\left(\sum_{i=1}^{N_\lambda} x_{\lambda i}\right)^2\right), \tag{1}$$

where $Z_{\boldsymbol{\beta},\boldsymbol{N}}$ is a normalisation constant called the partition function which depends on $\boldsymbol{\beta}$ and $\boldsymbol{N}$. The constants β_λ are called coupling parameters.

In the physical context of the CWM as a model of ferromagnetism, where $M = 1$, there is a single coupling parameter β which is the inverse temperature. As such, the range of values for β is usually $[0, \infty)$. For technical reasons to do with the range of the statistic employed to calculate the maximum likelihood estimator, we will consider the more general definition given above.

However, as a model of voting, the CWM has non-negative coupling parameters to reflect social cohesion. In the voting context, the coupling parameters β_λ measure the degree of influence the voters in group λ exert over each other, with the influence becoming stronger the larger β_λ is. As we see, the most probable voting configurations are those with unanimous votes in favour of or against the proposal. However, there are only two of these extreme configurations, whereas there is a multitude of low probability configurations with roughly equal numbers of votes for and against. This is the 'conflict between energy and entropy.' Which one of these pseudo forces dominates depends on the magnitude of the coupling parameters.

The partition function $Z_{\boldsymbol{\beta},\boldsymbol{N}}$ is defined by

$$Z_{\boldsymbol{\beta},\boldsymbol{N}} = \sum_{x \in \Omega_{N_1+\cdots+N_M}} \exp\left(\frac{1}{2}\sum_{\lambda=1}^{M} \frac{\beta_\lambda}{N_\lambda}\left(\sum_{i=1}^{N_\lambda} x_{\lambda i}\right)^2\right). \tag{2}$$

We set

$$S_\lambda := \sum_{i=1}^{N_\lambda} X_{\lambda i}, \quad \lambda \in \mathbb{N}_M. \tag{3}$$

The key to understanding the behaviour of the CWM is the random vector

$$(S_1, \ldots, S_M),$$

which represents the voting margins, i.e. the difference between the numbers of yes and no votes, in each group.

Notation 2. Throughout this article, we will use the symbol $\mathbb{E}X$ for the expectation and $\mathbb{V}X$ for the variance of some random variable X. Capital letters such as X will denote random variables, while lower case letters such as x will denote realisations of the corresponding random variable.

3 Maximum Likelihood Estimation of $\boldsymbol{\beta}$

3.1 The Maximum Likelihood Estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}$

We will denote by $n \in \mathbb{N}$ the size of a sample of observations. Then each sample takes values in the space

$$\Omega^n_{N_1+\cdots+N_M} := \prod_{i=1}^{n} \Omega_{N_1+\cdots+N_M}.$$

We assume that we have access to a sample of voting configurations $\left(x^{(1)}, \ldots, x^{(n)}\right) \in \Omega^n_{N_1+\cdots+N_M}$ composed of $n \in \mathbb{N}$ i.i.d. realisations of $(X_{11}, \ldots, X_{MN_\lambda})$ from the CWM. The density function for such a sample is given by the n-fold product of (1):

$$f\left(x^{(1)}, \ldots, x^{(n)}; \boldsymbol{\beta}\right) := Z^{-n}_{\boldsymbol{\beta},\boldsymbol{N}} \prod_{t=1}^{n} \exp\left(\frac{1}{2}\sum_{\lambda=1}^{M} \frac{\beta_\lambda}{N_\lambda}\left(\sum_{i=1}^{N_\lambda} x^{(t)}_{\lambda i}\right)^2\right). \tag{4}$$

For fixed $\left(x^{(1)}, \ldots, x^{(n)}\right)$, the function $\boldsymbol{\beta} \in \mathbb{R}^M \mapsto f\left(x^{(1)}, \ldots, x^{(n)}; \boldsymbol{\beta}\right) \in \mathbb{R}$ is called the likelihood function, and

$$\ln f\left(x^{(1)}, \ldots, x^{(n)}; \boldsymbol{\beta}\right) = -n \ln Z_{\boldsymbol{\beta},\boldsymbol{N}} + \frac{1}{2}\sum_{t=1}^{n}\sum_{\lambda=1}^{M} \frac{\beta_\lambda}{N_\lambda}\left(\sum_{i=1}^{N_\lambda} x^{(t)}_{\lambda i}\right)^2 \tag{5}$$

is the log-likelihood function.

The maximum likelihood estimator $\hat{\boldsymbol{\beta}}_{ML}$ of $\boldsymbol{\beta}$ given the sample $\left(x^{(1)}, \ldots, x^{(n)}\right)$ is the value which maximises the likelihood function, i.e.

$$\hat{\boldsymbol{\beta}}_{ML} := \arg\max_{\beta'} f\left(x^{(1)}, \ldots, x^{(n)}; \boldsymbol{\beta}'\right).$$

Since $x \mapsto \ln x$ is a strictly increasing function, we can instead identify $\hat{\boldsymbol{\beta}}_{ML}$ as the value which maximises the log-likelihood function

$$\hat{\boldsymbol{\beta}}_{ML} = \arg\max_{\beta'} \ln f\left(x^{(1)}, \ldots, x^{(n)}; \boldsymbol{\beta}'\right).$$

To find the maximum of the log-likelihood function, we derive with respect to each β_λ and equate to 0:

$$\frac{\mathrm{d} \ln f\left(x^{(1)}, \ldots, x^{(n)}; \boldsymbol{\beta}\right)}{\mathrm{d}\beta_\lambda} = -\frac{n}{Z_{\boldsymbol{\beta},\boldsymbol{N}}}\frac{\mathrm{d}Z_{\boldsymbol{\beta},\boldsymbol{N}}}{\mathrm{d}\beta_\lambda} + \frac{1}{2N_\lambda}\sum_{t=1}^{n}\left(\sum_{i=1}^{N_\lambda} x^{(t)}_{\lambda i}\right)^2 \stackrel{!}{=} 0. \tag{6}$$

We continue with our calculation of the maximum likelihood estimator. The squared sums S^2_λ defined in (3) have expectation

$$\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{N}} S^2_\lambda = \frac{\mathrm{d}Z_{\boldsymbol{\beta},\boldsymbol{N}}}{\mathrm{d}\beta_\lambda} \cdot 2N_\lambda\, Z^{-1}_{\boldsymbol{\beta},\boldsymbol{N}} \tag{7}$$

by Lemma 51. We substitute (7) into (6):

$$\frac{\mathrm{d} \ln f\left(x^{(1)}, \ldots, x^{(n)}; \boldsymbol{\beta}\right)}{\mathrm{d}\beta_\lambda} = -\frac{n}{Z_{\boldsymbol{\beta},\boldsymbol{N}}}\frac{1}{2N_\lambda} Z_{\boldsymbol{\beta},\boldsymbol{N}} \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{N}} S^2_\lambda + \frac{1}{2N_\lambda}\sum_{t=1}^{n}\left(\sum_{i=1}^{N} x^{(t)}_{i}\right)^2.$$

Then the optimality condition $\left.\frac{\mathrm{d}\ln f\left(x^{(1)},\dots,x^{(n)};\boldsymbol{\beta}\right)}{\mathrm{d}\beta}\right|_{\beta=\hat{\beta}_{ML}} = 0$ is equivalent to

$$\mathbb{E}_{\hat{\boldsymbol{\beta}}_{ML},N} S_\lambda^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\sum_{i=1}^{N} x_i^{(t)}\right)^2. \tag{8}$$

Definition 3. We define the statistic $\boldsymbol{T} : \Omega^n_{N_1+\cdots+N_M} \to \mathbb{R}$ for any realisation of the sample $x = \left(x^{(1)},\dots,x^{(n)}\right) \in \Omega^n_{N_1+\cdots+N_M}$ by

$$\boldsymbol{T}(x) \coloneqq \frac{1}{n}\sum_{t=1}^{n}\left(\left(\sum_{i=1}^{N_1} x_{1i}^{(t)}\right)^2,\dots,\left(\sum_{i=1}^{N_M} x_{Mi}^{(t)}\right)^2\right).$$

Remark 4. $\boldsymbol{T}$ is a random vector on the probability space $\Omega^n_{N_1+\cdots+N_M}$ with the power set of $\Omega^n_{N_1+\cdots+N_M}$ as the σ-algebra and the probability measure defined by the density function (4). We will write $\boldsymbol{T}$ for this random variable and $\boldsymbol{T}(x)$ for its realisation given a sample $x \in \Omega^n_{N_1+\cdots+N_M}$.

Proposition 5. *$\boldsymbol{T}$ is a sufficient statistic for $\boldsymbol{\beta}$.*

We will prove this proposition in Section 4.

Notation 6. We will write $[-\infty,\infty]$ for the compactification $\mathbb{R}\cup\{-\infty,\infty\}$ and $[0,\infty]$ for $[0,\infty)\cup\{\infty\}$.

Definition 7 (Maximum Likelihood Estimator)**.** Let $\boldsymbol{N} \in \mathbb{N}^M$ and $\boldsymbol{\beta} \in \mathbb{R}^M$. The *maximum likelihood estimator* of the parameter $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}}_{\boldsymbol{N}} : \Omega^n_{N_1+\cdots+N_M} \to [-\infty,\infty]^M$ such that the optimality condition (8) holds:

$$\mathbb{E}_{\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}(x),\boldsymbol{N}}\left(S_1^2,\dots,S_M^2\right) = \boldsymbol{T}(x), \quad x \in \Omega^n_{N_1+\cdots+N_M}.$$

3.2 Main Results of the Article

In the remainder of this section, we will state the main results concerning the estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}$. Symbols such as '$\xrightarrow[n\to\infty]{\mathrm{p}}$' and '$\mathcal{N}\left(0,\sigma^2\right)$' are used in a standard way (cf. Notation 29).

Proposition 8. *Let $\boldsymbol{N} \in \mathbb{N}^M$ and $n \in \mathbb{N}$. For each sample $x \in \Omega^n_{N_1+\cdots+N_M}$, there is a unique value $y \in [-\infty,\infty]^M$ such that $\mathbb{E}_{y,\boldsymbol{N}}\left(S_1^2,\dots,S_M^2\right) = \boldsymbol{T}(x)$ holds.*

Therefore, the estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}$ is uniquely determined for any realisation $x \in \Omega^n_{N_1+\cdots+N_M}$ of the sample.

Proposition 9. *Let $\boldsymbol{N} \in \mathbb{N}^M$. For each value of the coupling constants $\boldsymbol{\beta} > 0$ [2], there is a constant $\bar{\delta} > 0$ such that*

$$\mathbb{P}\left\{\hat{\boldsymbol{\beta}}_{\boldsymbol{N}} \notin [0,\infty)\right\} \le 2^M \exp\left(-\bar{\delta} n\right)$$

holds for all $n \in \mathbb{N}$.

In (16), we will state the value of the constant $\bar{\delta}$ in the proposition above. Proposition 9 says that although the estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}$ can assume negative and even infinite values, these realisations are exponentially unlikely as the sample size increases given our assumption on the true value of $\boldsymbol{\beta}$ meant to reflect social cohesion.

Finally, we state a theorem about the statistical properties of the estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}$.

Theorem 10. *Fix $\boldsymbol{N} \in \mathbb{N}^M$ and $\boldsymbol{\beta} > 0$. The estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}$ has the following properties:*

1. *$\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}$ is consistent: $\hat{\boldsymbol{\beta}}_{\boldsymbol{N}} \xrightarrow[n\to\infty]{\mathrm{p}} \boldsymbol{\beta}$.*

[2] For all $x \in \mathbb{R}^M$, $x > 0$ is to be interpreted coordinate-wise, i.e. $x_i > 0, i \in \mathbb{N}_M$.

2. $\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}$ *is asymptotically normal:* $\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}-\boldsymbol{\beta}\right)\xrightarrow[n\to\infty]{\mathrm{d}}\mathcal{N}(0,\Sigma)$, *where the covariance matrix* Σ *is diagonal with entries* $\Sigma_{\lambda\lambda}=\frac{4N_\lambda^2}{\mathbb{V}_{\beta_\lambda,N_\lambda}S_\lambda^2},\lambda\in\mathbb{N}_M$.
3. $\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}$ *satisfies a large deviations principle with rate* n *and rate function* $\boldsymbol{J}$ *defined in (19).* $\boldsymbol{J}$ *has a unique minimum at* $\boldsymbol{\beta}$*, and we have for each closed set* $K\subset[-\infty,\infty]^M$ *that does not contain* $\boldsymbol{\beta}$, $\inf_{y\in K}\boldsymbol{J}(y)>0$ *and*
$$\mathbb{P}\left\{\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}\in K\right\}\leq 2^M\exp\left(-n\inf_{y\in K}\boldsymbol{J}(y)\right)$$
for all $n\in\mathbb{N}$.

One could jump to the conclusion that given these results, the estimation problem of the parameter $\boldsymbol{\beta}$ using the maximum likelihood estimator from Definition 7 is solved in satisfactory fashion. However, there are computational problems with the calculation of the expectation $\mathbb{E}_{\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}(x),\boldsymbol{N}}\left(S_1^2,\ldots,S_M^2\right)$ in (8) for all but small $N_1,\ldots,N_M$. The main difficulty is the calculation of the normalisation constant $Z_{\boldsymbol{\beta},\boldsymbol{N}}$ in (2) which is of order $2^{N_1+\cdots+N_M}$ as $N_1+\cdots+N_M\to\infty$. Two possible solutions to this problem are:

1. Find and use an asymptotic approximation of $\mathbb{E}_{\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}(x),\boldsymbol{N}}\left(S_1^2,\ldots,S_M^2\right)$ valid for large $N_1+\cdots+N_M$ which is less costly to calculate than the exact moment $\mathbb{E}_{\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}(x),\boldsymbol{N}}\left(S_1^2,\ldots,S_M^2\right)$.
2. Employ alternate estimators based on small subsets of voters so that instead of the moment $\mathbb{E}_{\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}(x),\boldsymbol{N}}\left(S_1^2,\ldots,S_M^2\right)$ some other expression can be employed which is less costly to calculate. This approach has the added benefit that we need less data to estimate $\boldsymbol{\beta}$. Instead of requiring access to a sample of voting configurations from the entire population, a sample containing observations of a subset of votes suffices.

We will explore both of these avenues in future work.

4 Proof of Propositions 8 and 9

We will analyse the properties of the estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}$, with Propositions 20 and 27 being the key insights for the proof of Propositions 8 and 9. Proposition 8 will follow from Proposition 20, and Proposition 9 from Propositions 20 and 27.

First, we will prove Proposition 5 about the sufficiency of the statistic $\boldsymbol{T}$:

Proof of Proposition 5. We observe that the density function of the sample distribution given in (4) only depends on the observations through the realisation of the statistic $\boldsymbol{T}$:

$$\begin{aligned}f\left(x^{(1)},\ldots,x^{(n)};\boldsymbol{\beta}\right)&=Z_{\boldsymbol{\beta},\boldsymbol{N}}^{-n}\prod_{t=1}^{n}\exp\left(\frac{1}{2}\sum_{\lambda=1}^{M}\frac{\beta_\lambda}{N_\lambda}\left(\sum_{i=1}^{N_\lambda}x_{\lambda i}^{(t)}\right)^2\right)\\&=Z_{\boldsymbol{\beta},\boldsymbol{N}}^{-n}\exp\left(\frac{n}{2}\boldsymbol{T}\left(x^{(1)},\ldots,x^{(n)}\right)\begin{pmatrix}\frac{\beta_1}{N_1}\\\vdots\\\frac{\beta_M}{N_M}\end{pmatrix}\right)=:g\left(T\left(x^{(1)},\ldots,x^{(n)}\right);\boldsymbol{\beta}\right).\end{aligned}$$

□

Due to the product structure of the CWM measure from Definition 1, which features non-interacting groups, we can estimate the coupling constants β_λ independently of each other. We will therefore work with the marginal distributions of each group λ given by

$$\mathbb{P}_{\beta_\lambda, N_\lambda}\left(X_{\lambda 1} = x_{\lambda 1}, \ldots, X_{\lambda N_\lambda} = x_{\lambda N_\lambda}\right) := Z_{\beta_\lambda, N_\lambda}^{-1} \exp\left(\frac{1}{2} \sum_{\lambda=1}^{M} \frac{\beta_\lambda}{N_\lambda} \left(\sum_{i=1}^{N_\lambda} x_{\lambda i}\right)^2\right)$$

for all $(x_{\lambda 1}, \ldots, x_{\lambda N_\lambda}) \in \Omega_{N_\lambda}$, where the partition function is

$$Z_{\beta_\lambda, N_\lambda} = \sum_{x_\lambda \in \Omega_{N_\lambda}} \exp\left(\frac{\beta_\lambda}{2N_\lambda} \left(\sum_{i=1}^{N_\lambda} x_{\lambda i}\right)^2\right).$$

We will return to the multi-group setting in Section 7. In the meantime, we will be working with a single group at a time, and therefore there will be no confusion if we omit the subindex λ from all our expressions to improve readability. As such, we will be writing $\mathbb{P}_{\beta,N}$ instead of $\mathbb{P}_{\beta_\lambda, N_\lambda}$, T instead of $(\boldsymbol{T})_\lambda$, etc.

Definition 11. The *Rademacher distribution* with parameter $t \in [-1, 1]$ is a probability measure P_t on $\{-1, 1\}$ given by $P_t\{1\} := \frac{1+t}{2}$.

We set $u := (1, \ldots, 1) \in \Omega_N$ and for all $x \in \Omega_N$

$$p_\beta(x) := \exp\left(\frac{\beta}{2N} \left(\sum_{i=1}^{N} x_i\right)^2\right). \tag{9}$$

We will use the following auxiliary results in the proof of Proposition 20:

Lemma 12. *The limits*

$$\lim_{\beta \to \infty} \frac{p_\beta(x)}{p_\beta(u)} = \begin{cases} 1 & \text{if } \left|\sum_{i=1}^{N} x_i\right| = N, \\ 0 & \text{otherwise,} \end{cases}$$

hold for all $x \in \Omega_N$.

Proof. First let $\left|\sum_{i=1}^{N} x_i\right| = N$. Then we have

$$\lim_{\beta \to \infty} \frac{p_\beta(x)}{p_\beta(u)} = \lim_{\beta \to \infty} \exp\left[\frac{\beta}{2N}\left(\left(\sum_{i=1}^{N} x_i\right)^2 - \left(\sum_{i=1}^{N} u_i\right)^2\right)\right] = \lim_{\beta \to \infty} \exp\left[\frac{\beta}{2N}\left(N^2 - N^2\right)\right] = 1.$$

Now let $\left|\sum_{i=1}^{N} x_i\right| < N$. Then

$$\lim_{\beta \to \infty} \frac{p_\beta(x)}{p_\beta(u)} = \lim_{\beta \to \infty} \exp\left[\frac{\beta}{2N}\left(\left(\sum_{i=1}^{N} x_i\right)^2 - \left(\sum_{i=1}^{N} u_i\right)^2\right)\right] = 0$$

because $\left(\sum_{i=1}^{N} x_i\right)^2 < N^2 = \left(\sum_{i=1}^{N} u_i\right)^2$. $\square$

The next statement follows directly from the lemma by noting that $\left|\sum_{i=1}^{N} x_i\right| = N$ is equivalent to $x \in \{-u, u\}$:

Corollary 13. *The limits*

$$\lim_{\beta\to\infty} \frac{p_\beta(x)}{\sum_{y\in\Omega_N} p_\beta(y)} = \begin{cases} \frac{1}{2} & \text{if } \left|\sum_{i=1}^{N} x_i\right| = N, \\ 0 & \text{otherwise,} \end{cases}$$

hold for all $x \in \Omega_N$.

Definition 14. We define the minimum of the range of S^2 to be $\kappa := \min_{x\in\Omega_N} \left\{\left|\sum_{i=1}^{N} x_i\right|\right\}$ and the set

$$\Upsilon := \left\{ x \in \Omega_N \;\middle|\; \left|\sum_{i=1}^{N} x_i\right| = \kappa \right\}.$$

Remark 15. Note that

$$\kappa = \begin{cases} 0 & \text{if } N \text{ is even,} \\ 1 & \text{otherwise,} \end{cases}$$

and the cardinality of Υ is

$$|\Upsilon| = \begin{cases} \begin{pmatrix} N \\ \frac{N}{2} \end{pmatrix} & \text{if } N \text{ is even,} \\ \begin{pmatrix} N \\ \frac{N+1}{2} \end{pmatrix} & \text{otherwise.} \end{cases}$$

Lemma 16. *Let $y \in \Upsilon$. Then the limits*

$$\lim_{\beta\to-\infty} \frac{p_\beta(x)}{p_\beta(y)} = \begin{cases} 1 & \text{if } \left|\sum_{i=1}^{N} x_i\right| = \kappa, \\ 0 & \text{otherwise,} \end{cases}$$

hold for all $x \in \Omega_N$.

Proof. Let $x, y \in \Upsilon$. Then we have

$$\lim_{\beta\to-\infty} \frac{p_\beta(x)}{p_\beta(y)} = \lim_{\beta\to-\infty} \exp\left[\frac{\beta}{2N}\left(\left(\sum_{i=1}^{N} x_i\right)^2 - \left(\sum_{i=1}^{N} y_i\right)^2\right)\right] = \lim_{\beta\to-\infty} \exp\left[\frac{\beta}{2N}(\kappa - \kappa)\right] = 1.$$

Now let $x \notin \Upsilon$. Then

$$\lim_{\beta\to-\infty} \frac{p_\beta(x)}{p_\beta(y)} = \lim_{\beta\to-\infty} \exp\left[\frac{\beta}{2N}\left(\left(\sum_{i=1}^{N} x_i\right)^2 - \left(\sum_{i=1}^{N} y_i\right)^2\right)\right] = 0$$

because $\left(\sum_{i=1}^{N} x_i\right)^2 - \left(\sum_{i=1}^{N} y_i\right)^2 > 0$. $\square$

The next corollary follows immediately from the last lemma:

Corollary 17. *The limits*

$$\lim_{\beta\to-\infty}\frac{p_\beta\left(x\right)}{\sum_{y\in\Omega_N}p_\beta\left(y\right)}=\begin{cases}\frac{1}{|\Upsilon|} & \text{if } \left|\sum_{i=1}^{N}x_i\right|=\kappa,\\ 0 & \text{otherwise,}\end{cases}$$

hold for all $x\in\Omega_N$.

Lemma 18. *The following statements hold:*

1. $\lim_{\beta\to-\infty}\mathbb{E}_{\beta,N}S^2=\kappa$.
2. $\lim_{\beta\to\infty}\mathbb{E}_{\beta,N}S^2=N^2$.

We will write

$$s\left(x\right):=\sum_{i=1}^{N}x_i$$

for all $x\in\Omega_N$.

Proof. We show each statement in turn.

1. We start by proving the statement

$$\lim_{\beta\to-\infty}\mathbb{E}_{\beta,N}S^2=\kappa.$$

The moment $\mathbb{E}_{\beta,N}S^2$ is calculated as the sum of 2^N summands of the type

$$Z_{\beta,N}^{-1}\,s\left(x\right)^2\exp\left(\frac{\beta}{2N}s\left(x\right)^2\right),\quad x\in\Omega_N.$$

Since the number of summands is fixed over all values of $\beta\in\mathbb{R}$, we have

$$\lim_{\beta\to-\infty}\mathbb{E}_{\beta,N}S^2=\sum_{x\in\Omega_N}s\left(x\right)^2\lim_{\beta\to-\infty}\frac{\exp\left(\frac{\beta}{2N}s\left(x\right)^2\right)}{Z_{\beta,N}}=\sum_{x\in\Omega_N}s\left(x\right)^2\lim_{\beta\to-\infty}\frac{p_\beta\left(x\right)}{\sum_{y\in\Omega}p_\beta\left(y\right)},$$

where $p_\beta\left(x\right)$ is defined as in (9). By Definition 14 of κ and Υ, for all $x\in\Upsilon$,

$$s\left(x\right)^2=\kappa.$$

According to Corollary 17, we obtain

$$\sum_{x\in\Omega_N}s\left(x\right)^2\lim_{\beta\to-\infty}\frac{p_\beta\left(x\right)}{\sum_{y\in\Omega_N}p_\beta\left(y\right)}=\sum_{x\in\Upsilon}s\left(x\right)^2\frac{1}{|\Upsilon|}=\kappa.$$

2. We next show the second statement employing Corollary 13, $s\left(u\right)^2=s\left(-u\right)^2=N^2$, and the fact that there are finitely many configurations $x\in\Omega_N$:

$$\begin{aligned}\lim_{\beta\to\infty}\mathbb{E}_{\beta,N}S^2&=\lim_{\beta\to\infty}\frac{\sum_{x\in\Omega_N}p_\beta\left(x\right)s(x)^2}{\sum_{y\in\Omega_N}p_\beta\left(y\right)}=\lim_{\beta\to\infty}\frac{p_\beta\left(u\right)}{p_\beta\left(u\right)+p_\beta\left(-u\right)}s(u)^2+\lim_{\beta\to\infty}\frac{p_\beta\left(-u\right)}{p_\beta\left(u\right)+p_\beta\left(-u\right)}s(-u)^2\\&=\frac{1}{2}s(u)^2+\frac{1}{2}s(-u)^2=N^2.\end{aligned}$$

□

As we will use the function $\beta \in \mathbb{R} \mapsto \mathbb{E}_{\beta,N} S^2 \in \mathbb{R}$ in our proofs later on, it is convenient to assign a letter to this function.

Definition 19. Let for each $N \in \mathbb{N}$ the function $\vartheta_N : [-\infty, \infty] \to \mathbb{R}$ be defined by

$$\vartheta_N(\beta) \coloneqq \mathbb{E}_{\beta,N} S^2,\ \beta \in \mathbb{R}, \qquad \vartheta_N(-\infty) \coloneqq \kappa, \qquad \text{and } \vartheta_N(\infty) \coloneqq N^2.$$

Proposition 20. *The function ϑ_N has the following properties:*

1. *ϑ_N is strictly increasing and continuous; it is continuously differentiable on $\mathbb{R}$. Its derivative is*

$$\vartheta_N'(\beta) = \frac{1}{2N} \mathbb{V}_{\beta,N} S^2, \quad \beta \in \mathbb{R}.$$

2. *We have for all $\beta \in (-\infty, 0)$ and all $N \in \mathbb{N}$:*

$$\kappa < \mathbb{E}_{\beta,N} S^2 < N.$$

3. *We have for all $\beta \in [0, \infty)$ and all $N \in \mathbb{N}$:*

$$N \leq \mathbb{E}_{\beta,N} S^2 < N^2.$$

Proof. 1. The continuity on $\mathbb{R}$ is clear. The function ϑ_N is a sum of differentiable functions of $\beta \in \mathbb{R}$, and we calculate

$$
\begin{aligned}
\frac{\mathrm{d}\mathbb{E}_{\beta,N}S^2}{\mathrm{d}\beta} &= \frac{\mathrm{d}}{\mathrm{d}\beta}\left[\sum_{x\in\Omega_N}\left(\sum_{i=1}^N x_i\right)^2 Z_{\beta,N}^{-1}\exp\left(\frac{\beta}{2N}\left(\sum_{i=1}^N x_i\right)^2\right)\right]\\
&= \sum_{x\in\Omega_N}\left[\frac{\frac{1}{2N}\left(\sum_{i=1}^N x_i\right)^4\exp\left(\frac{\beta}{2N}\left(\sum_{i=1}^N x_i\right)^2\right)Z_{\beta,N}}{Z_{\beta,N}^2}\right.\\
&\qquad\left.-\frac{\left(\sum_{i=1}^N x_i\right)^2\exp\left(\frac{\beta}{2N}\left(\sum_{i=1}^N x_i\right)^2\right)\frac{\mathrm{d}Z_{\beta,N}}{\mathrm{d}\beta}}{Z_{\beta,N}^2}\right]\\
&= \frac{1}{2N}\mathbb{E}_{\beta,N}S^4 - Z_{\beta,N}^{-2}\sum_{x\in\Omega_N}\left(\sum_{i=1}^N x_i\right)^2\exp\left(\frac{\beta}{2N}\left(\sum_{i=1}^N x_i\right)^2\right)\\
&\quad\cdot\sum_{y\in\Omega_N}\frac{1}{2N}\left(\sum_{i=1}^N y_i\right)^2\exp\left(\frac{\beta}{2N}\left(\sum_{i=1}^N y_i\right)^2\right)\\
&= \frac{1}{2N}\mathbb{E}_{\beta,N}S^4 - \frac{1}{2N}\sum_{x\in\Omega_N}\left(\sum_{i=1}^N x_i\right)^2 Z_{\beta,N}^{-1}\exp\left(\frac{\beta}{2N}\left(\sum_{i=1}^N x_i\right)^2\right)\\
&\quad\cdot\sum_{y\in\Omega_N}\left(\sum_{i=1}^N y_i\right)^2 Z_{\beta,N}^{-1}\exp\left(\frac{\beta}{2N}\left(\sum_{i=1}^N y_i\right)^2\right)\\
&= \frac{1}{2N}\mathbb{E}_{\beta,N}S^4 - \frac{1}{2N}\left(\mathbb{E}_{\beta,N}S^2\right)^2 = \frac{1}{2N}\mathbb{V}_{\beta,N}S^2 > 0,
\end{aligned}
$$

where we used the derivative of $Z_{\beta,N}$ provided in Lemma 51, and the strict inequality stems from the fact that the random variable S^2 is not almost surely constant for any $\beta \in \mathbb{R}$. Thus, $\vartheta_N|\mathbb{R}$ is continuously differentiable and strictly increasing. The continuity of ϑ_N follows from the limit statements in Lemma 18. Lemma 18 and the inequalities in points 2 and 3, which we will show now, yield the strict monotonicity of ϑ_N on its entire domain.

2. For $\beta = 0$, the X_i, $i \in \mathbb{N}_N$, are independent Rademacher random variables (see Definition 11) with

parameter $t = 0$. We calculate the second moment of S:

$$\begin{aligned}\mathbb{E}_{0,N} S^2 &= \mathbb{E}_{0,N} \left(\sum_{i=1}^{N} X_i\right)^2 = \mathbb{E}_{0,N} \left(\sum_{i=1}^{N} X_i^2 + \sum_{i\neq j} X_i X_j\right)\\ &= \sum_{i=1}^{N} \mathbb{E}_{0,N} X_i^2 + \sum_{i\neq j} \mathbb{E}_{0,N} \left(X_i X_j\right)\\ &= \sum_{i=1}^{N} 1 + \sum_{i\neq j} \mathbb{E}_{0,N} X_i \, \mathbb{E}_{0,N} X_j\\ &= N,\end{aligned}$$

where we used $\mathbb{E}_{0,N} X_i = 0$, $X_i^2 = 1$ and the independence of all X_i, X_j with $i \neq j$ under $\mathbb{P}_{0,N}$.
Next, we use $S^2 \geq \kappa$, $\mathbb{E}_{0,N} S^2 = N$, and the strict monotonicity of the function $\vartheta_N|\mathbb{R}$ to establish that for all $\beta \in (-\infty, 0)$,

$$\kappa \leq \lim_{\beta' \to -\infty} \mathbb{E}_{\beta',N} S^2 < \mathbb{E}_{\beta,N} S^2 < \mathbb{E}_{0,N} S^2 = N.$$

This proves the second statement.

3. Using the first property and $\lim_{\beta\to\infty} \mathbb{E}_{\beta,N} S^2 = N^2$ from Lemma 18, we have for all $N \in \mathbb{N}$ and all $\beta \geq 0$,

$$N = \mathbb{E}_{0,N} S^2 \leq \mathbb{E}_{\beta,N} S^2 < \lim_{\beta' \to \infty} \mathbb{E}_{\beta',N} S^2 = N^2.$$

□

Recall that we convened to write $T = (\boldsymbol{T})_\lambda$. Proposition 20 assures us that any realisation of the statistic T in the interval $\left[N, N^2\right)$ allows us to identify a unique value of $\hat{\beta}_N \in [0, \infty)$ such that the optimality condition (8) holds. It should be noted that realisations $T(x)$ in $[\kappa, N) \cup \left\{N^2\right\}$ are possible since the range of S^2 is $\left\{\kappa, (\kappa+2)^2, \ldots, N^2\right\}$. As T is defined to be the average of the realisations of S^2 over all observations $x^{(t)}$ in the sample $x \in \Omega_N^n$, this implies that the range of T includes N^2 and a subset of $[\kappa, N)$.

Taking into account that $\mathbb{E}\, T = \mathbb{E}_{\beta,N} S^2$, we obtain the following realisations of $\hat{\beta}_N$ depending on the value of T:

1. If $T(x) \in (\kappa, N)$, then a negative value for $\hat{\beta}_N$ satisfies (8), and for $T(x) = \kappa$, (8) holds if $\hat{\beta}_N = -\infty$.
2. If $T(x) \in \left[N, N^2\right)$, then (8) is satisfied for a unique value $\hat{\beta}_N \in [0, \infty)$.
3. If $T(x) = N^2$, then (8) holds if $\hat{\beta}_N = \infty$ but not for any finite value of $\hat{\beta}_N$.

These observations are the reason we defined $\hat{\beta}_N$ as a function which takes values in the extended real numbers, including $-\infty$ and ∞. This concludes the proof of Proposition 8.

We are interested in the case that the coupling constant β lies in $[0, \infty)$, as this models social cohesion. If $\beta > 0$, the edge case $T(x) \in [\kappa, N) \cup \left\{N^2\right\}$ is of little practical importance due to Proposition 27 below. We will use some concepts and results in the proof of said proposition, which we state before the proposition itself.

We will work with rate functions which are defined as Legendre transforms.

Definition 21. Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. Then the Legendre transform $f^* : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ of f is defined by $f^*(t) \coloneqq \sup_{x\in\mathbb{R}} \{xt - f(x)\}, t \in \mathbb{R}$.

We will employ two lemmas concerning the convexity of the Legendre transformation of convex functions to prove Proposition 27 below.

Lemma 22. *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. Then the Legendre transform f^* is convex.*

Proof. Let $\theta \in [0, 1]$ and $t_1, t_2 \in \mathbb{R}$. Then

$$\begin{aligned}
\theta f^*(t_1) + (1-\theta) f^*(t_2) &= \theta \sup_{x\in\mathbb{R}} \{t_1 x - f(x)\} + (1-\theta) \sup_{x\in\mathbb{R}} \{t_2 x - f(x)\} \\
&= \sup_{x\in\mathbb{R}} \{\theta t_1 x - \theta f(x)\} + \sup_{x\in\mathbb{R}} \{(1-\theta) t_2 x - (1-\theta) f(x)\} \\
&\geq \sup_{x\in\mathbb{R}} \{(\theta t_1 + (1-\theta) t_2) x - (\theta + (1-\theta)) f(x)\} \\
&= f^*(\theta t_1 + (1-\theta) t_2),
\end{aligned}$$

where we used first the definition of f^*, then $\theta \in [0, 1]$, and finally $\sup_{x\in\mathbb{R}} g(x) + \sup_{x\in\mathbb{R}} h(x) \geq \sup_{x\in\mathbb{R}} \{g(x) + h(x)\}$ for any functions $g, h : \mathbb{R} \to \mathbb{R}$. ☐

Lemma 23. *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex and differentiable function. Then the Legendre transform f^* is strictly convex on the set $\{t \in \mathbb{R} \mid f^*(t) < \infty\}$.*

Proof. We already know from Lemma 22 that f^* is convex. Assume it is not strictly convex on $F := \{t \in \mathbb{R} \mid f^*(t) < \infty\}$. Since f is assumed to be differentiable, and hence continuous, this implies the existence of $t_1, t_2 \in F, t_1 \neq t_2$, such that

$$f^*\left(\frac{t_1 + t_2}{2}\right) = \frac{1}{2} f^*(t_1) + \frac{1}{2} f^*(t_2). \tag{10}$$

See [9, p. 12] for a reference that for continuous real-valued functions defined on open intervals midpoint convexity is equivalent to convexity.

Let z be a maximiser of the set $\left\{x\frac{t_1+t_2}{2} - f(x) \,\middle|\, x \in \mathbb{R}\right\}$. Then the inequality

$$f^*\left(\frac{t_1 + t_2}{2}\right) = z\frac{t_1 + t_2}{2} - f(z) \geq x\frac{t_1 + t_2}{2} - f(x), \quad x \in \mathbb{R},$$

holds. By (10),

$$\begin{aligned}
\frac{1}{2} f^*(t_1) + \frac{1}{2} f^*(t_2) = f^*\left(\frac{t_1 + t_2}{2}\right) &= z\frac{t_1 + t_2}{2} - f(z) \\
&= \frac{1}{2}(z t_1 - f(z)) + \frac{1}{2}(z t_2 - f(z)) \\
&\leq \frac{1}{2} f^*(t_1) + \frac{1}{2} f^*(t_2),
\end{aligned}$$

and hence z is also a maximiser of the sets $\{x t_i - f(x) \mid x \in \mathbb{R}\}$, $i = 1, 2$, i.e.

$$z t_i - f(z) \geq x t_i - f(x), \quad x \in \mathbb{R},\ i = 1, 2.$$

By rearranging terms, we obtain

$$f(x) \geq x t_i - z t_i + f(z), \quad x \in \mathbb{R},\ i = 1, 2.$$

Thus we have two affine functions $x \mapsto g_i(x) := xt_i - zt_i + f(z), i = 1, 2$, with the properties

$$g_1 \neq g_2, \quad g_i \leq f, \quad g_i(z) = f(z), \quad i = 1, 2.$$

Hence each g_i is tangent to f at z. However, due to the assumed differentiability of f and $g_1 \neq g_2$ this is impossible, yielding a contradiction. Therefore, (10) cannot hold. $\square$

Definition 24. Let P be a probability measure on $\mathbb{R}$. Then let $\Lambda_P(t) := \ln \int_{\mathbb{R}} \exp(tx) P(\mathrm{d}x)$ for all $t \in \mathbb{R}$ such that the expression is finite. We call Λ_P the cumulant generating function of P and the Legendre transform Λ_P^* its entropy function. Let Y be a real random variable with distribution P. We will then say that $\Lambda_Y := \Lambda_P$ and $\Lambda_Y^* := \Lambda_P^*$ are the cumulant generating and entropy function of Y, respectively.

As a last ingredient we will need in the proof of Proposition 27, we prove some properties of Λ_{S^2} and $\Lambda_{S^2}^*$.

Definition 25. Let Y be a random variable. The value

$$\operatorname{ess\,inf} Y := \sup\{a \in \mathbb{R} \,|\, \mathbb{P}\{Y < a\} = 0\}$$

is called the essential infimum of Y. We convene that $\operatorname{ess\,inf} Y := -\infty$ if the set of essential lower bounds for Y on the right hand side of the display above is empty. The value

$$\operatorname{ess\,sup} Y := \inf\{a \in \mathbb{R} \,|\, \mathbb{P}\{Y > a\} = 0\}$$

is called the essential supremum of Y. We convene that $\operatorname{ess\,sup} Y := \infty$ if the set on the right hand side above is empty.

Lemma 26. *Let Y be a bounded random variable which is not almost surely constant. The cumulant generating function Λ_Y of Y and the entropy function Λ_Y^* of Y have the following properties:*

1. *Λ_Y is convex and differentiable.*
2. *Λ_Y^* is finite on the interval $(\operatorname{ess\,inf} Y, \operatorname{ess\,sup} Y)$ and infinite on $[\operatorname{ess\,inf} Y, \operatorname{ess\,sup} Y]^c$.*
3. *Λ_Y^* is strictly convex on $(\operatorname{ess\,inf} Y, \operatorname{ess\,sup} Y)$.*
4. *Λ_Y^* is strictly decreasing on the interval $(\operatorname{ess\,inf} Y, \mathbb{E}\, Y)$ and strictly increasing on $(\mathbb{E}\, Y, \operatorname{ess\,sup} Y)$.*
5. *Λ_Y^* has a unique global minimum at $\mathbb{E}\, Y$ with $\Lambda_Y^*(\mathbb{E}\, Y) = 0$.*

Proof. Since Y is bounded, Λ_Y is well-defined and finite for all $t \in \mathbb{R}$. We now show that Λ_Y is convex and differentiable. Let $\theta \in [0, 1]$ and $t_1, t_2 \in \mathbb{R}$. Then

$$\begin{aligned}
\Lambda_Y(\theta t_1 + (1-\theta) t_2) &= \ln \mathbb{E} \exp((\theta t_1 + (1-\theta) t_2) Y) \\
&= \ln \mathbb{E}\left[\exp(t_1 Y)^{\theta} \exp(t_2 Y)^{1-\theta}\right] \\
&\leq \ln\left[(\mathbb{E} \exp(t_1 Y))^{\theta} (\mathbb{E} \exp(t_2 Y))^{1-\theta}\right] \\
&= \theta \Lambda_Y(t_1) + (1-\theta) \Lambda_Y(t_2),
\end{aligned}$$

where we used Hölder's inequality. Note that the function $(t, Y) \mapsto g(t, Y) := \exp(tY)$ has the properties

1. $g(t, \cdot)$ is an integrable function with respect to the push-forward measure $\mathbb{P} \circ Y^{-1}$ for all $t \in \mathbb{R}$.
2. The partial derivative $\frac{\partial g(t,Y)}{\partial t}$ exists $\mathbb{P} \circ Y^{-1}$-almost surely for all $t \in \mathbb{R}$.
3. $\left|\frac{\partial g(t,Y)}{\partial t}\right| = |Y \exp(tY)|$ is $\mathbb{P} \circ Y^{-1}$-integrable for all $t \in \mathbb{R}$.

By Leibniz's integral rule, $\Lambda_Y = \ln \int g(\cdot, Y)\, \mathrm{d}\, \mathbb{P} \circ Y^{-1}$ is differentiable with

$$\Lambda_Y'(t) = \frac{\mathbb{E}\left(Y \exp\left(tY\right)\right)}{\mathbb{E} \exp\left(tY\right)}.$$

Since Λ_Y is convex and differentiable, Λ_Y^* is strictly convex by Lemma 23. This shows statements 1 and 3.
Let $y \in (\operatorname{ess\,inf} Y, \operatorname{ess\,sup} Y)$. We show that $\Lambda_Y^*(y) < \infty$ holds. We have $\mathbb{P}(Y < y), \mathbb{P}(Y > y) > 0$. Let for all $t \in \mathbb{R}$

$$f(t) \coloneqq yt - \Lambda_Y(t)$$

and

$$g(t) \coloneqq \exp f(t) = \frac{\exp(yt)}{\mathbb{E} \exp(tY)}.$$

We write

$$\mathbb{E} \exp(tY) = \mathbb{E} \exp(tY)\, \mathbb{1}_{\{Y \leq y\}} + \mathbb{E} \exp(tY)\, \mathbb{1}_{\{y < Y \leq \operatorname{ess\,sup} Y\}} + \mathbb{E} \exp(tY)\, \mathbb{1}_{\{\operatorname{ess\,sup} Y < Y\}}.$$

By Definition 25, we have $\mathbb{P}\{\operatorname{ess\,sup} Y < Y\} = 0$. Therefore, by dividing numerator and denominator by $\exp(yt)$, we obtain

$$g(t) = \frac{1}{\mathbb{E} \exp(t(Y - y))\, \mathbb{1}_{\{Y \leq y\}} + \mathbb{E} \exp(t(Y - y))\, \mathbb{1}_{\{y < Y \leq \operatorname{ess\,sup} Y\}}}.$$

We see that $\lim_{t \to \infty} \mathbb{E} \exp(t(Y - y))\, \mathbb{1}_{\{Y \leq y\}} = 0$ due to $\mathbb{P}(Y < y) > 0$, and
$\lim_{t \to \infty} \mathbb{E} \exp(t(Y - y))\, \mathbb{1}_{\{y < Y \leq \operatorname{ess\,sup} Y\}} = \infty$ due to $\mathbb{P}(Y > y) > 0$. It follows that $\lim_{t \to \infty} g(t) = 0$ and $\lim_{t \to \infty} f(t) = -\infty$.
Next we note that $\lim_{t \to -\infty} \mathbb{E} \exp(t(Y - y))\, \mathbb{1}_{\{Y \leq y\}} = \infty$ due to $\mathbb{P}(Y < y) > 0$,
and $\lim_{t \to -\infty} \mathbb{E} \exp(t(Y - y))\, \mathbb{1}_{\{y < Y \leq \operatorname{ess\,sup} Y\}} = 0$ due to $\mathbb{P}(Y > y) > 0$. Therefore, $\lim_{t \to -\infty} g(t) = 0$ and $\lim_{t \to -\infty} f(t) = -\infty$. Hence, the continuous function f reaches its maximum at some point $t_0 \in \mathbb{R}$, and this implies $\Lambda_Y^*(y) = \sup_{t \in \mathbb{R}} \{yt - \Lambda_Y(t)\} = \max_{t \in \mathbb{R}} f(t) = f(t_0) < \infty$. This shows that Λ_Y^* is finite on the interval $(\operatorname{ess\,inf} Y, \operatorname{ess\,sup} Y)$.
Now let $y \in [\operatorname{ess\,inf} Y, \operatorname{ess\,sup} Y]^c$. Assume first $y > \operatorname{ess\,sup} Y$. As $\mathbb{P}\{\operatorname{ess\,sup} Y < Y\} = 0$, we can write

$$\mathbb{E} \exp(tY) = \mathbb{E} \exp(tY)\, \mathbb{1}_{\{Y \leq \operatorname{ess\,sup} Y\}}$$

and

$$g(t) = \frac{1}{\mathbb{E} \exp(t(Y - y))\, \mathbb{1}_{\{Y \leq \operatorname{ess\,sup} Y\}}}.$$

Then $\lim_{t \to \infty} \mathbb{E} \exp(t(Y - y))\, \mathbb{1}_{\{Y \leq \operatorname{ess\,sup} Y\}} = 0$ and $\lim_{t \to -\infty} \mathbb{E} \exp(t(Y - y))\, \mathbb{1}_{\{Y \leq \operatorname{ess\,sup} Y\}} = \infty$. Therefore, $\lim_{t \to \infty} g(t) = \infty$ and $\lim_{t \to -\infty} g(t) = 0$. Hence, $\Lambda_Y^*(y) = \sup_{t \in \mathbb{R}} \{yt - \Lambda_Y(t)\} = \lim_{t \to \infty} f(t) = \infty$. Analogously, one can show $\Lambda_Y^*(y) = \sup_{t \in \mathbb{R}} \{yt - \Lambda_Y(t)\} = \lim_{t \to \infty} f(t) = \infty$ for all $y < \operatorname{ess\,inf} Y$. This concludes the proof of statement 2.
Next we show that $\Lambda_Y^*(\mathbb{E}Y) = 0$ and $\Lambda_Y^*(x) > 0$ for all $x \neq \mathbb{E}\, Y$. Jensen's inequality yields

$$\Lambda_Y(t) = \ln \mathbb{E} \exp(tY) \geq \mathbb{E}(\ln \exp(tY)) = t\, \mathbb{E}\, Y. \tag{11}$$

It follows directly from $\Lambda(0) = 0$ and Definition 21 that $\Lambda_Y^*(\mathbb{E}\, Y) = 0$, and the second part of statement 5 is proved.

We rearrange terms in (11) to obtain for all $t < 0$ and $x \geq \mathbb{E}\,Y$

$$xt - \Lambda_Y(t) \leq t\mathbb{E}\,Y - \Lambda_Y(t) \leq 0.$$

Since $\Lambda_Y(0) = 0$, $\Lambda_Y^*(x) \geq 0$ for all $x \in \mathbb{R}$ is a consequence of Definition 21. This and the last display yield

$$\Lambda_Y^*(x) = \sup_{t \geq 0} \{xt - \Lambda_Y(t)\}$$

for all $x \geq \mathbb{E}\,Y$. As the function $x \mapsto xt - \Lambda_Y(t)$ given $t \geq 0$ is increasing, we have shown that Λ_Y^* is increasing on the interval $(\mathbb{E}\,Y, \operatorname{ess\,sup} Y)$. Together with the strict convexity of Λ_Y^*, this implies that Λ_Y^* is strictly increasing on the interval $(\mathbb{E}\,Y, \operatorname{ess\,sup} Y)$. Analogously, one can show

$$\Lambda_Y^*(t) = \sup_{t \leq 0} \{xt - \Lambda_Y(t)\} \tag{12}$$

for all $x \leq \mathbb{E}\,Y$ and that Λ_Y^* is strictly decreasing on the interval $(\operatorname{ess\,inf} Y, \mathbb{E}\,Y)$, which completes the proof of statement 4 and the first part of statement 5. □

As an application of Lemma 26, we obtain the exponential convergence to 0 of the probability of the set of atypical realisations $\left\{T \notin \left[N, N^2\right)\right\}$.

Proposition 27. *For any value of the coupling constant $\beta > 0$, there is a constant $\delta := \inf\left\{\Lambda_{S^2}^*(t) \mid t \notin \left[N, N^2\right)\right\} > 0$ such that*

$$\mathbb{P}\left\{T \notin \left[N, N^2\right)\right\} \leq 2\exp(-\delta n)$$

holds for all $n \in \mathbb{N}$.

Proof. The random variable S^2 is bounded and not almost surely constant, so Lemma 26 applies to $\Lambda_{S^2}^*$. Set

$$\delta := \inf\left\{\Lambda_{S^2}^*(t) \mid t \notin \left[N, N^2\right)\right\}. \tag{13}$$

Due to $N < \mathbb{E}_{\beta,N} S^2 < N^2$ by Proposition 20 and the strict monotonicity of $\Lambda_{S^2}^*$ on each of the intervals $\left(\kappa, \mathbb{E}_{\beta,N} S^2\right)$ and $\left(\mathbb{E}_{\beta,N} S^2, N^2\right)$ by Lemma 26, we conclude that $\Lambda_{S^2}^*(N) > 0$ and $\Lambda_{S^2}^*\left(N^2\right) > 0$, and hence

$$\delta = \min\left\{\Lambda_{S^2}^*(N), \Lambda_{S^2}^*\left(N^2\right)\right\} > 0$$

holds. We write

$$\mathbb{P}\left\{T \notin \left[N, N^2\right)\right\} = \mathbb{P}\{T \in (-\infty, N)\} + \mathbb{P}\left\{T \in \left[N^2, \infty\right)\right\}.$$

An application of Markov's inequality yields for all $x \leq 0$

$$\begin{aligned}
\mathbb{P}\{T \in (-\infty, N)\} &= \mathbb{P}\{T - N < 0\} \leq \mathbb{P}\{\exp(nx(T-N)) \geq 1\} \leq \mathbb{E}\exp(nx(T-N)) \\
&= \exp(-nxN) \prod_{s=1}^{n} \mathbb{E}\exp\left(x\left(\sum_{i=1}^{N} X_i^{(s)}\right)^2\right) = \exp(-nxN)\left[\mathbb{E}\exp\left(xS^2\right)\right]^n \\
&= \exp(-nxN)\exp(n\Lambda_{S^2}(x)) = \exp(-n(xN - \Lambda_{S^2}(x))).
\end{aligned}$$

As this holds for all $x \leq 0$, we use $N < \mathbb{E}_{\beta,N} S^2$ and (12) to arrive at

$$\mathbb{P}\{T \in (-\infty, N)\} \leq \exp\left(-n\Lambda_{S^2}^*(N)\right). \tag{14}$$

Similarly, we calculate the upper bound

$$\mathbb{P}\left\{T \in \left[N^2, \infty\right)\right\} \leq \exp\left(-n\Lambda^*_{S^2}\left(N^2\right)\right). \tag{15}$$

Combining (14) and (15) yields the claim considering the definition of δ in (13). □

Remark 28. We define the constant $\bar{\delta}$ from Proposition 9 by setting

$$\bar{\delta} := \sum_{\lambda=1}^{M} \delta_\lambda, \tag{16}$$

with each of the δ_λ being of the form (13) with the entropy function in the definition being $\Lambda^*_{S^2_\lambda}$.

Proof of Proposition 9. The equivalences

$$\begin{aligned} T \in (-\infty, N) &\iff \hat{\beta}_N \in [-\infty, 0), \\ T = N^2 &\iff \hat{\beta}_N = \infty \end{aligned}$$

hold by Proposition 20. Proposition 9 now follows from Proposition 27. □

5 Proof of Theorem 10

Notation 29. Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of random variables and Y a random variable. We will write $Y_n \xrightarrow[n\to\infty]{\mathrm{p}} Y$ for the statement 'Y_n converges in probability to Y,' i.e.

$$\mathbb{P}\left\{|Y_n - Y| > \varepsilon\right\} \xrightarrow[n\to\infty]{} 0$$

holds for all $\varepsilon > 0$.
Let P_n be the distribution of Y_n, $n \in \mathbb{N}$, and P the distribution of Y. We will write $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$ or $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} P$ for the statement 'Y_n converges in distribution to Y,' i.e.

$$\int f \,\mathrm{d}P_n \xrightarrow[n\to\infty]{} \int f \,\mathrm{d}P$$

for all continuous and bounded functions $f : \mathbb{R} \to \mathbb{R}$.
We will refer to a normal distribution with mean $\eta \in \mathbb{R}$ and variance $\sigma^2 > 0$ as $\mathcal{N}\left(\eta, \sigma^2\right)$.
Let for any random variable X and any probability measure P the expression $X \sim P$ stand for 'X is distributed according to P.'

In the following, we will state some auxiliary results we will use in the proof of Theorem 10.

Recall that we use the symbol $\kappa \in \{0, 1\}$ for the minimum value the random variable S^2 can assume.

Recall Definition 19 of the function ϑ_N.

Lemma 30. *The function ϑ_N from Definition 19 has an inverse function $\vartheta_N^{-1} : \left[\kappa, N^2\right] \to [-\infty, \infty]$ which is strictly increasing and continuously differentiable on $\left(\kappa, N^2\right)$.*

Proof. The statements follow from Proposition 20 and the inverse function theorem. In particular, the continuous differentiability of ϑ_N^{-1} follows from the continuous differentiability of ϑ_N:

$$\vartheta_N'(x) > 0, \quad x \in \mathbb{R}.$$

So ϑ_N^{-1} is differentiable and

$$\left(\vartheta_N^{-1}\right)'(y) = \frac{1}{\vartheta_N'\left(\vartheta_N^{-1}(y)\right)}, \quad y \in \left(\kappa, N^2\right).$$

□

Remark 31. With the previous lemma, we can express the estimator $\hat{\beta}_N : \Omega_N^n \to [-\infty, \infty]$ as

$$\hat{\beta}_N(x) = \left(\vartheta_N^{-1} \circ T\right)(x), \quad x \in \Omega_N^n.$$

Notation 32. Let $(F, \mathcal{F})$ be a measurable space. We will write for any set $A \in \mathcal{F}$, the indicator function of A as $\mathbb{1}_A : F \to \mathbb{R}$,

$$\mathbb{1}_A(x) := \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \in F \setminus A. \end{cases}$$

Next we present an auxiliary lemma about the convergence in distribution of sequences of random variables.

Lemma 33. *Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of random variables and $(M_n)_{n\in\mathbb{N}}$ a sequence of positive numbers such that*

$$|Y_n| \leq M_n, \quad n \in \mathbb{N},$$

is satisfied. Let ν be a probability measure on $\mathbb{R}$, and assume the convergence $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} \nu$. Finally, let $(B_n)_{n\in\mathbb{N}}$ be a sequence of measurable sets which satisfies

$$\mathbb{P}\{Y_n \in B_n\} = o\left(\frac{1}{M_n}\right).$$

Then we have for all $z \in \mathbb{R}$,

$$\mathbb{1}_{\{Y_n \in B_n^c\}} Y_n + z \mathbb{1}_{\{Y_n \in B_n\}} \xrightarrow[n\to\infty]{\mathrm{d}} \nu.$$

Proof. Set $A_n := \mathrm{Range}(Y_n) \setminus B_n$ and

$$\begin{aligned} W_n &:= \mathbb{1}_{\{Y_n \in A_n\}} Y_n + z \mathbb{1}_{\{Y_n \in B_n\}}, \\ U_n &:= \mathbb{1}_{\{Y_n \in B_n\}} Y_n - z \mathbb{1}_{\{Y_n \in B_n\}} \end{aligned}$$

for all $n \in \mathbb{N}$. We have

$$Y_n = W_n + U_n, \tag{17}$$

and hence if we show

$$U_n \xrightarrow[n\to\infty]{\mathrm{p}} 0,$$

then

$$W_n \xrightarrow[n\to\infty]{\mathrm{d}} \nu$$

follows from $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} \nu$, (17), and Theorem 52.

Let $\varepsilon > 0$. We calculate

$$\begin{aligned}\mathbb{P}\left\{|U_n| \geq \varepsilon\right\} &\leq \frac{\mathbb{E}\,|U_n|}{\varepsilon} = \frac{1}{\varepsilon}\int_{\{Y_n\in B_n\}} |Y_n|\, \mathrm{d}\mathbb{P} \\ &\leq \frac{1}{\varepsilon}\, M_n\, \mathbb{P}\left\{Y_n \in B_n\right\} \xrightarrow[n\to\infty]{} 0,\end{aligned}$$

where the convergence to 0 follows from the assumption $\mathbb{P}\left\{Y_n \in B_n\right\} = o\left(\frac{1}{M_n}\right)$. Finally,

$$\mathbb{P}\left\{|z|\, \mathbb{1}_{\{Y_n\in B_n\}} \geq \varepsilon\right\} \leq \frac{|z|}{\varepsilon}\, \mathbb{P}\left\{Y_n \in B_n\right\} = \xrightarrow[n\to\infty]{} 0,$$

and thus $U_n \xrightarrow[n\to\infty]{\mathrm{p}} 0$ holds. $\square$

The next auxiliary results relate to large deviation principles and contraction principles (see Definition 55 and Theorem 57).

Lemma 34. *Let $\mathcal{X}$ be a metric space and $I : \mathcal{X} \to [0, \infty]$ a good rate function. Then, for any non-empty closed set $K \subset \mathcal{X}$, there is a point $x_K \in K$ such that*

$$I\left(x_K\right) = \inf_{x\in K} I\left(x\right).$$

Proof. If $\inf_{x\in K} I\left(x\right) = \infty$, then set x_K equal to an arbitrary point in K. So assume $\inf_{x\in K} I\left(x\right) < \infty$. Let $\left(x_n\right)_{n\in\mathbb{N}}$ be a sequence in K with $\lim_{n\to\infty} I\left(x_n\right) = \inf_{x\in K} I\left(x\right)$. Then there is a constant $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$,

$$I\left(x_n\right) \leq \inf_{x\in K} I\left(x\right) + 1 =: \alpha < \infty.$$

We have

$$x_n \in \left\{x \in \mathcal{X} \,|\, I\left(x\right) \leq \alpha\right\}, \quad n \geq n_0.$$

As I is a good rate function, the level set $\left\{x \in \mathcal{X} \,|\, I\left(x\right) \leq \alpha\right\}$ is compact, and therefore $\left(x_n\right)_{n\geq n_0}$ has a subsequence $\left(x_{n_k}\right)_{k\in\mathbb{N}}$ that converges to some $x_0 \in \left\{x \in \mathcal{X} \,|\, I\left(x\right) \leq \alpha\right\}$. Since $x_{n_k} \in K$ for all $k \in \mathbb{N}$, and K is closed, x_0 must belong to K. Next we note that

$$\begin{aligned}\inf_{x\in K} I\left(x\right) &= \lim_{n\to\infty} I\left(x_n\right) = \lim_{k\to\infty} I\left(x_{n_k}\right) \\ &= \liminf_{k\to\infty} I\left(x_{n_k}\right) \geq I\left(x_0\right),\end{aligned}$$

where the inequality is due to $x_{n_k} \xrightarrow[k\to\infty]{} x_0$ and the lower semi-continuity of I. As $x_0 \in K$, we also have $I\left(x_0\right) \geq \inf_{x\in K} I\left(x\right)$. Thus $I\left(x_0\right) = \inf_{x\in K} I\left(x\right)$ holds. $\square$

Lemma 35. *Let $\mathcal{X}$ and $\mathcal{Y}$ be metric spaces, $I : \mathcal{X} \to [0, \infty]$ a good rate function, and $f : \mathcal{X} \to \mathcal{Y}$ a continuous function. We define $J : \mathcal{Y} \to [0, \infty]$ by*

$$J\left(y\right) := \inf\left\{I\left(x\right) \,|\, x \in \mathcal{X}, f\left(x\right) = y\right\}, \quad y \in \mathcal{Y}.$$

Let $M_I \subset \mathcal{X}$ be the set of minima of I and $M_J \subset \mathcal{Y}$ the set of minima of J. Then $f\left(M_I\right) = M_J$ holds. In particular, if f is injective and I has a unique minimum at x_0, then J has a unique minimum at $f\left(x_0\right)$.

Proof. Let $y_0 \in f(M_I)$ and $x_0 \in M_I$ be such that $y_0 = f(x_0)$. By definition of J, we have

$$J(y_0) = \inf\{I(x) \mid x \in \mathcal{X}, f(x) = y_0\} \leq I(x_0)$$

because of $f(x_0) = y_0$. On the other hand, we have for all $x \in \mathcal{X}$

$$I(x) \geq I(x_0)$$

as $x_0 \in M_I$, and therefore,

$$J(y_0) = \inf\{I(x) \mid x \in \mathcal{X}, f(x) = y_0\} \geq I(x_0).$$

So we have established $J(y_0) = I(x_0)$. Let $y \in \mathcal{Y}$. We have

$$J(y) = \inf\{I(x) \mid x \in \mathcal{X}, f(x) = y\} \geq I(x_0) = J(y_0),$$

and thus $y_0 \in M_J$.
Now let $y_0 \in M_J$. If $y_0 \notin f(\mathcal{X})$, then

$$J(y_0) = \inf\{I(x) \mid x \in \mathcal{X}, f(x) = y_0\} = \inf\emptyset = \infty.$$

Since I is a rate function, there is some $x_1 \in \mathcal{X}$ such that $I(x_1) < \infty$, and $J(f(x_1)) \leq I(x_1) < \infty = J(y_0)$, contradicting the assumption $y_0 \in M_J$. Hence, $y_0 \in f(\mathcal{X})$ must hold, and $f^{-1}(\{y_0\})$ is a non-empty closed subset of $\mathcal{X}$ because f is continuous. We apply the previous lemma to obtain a point $x_0 \in f^{-1}(\{y_0\})$ with the property

$$I(x_0) = \inf\left\{I(x) \mid x \in f^{-1}(\{y_0\})\right\} = J(y_0) \leq J(y), \quad y \in \mathcal{Y}.$$

Therefore,

$$I(x_0) \leq I(x), \quad x \in f^{-1}(\{y\}), \quad y \in \mathcal{Y}.$$

Using the identity $\mathcal{X} = \bigcup_{y \in f(\mathcal{X})} f^{-1}(\{y\})$, we obtain

$$I(x_0) \leq I(x) \quad x \in \mathcal{X},$$

and thus $x_0 \in M_I$ and $y_0 = f(x_0) \in f(M_I)$ follow. □

Now are ready to begin the proof proper of Theorem 10. Recall that N is the number of voters, n is the number of observations in the sample, $\Lambda^*_{S^2}$ is the entropy function of S^2, and see Definition 55 of large deviation principles. Also recall Lemma 30 which states that the function ϑ_N^{-1} exists and is continuous and strictly increasing. In preparation for statement 3 of Theorem 10, we define the good rate function $J : [-\infty, \infty] \to [0, \infty]$ by

$$J(y) := \inf\left\{\Lambda^*_{S^2}(x) \mid x \in \mathbb{R}, \vartheta_N^{-1}(x) = y\right\}, \quad y \in [-\infty, \infty]. \tag{18}$$

Remark 36. We define the multivariate good rate function $\boldsymbol{J} : [-\infty, \infty]^M \to [0, \infty]$ from the statement of Theorem 10 by setting

$$\boldsymbol{J}(x) := \sum_{\lambda=1}^{M} J_\lambda(x_\lambda), \quad x \in [-\infty, \infty]^M, \tag{19}$$

with each of the J_λ being of the form (18).

Proof of Theorem 10. We prove each statement in turn.

1. Recall that $\left(x^{(1)}, \ldots, x^{(n)}\right) \in \Omega_N^n$ refers to the sample of voting configurations we observe. $x_i^{(t)} \in \{-1,1\}$, $t \in \mathbb{N}_n$, $i \in \mathbb{N}_N$, is the vote of individual i in the t-th vote. The weak law of large numbers

$$\frac{1}{n} \sum_{t=1}^{n} \left(\sum_{i=1}^{N} X_i^{(t)} \right)^2 \xrightarrow[n\to\infty]{\mathrm{p}} \mathbb{E}_{\beta,N} S^2 \tag{20}$$

holds because $\left(\sum_{i=1}^{N} X_i^{(t)}\right)^2$ is a bounded random variable, and thus the second moment exists. By Definition 7, we have

$$\mathbb{E}_{\hat{\beta}_N,N} S^2 = \frac{1}{n} \sum_{t=1}^{n} \left(\sum_{i=1}^{N} x_i^{(t)} \right)^2 .$$

This and (20) yield

$$\mathbb{E}_{\hat{\beta}_N,N} S^2 \xrightarrow[n\to\infty]{\mathrm{p}} \mathbb{E}_{\beta,N} S^2.$$

As noted above the inverse function ϑ_N^{-1}, which maps an expectation $\mathbb{E}_{\beta,N} S^2$ to the value β, is continuous. $\hat{\beta}_N \xrightarrow[n\to\infty]{\mathrm{p}} \beta$ follows by Theorem 53.

2. By Definition 3, the statistic T has the form specified in Proposition 56 with the function $f : \Omega_N \to \mathbb{R}$ given by

$$f(x_1, \ldots, x_N) := \left(\sum_{i=1}^{N} x_i \right)^2$$

for all $(x_1, \ldots, x_N) \in \Omega_N$. Hence, we have $\mu = \mathbb{E}_{\beta,N} S^2 = \mathbb{E}\, T$ and $\sigma^2 = \mathbb{V}_{\beta,N} S^2$. By Proposition 56,

$$\sqrt{n}\left(T - \mathbb{E}_{\beta,N} S^2\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \mathbb{V}_{\beta,N} S^2\right)$$

holds. We want to apply Theorem 54 to the transformation $f := \vartheta_N^{-1}$, but face the difficulty that ϑ_N^{-1} is not differentiable on the entire range of T. By assumption, $0 < \beta < \infty$. Therefore, by Proposition 20, we have $\kappa < \mathbb{E}_{\beta,N} S^2 < N^2$. Let $a < b$ be real numbers such that

$$\kappa < a < \mathbb{E}_{\beta,N} S^2 < b < N^2.$$

Set $K := (a,b)^c$ and $B_n := \left\{\sqrt{n}\left(T - \mathbb{E}_{\beta,N} S^2\right) \mid T \in K\right\}$ for all $n \in \mathbb{N}$. We define

$$W_n := \sqrt{n}\left(T - \mathbb{E}_{\beta,N} S^2\right) \mathbb{1}_{\{T \in K^c\}}, \quad n \in \mathbb{N}$$

and apply Lemma 33 to the sequence $Y_n := \sqrt{n}\left(T - \mathbb{E}_{\beta,N} S^2\right)$.
Note that B_n is finite for all $n \in \mathbb{N}$, so $B := \bigcup_{n=1}^{\infty} B_n$ is countable hence closed. Set $\nu := \mathcal{N}\left(0, \mathbb{V}_{\beta,N} S^2\right)$ and $M_n := \sqrt{n}$ for each $n \in \mathbb{N}$. The set K is closed and $T \in K$ if and only if $Y_n \in B_n$. By statement 4 of Proposition 56, we have

$$\mathbb{P}\{Y_n \in B\} \leq 2 \exp\left(-n \inf_{x \in K} \Lambda_{S^2}^{*}(x)\right), \quad n \in \mathbb{N}.$$

Therefore, $\mathbb{P}\left\{Y_n \in B\right\} = o\left(\frac{1}{M_n}\right)$ holds, and we can apply Lemma 33 to conclude

$$W_n \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \mathbb{V}_{\beta,N} S^2\right).$$

We apply Theorem 54. Set $D \coloneqq [a,b]$ and $f : D \to \mathbb{R}$

$$f(y) \coloneqq \vartheta_N^{-1}(y), \quad y \in D.$$

f is continuously differentiable and the derivative f' is strictly positive on the compact set D. Therefore,

$$\begin{aligned}
&\sqrt{n}\left(\hat{\beta}_N - \beta\right) \mathbb{1}_{\{T \in K^c\}} \\
&= \sqrt{n}\left(\hat{\beta}_N - \beta\right) \mathbb{1}_{\{T \in K^c\}} + \sqrt{n}\left(\beta - \beta\right) \mathbb{1}_{\{T \in K\}} \\
&= \sqrt{n}\left(\left(\hat{\beta}_N \mathbb{1}_{\{T \in K^c\}} + \beta \mathbb{1}_{\{T \in K\}}\right) - \beta\right) \\
&= \sqrt{n}\left(f\left(T \mathbb{1}_{\{T \in K^c\}} + \mathbb{E}_{\beta,N} S^2 \mathbb{1}_{\{T \in K\}}\right) - f\left(\mathbb{E}_{\beta,N} S^2\right)\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \left(f'(\mu)\right)^2 \sigma^2\right)
\end{aligned}$$

holds. We apply Lemma 33 once more to conclude that

$$\sqrt{n}\left(\hat{\beta}_N - \beta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \left(f'(\mu)\right)^2 \sigma^2\right)$$

is satisfied.
We use Proposition 20 which states that $\vartheta_N'(\beta) = \frac{1}{2N} \mathbb{V}_{\beta,N} S^2$, and Lemma 30 provides the derivative

$$\left(\vartheta_N^{-1}\right)'\left(\mathbb{E}_{\beta,N} S^2\right) = \frac{1}{\vartheta_N'\left(\vartheta_N^{-1}\left(\mathbb{E}_{\beta,N} S^2\right)\right)} = \frac{1}{\vartheta_N'(\beta)} = \frac{2N}{\mathbb{V}_{\beta,N} S^2}.$$

Substituting the value of the derivative in the previous display yields the claim concerning the asymptotic normality of the estimator $\hat{\beta}_N$.

3. By Definition 3, the statistic T has the form specified in Proposition 56 with the function $f : \Omega_N \to \mathbb{R}$ given by

$$f(x_1, \ldots, x_N) \coloneqq \left(\sum_{i=1}^{N} x_i\right)^2$$

for all $(x_1, \ldots, x_N) \in \Omega_N$. Hence, we have $\mu = \mathbb{E}_{\beta,N} S^2$, $\sigma^2 = \mathbb{V}_{\beta,N} S^2$, and $\Lambda_{S^2}^*$ for the objects in Proposition 56, and, therefore, T satisfies a large deviations principle with rate n and rate function $\Lambda_{S^2}^*$. As noted before, the function ϑ_N^{-1} is continuous, we have $T\left(x^{(1)}, \ldots, x^{(n)}\right) \in \left[\kappa, N^2\right]$ for all $n \in \mathbb{N}$ and all $\left(x^{(1)}, \ldots, x^{(n)}\right) \in \Omega_N^n$, and $\left[\kappa, N^2\right]$ is the domain of ϑ_N^{-1}. By Theorem 57, $\hat{\beta}_N = \vartheta_N^{-1} \circ T$ satisfies a large deviations principle with rate n and rate function J.
By Lemma 26, $\Lambda_{S^2}^*$ has a unique minimum at $\mathbb{E}_{\beta,N} S^2$. Since Lemma 30 also says that ϑ_N^{-1} is strictly increasing, by Lemma 35, J has a unique minimum at $\beta = \vartheta_N^{-1}\left(\mathbb{E}_{\beta,N} S^2\right)$. Let $K \subset \mathbb{R}$ be a closed set that does not contain β. We have

$$\mathbb{P}\left\{\hat{\beta}_N \in K\right\} = \mathbb{P}\left\{\vartheta_N\left(\hat{\beta}_N\right) \in \vartheta_N(K)\right\} = \mathbb{P}\left\{\mathbb{E}_{\hat{\beta}_N,N} S^2 \in \vartheta_N(K)\right\} = \mathbb{P}\left\{T \in \vartheta_N(K)\right\},$$

where the last step uses Definition 7. Now note that K is a closed subset of the compact space $[-\infty, \infty]$ and therefore compact. The continuous function ϑ_N maps it to the compact set $\vartheta_N(K) \subset \mathbb{R}$. Hence, $\vartheta_N(K)$ is a closed set. Since K does not contain β and ϑ_N is strictly increasing, $\vartheta_N(K)$ does not contain $\mathbb{E}_{\beta,N} S^2$. By Proposition 56, we have

$$\mathbb{P}\{T \in \vartheta_N(K)\} \leq 2\exp(-\delta' n)$$

for all $n \in \mathbb{N}$ with $\delta' := \inf_{x \in \vartheta_N(K)} \Lambda^*_{S^2}(x) > 0$. The equality

$$\begin{aligned}\delta = \inf_{y \in K} J(y) &= \inf_{y \in K} \inf\left\{\Lambda^*_{S^2}(x) \mid x \in \mathbb{R}, \vartheta_N^{-1}(x) = y\right\} = \inf\left\{\Lambda^*_{S^2}(x) \mid x \in \mathbb{R}, \vartheta_N^{-1}(x) \in K\right\} \\ &= \inf\left\{\Lambda^*_{S^2}(x) \mid x \in \vartheta_N(K)\right\} = \delta'\end{aligned}$$

yields the final claim.

□

6 Standard Error of the Statistic T and the Estimator $\hat{\beta}_N$

In order to calculate moments in the proof of Lemma 40 below, we introduce the concept of profile vectors.

Definition 37. Let $k, n \in \mathbb{N}$ with $k \leq n$. We will call all $\underline{\boldsymbol{i}} \in \mathbb{N}_n^k$ *index vectors* and set

$$\Pi := \left\{(r_1, \ldots, r_k) \in (\mathbb{N}_k)^k \;\middle|\; \sum_{\ell=1}^{k} \ell r_\ell = k\right\},$$

and we will refer to Π as the set of profile vectors and to the elements $\underline{\boldsymbol{r}} \in \Pi$ as *profile vectors*. For any index vector $\underline{\boldsymbol{i}} \in \mathbb{N}_n^k$, the expression $\underline{\boldsymbol{r}} := (r_1, \ldots, r_k) := \boldsymbol{\rho}(\underline{\boldsymbol{i}})$ is defined as follows: for each $\ell \in \mathbb{N}_k$, r_ℓ is the number of indices in $\underline{\boldsymbol{i}}$ that appear exactly ℓ times. We will call $\underline{\boldsymbol{r}}$ the profile vector of $\underline{\boldsymbol{i}}$.

Remark 38. It is an elementary fact that for each index vector $\underline{\boldsymbol{i}} \in \mathbb{N}_n^k$, $\boldsymbol{\rho}(\underline{\boldsymbol{i}}) \in \Pi$ holds.

Lemma 39. *Let $k \in \mathbb{N}$ and $\underline{\boldsymbol{r}} \in \Pi$. The number of index vectors $\underline{\boldsymbol{i}} \in \mathbb{N}_n^k$ such that $\underline{\boldsymbol{r}} = (r_1, \ldots, r_k) = \boldsymbol{\rho}(\underline{\boldsymbol{i}})$ is given by*

$$\frac{n!}{r_1! \cdots r_k! \left(n - \sum_{\ell=1}^k r_\ell\right)!} \frac{k!}{1!^{r_1} \cdots k!^{r_k}}.$$

Proof. We construct an index vector $\underline{\boldsymbol{i}}$ with entries in $\mathbb{N}_n$ and profile $\underline{\boldsymbol{r}} = \boldsymbol{\rho}(\underline{\boldsymbol{i}})$ in two steps:

1. We partition $\mathbb{N}_n$ into $k+1$ sets B_j, $j \in \mathbb{N}_{k+1}$. Each set B_j contains the indices which occur exactly j times in $\underline{\boldsymbol{i}}$ for $j \leq k$ and $B_{k+1} := \mathbb{N}_n \setminus \bigcup_{j=1}^k B_j$. Hence, $|B_j| = r_j$ holds for all $j \leq k$. There are

$$\begin{pmatrix} & & n! & \\ r_1! & r_2! & r_k! & \left(N - \sum_{\ell=1}^k r_\ell\right)! \end{pmatrix} \tag{21}$$

   ways to realise this partition of $\mathbb{N}_n$.
2. The index vector $\underline{\boldsymbol{i}}$ is of length k. We can think of the elements of $\bigcup_{j=1}^k B_j$ as a finite alphabet. Our task is to assemble an ordered k-tuple $\underline{\boldsymbol{i}}$ of elements of $\bigcup_{j=1}^k B_j$ in such a fashion that for each $j \leq k$,

each of the r_j elements of B_j occurs exactly j times. So there are r_1 elements of $\bigcup_{j=1}^k B_j$ selected once, r_2 elements selected twice,..., r_k elements selected k times. There are

$$\binom{k!}{1 \;\; \ldots \;\; 1 \;\; 2 \;\; \ldots \;\; 2 \;\; \ldots \;\; k! \;\; \ldots \;\; k!} = \frac{k!}{1!^{r_1} \cdots k!^{r_k}}$$

ways to accomplish this task.

□

The statistic T from Definition 3 is an unbiased estimator of the expectation $\mathbb{E}_{\beta,N} S^2$. Since each summand composing T takes values in $\left[\kappa, N^2\right]$, T is a bounded random variable.

Lemma 40. *Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables with mean $\mathbb{E}\, Y_n = \mu$, variance $\mathbb{V}\, Y_n = \sigma^2$, and existing moments $\mathbb{E}\,|X_n|^k < \infty$ for all $k, n \in \mathbb{N}$. Let Z be a random variable with distribution $\mathcal{N}\left(0, \sigma^2\right)$. Then*

$$\mathbb{E}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \mu)\right)^k \xrightarrow[n\to\infty]{} \mathbb{E}\, Z^k$$

holds for all $k \in \mathbb{N}$.

Proof. We will use the concepts from Definition 37 and Lemma 39 to calculate the expectation $\mathbb{E}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \mu)\right)^k$. To simplify the notation, we set $U_i \coloneqq Y_i - \mu$ for all $i \in \mathbb{N}$.

As the random variables $(U_n)_{n\in\mathbb{N}}$ are i.i.d., we have $\mathbb{E}\, U_{i_1} \cdots U_{i_k} = \mathbb{E}\, U_{j_1} \cdots U_{j_k}$ for all index vectors $\underline{\boldsymbol{i}}, \underline{\boldsymbol{j}}$ with $\boldsymbol{\rho}(\underline{\boldsymbol{i}}) = \boldsymbol{\rho}(\underline{\boldsymbol{j}})$. Therefore, we can write for any profile index vector $\underline{\boldsymbol{i}}$ with profile vector $\underline{\boldsymbol{r}} = \boldsymbol{\rho}(\underline{\boldsymbol{i}})$

$$\mathbb{E}\, U(\underline{\boldsymbol{r}}) \coloneqq \mathbb{E}\, U_{i_1} \cdots X_{i_k}.$$

Using Lemma 39, we express the expectation $\mathbb{E}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n U_i\right)^k$ as

$$\begin{aligned} \mathbb{E}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n U_i\right)^k &= \frac{1}{n^{\frac{k}{2}}} \sum_{i_1,\ldots,i_k=1}^n \mathbb{E}\, U_{i_1} \cdots U_{i_k} \\ &= \frac{1}{n^{\frac{k}{2}}} \sum_{\underline{\boldsymbol{r}}\in\Pi} \frac{n!}{r_1! \cdots r_k! \left(n - \sum_{\ell=1}^k r_\ell\right)!} \frac{k!}{1!^{r_1} \cdots k!^{r_k}} \mathbb{E}\, U(\underline{\boldsymbol{r}}). \end{aligned} \tag{22}$$

As a first step, we note that since $\mathbb{E}\, U_i = 0$ and $\mathbb{E}\, U_i^\ell U_j^m = \mathbb{E}\, U_i^\ell\, \mathbb{E}\, U_j^m$ for all $i \neq j$ and all $\ell, m \in \mathbb{N}$, we have for all $\underline{\boldsymbol{r}} \in \Pi$ with $r_1 \geq 1$,

$$\mathbb{E}\, U(\underline{\boldsymbol{r}}) = 0. \tag{23}$$

Now let $r_1 = 0$ and $r_{\ell^*} > 0$ for some $\ell^* > 2$. Set $\lfloor x \rfloor \coloneqq \max\{m \in \mathbb{Z} \,|\, m \leq x\}$ for all $x \in \mathbb{R}$. Suppose first that $\ell^* = 2j^* + 1$ for some $j^* \geq 1$. Then

$$\begin{aligned}
\sum_{\ell=1}^{k} r_\ell &= \frac{1}{2}\left[\sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} 2r_{2\ell+1} + \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} 2r_{2\ell}\right] \\
&= \frac{1}{2}\left[2r_{2j^*+1} + \sum_{1\leq\ell\leq\lfloor \frac{k}{2} \rfloor, \ell\neq j^*} 2r_{2\ell+1} + \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} 2r_{2\ell}\right] \\
&\leq \frac{1}{2}\left[(2j^*+1)\, r_{2j^*+1} + \sum_{1\leq\ell\leq\lfloor \frac{k}{2} \rfloor, \ell\neq j^*} 2r_{2\ell+1} + \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} 2r_{2\ell}\right] - \frac{1}{2} r_{2j^*+1} \\
&\leq \frac{1}{2}\left[(2j^*+1)\, r_{2j^*+1} + \sum_{1\leq\ell\leq\lfloor \frac{k}{2} \rfloor, \ell\neq j^*} 2r_{2\ell+1} + \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} 2r_{2\ell}\right] - \frac{1}{2} \\
&\leq \frac{1}{2}\left[\sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} (2\ell+1)\, r_{2\ell+1} + \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} 2\ell r_{2\ell}\right] - \frac{1}{2} \\
&= \frac{k}{2} - \frac{1}{2}.
\end{aligned}$$

Now suppose that $\ell^* = 2j^*$ for some $j^* \geq 2$. Then

$$\begin{aligned}
\sum_{\ell=1}^{k} r_\ell &= \frac{1}{2}\left[\sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} 2r_{2\ell+1} + \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} 2r_{2\ell}\right] \\
&= \frac{1}{2}\left[2r_{2j^*} + \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} 2r_{2\ell+1} + \sum_{1\leq\ell\leq\lfloor \frac{k}{2} \rfloor, \ell\neq j^*} 2r_{2\ell}\right] \\
&\leq \frac{1}{2}\left[2j^* r_{2j^*} + \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} 2r_{2\ell+1} + \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} 2r_{2\ell}\right] - r_{2j^*} \\
&\leq \frac{1}{2}\left[2j^* r_{2j^*} + \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} 2r_{2\ell+1} + \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} 2r_{2\ell}\right] - 1 \\
&= \frac{1}{2}\left[\sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} (2\ell+1)\, r_{2\ell+1} + \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor} 2\ell r_{2\ell}\right] - 1 \\
&= \frac{k}{2} - 1.
\end{aligned}$$

So we have the upper bound $\frac{k}{2}-\frac{1}{2}$ for $\sum_{\ell=1}^{k} r_\ell$ independently of the parity of ℓ^*. Hence,

$$\begin{aligned}
\frac{1}{n^{\frac{k}{2}}}\frac{n!}{r_1!\cdots r_k!\left(n-\sum_{\ell=1}^{k} r_\ell\right)!}\frac{k!}{1!^{r_1}\cdots k!^{r_k}}\mathbb{E}\,U\left(\underline{\boldsymbol{r}}\right) &\leq C\frac{1}{n^{\frac{k}{2}}}\frac{n!}{\left(n-\sum_{\ell=1}^{k} r_\ell\right)!}\\
&\leq C\frac{1}{n^{\frac{k}{2}}}n^{\sum_{\ell=1}^{k} r_\ell}\\
&\leq C\frac{1}{\sqrt{n}}. \qquad (24)
\end{aligned}$$

Finally, we set

$$\Pi_1 := \left\{\underline{\boldsymbol{r}}\in\Pi\,|\,r_\ell=0,\ell\neq 2\right\}.$$

Note that $\Pi_1=\emptyset$ if k is odd and $\Pi_1=\left\{\left(0,\frac{k}{2},0,\ldots,0\right)\right\}$ if k is even. This means by (23) and (24) that

$$\mathbb{E}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i\right)^k = \frac{1}{n^{\frac{k}{2}}}\sum_{\underline{\boldsymbol{r}}\in\Pi}\frac{n!}{r_1!\cdots r_k!\left(n-\sum_{\ell=1}^{k} r_\ell\right)!}\frac{k!}{1!^{r_1}\cdots k!^{r_k}}\mathbb{E}\,U\left(\underline{\boldsymbol{r}}\right)\leq C\frac{1}{\sqrt{n}}$$

if k is odd. Now let k be even, and hence $\Pi_1=\left\{\left(0,\frac{k}{2},0,\ldots,0\right)\right\}$. Then

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i\right)^k &\approx \frac{1}{n^{\frac{k}{2}}}\frac{n^{\sum_{\ell=1}^{k} r_\ell}}{\left(\frac{k}{2}\right)!}\frac{k!}{(2!)^{\frac{k}{2}}}\mathbb{E}\,U_1^2\cdots U_{\frac{k}{2}}^2\\
&= \frac{1}{n^{\frac{k}{2}}}\frac{n^{\frac{k}{2}}}{\left(\frac{k}{2}\right)!}\frac{k!}{2^{\frac{k}{2}}}\sigma^k\\
&= (k-1)!!\,\sigma^k,
\end{aligned}$$

where we used the identity $\frac{k!}{\left(\frac{k}{2}\right)!2^{\frac{k}{2}}}=(k-1)!!$ for even k. The last expression above is the moment of order k, for k even, of the distribution $\mathcal{N}\left(0,\sigma^2\right)$. The odd moments of $\mathcal{N}\left(0,\sigma^2\right)$ are 0 and by (23) and (24) for odd k all summands of (22) converge to 0. This concludes the proof. ☐

The previous lemma serves to determine the limiting standard error of the statistic T.

Proposition 41. *The standard error of the statistic T satisfies*

$$\lim_{n\to\infty}\sqrt{n}\sqrt{\mathbb{V}\,T}=\sqrt{\mathbb{V}_{\beta,N}S^2}.$$

Proof. This follows directly from Theorem 10 and Lemma 40. ☐

While the statistic T has a finite standard deviation for all population sizes N and all sample sizes n, the same is not true for the estimator $\hat{\beta}_N$ as this estimator assumes the values $\pm\infty$ with positive probability for all N and all n. Thus a similar statement does not hold for finite n nor in the limit $n\to\infty$. However, instead of using Lemma 40 to determine the limiting standard error, we can employ statement 2 of Theorem 10 to obtain the limiting probabilities for a certain class of events. Below, $\mathcal{N}\left(0,\sigma^2\right)A$ stands for the probability of any measurable set A under the $\mathcal{N}\left(0,\sigma^2\right)$ distribution.

Proposition 42. *Let $A \subset \mathbb{R}$ be a Lebesgue measurable set. Then we have*

$$\mathbb{P}\left\{\hat{\beta}_N \in \beta + \frac{1}{\sqrt{n}} A\right\} \xrightarrow[n\to\infty]{} \mathcal{N}\left(0, \frac{4N^2}{\mathbb{V}_{\beta,N} S^2}\right) A.$$

In particular, we have for any $\varepsilon > 0$, $\mathbb{P}\left\{\left|\hat{\beta}_N - \beta\right| \geq \frac{\varepsilon}{\sqrt{n}}\right\} \xrightarrow[n\to\infty]{} \mathcal{N}\left(0, \frac{4N^2}{\mathbb{V}_{\beta,N} S^2}\right)(-\varepsilon, \varepsilon)^c$.

Proof. Theorem 10 states that

$$\sqrt{n}\left(\hat{\beta}_N - \beta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \frac{4N^2}{\mathbb{V}_{\beta,N} S^2}\right).$$

By the definition of convergence in distribution and the absolute continuity of the normal distribution, we have for any measurable set A,

$$\begin{aligned}
\mathbb{P}\left\{\hat{\beta}_N \in \beta + \frac{1}{\sqrt{n}} A\right\} &= \mathbb{P}\left\{\hat{\beta}_N - \beta \in \frac{1}{\sqrt{n}} A\right\} \\
&= \mathbb{P}\left\{\sqrt{n}\left(\hat{\beta}_N - \beta\right) \in A\right\} \\
&\xrightarrow[n\to\infty]{} \mathcal{N}\left(0, \frac{4N^2}{\mathbb{V}_{\beta,N} S^2}\right) A.
\end{aligned}$$

□

Even in the absence of a finite standard deviation for the estimator $\hat{\beta}_N$, the previous proposition gives us an explicit limit for the probability of a deviation of order $\frac{1}{\sqrt{n}}$ of the estimator $\hat{\beta}_N$ from the true parameter value β.

7 Optimal Weights in a Two-Tier Voting System

A two-tier voting system describes a scenario where the population of a state or union of states is divided into M groups (e.g., member states). Each group sends a representative to a common council that makes decisions for the union. These representatives cast their votes ('yes' or 'no') based on the majority opinion within their respective group. Each group $\lambda \in \mathbb{N}_M$ is assumed to be of size $N_\lambda \in \mathbb{N}$. The votes of individual voter are represented by the random variable $X_{\lambda i}$, where $\lambda \in \mathbb{N}_M$ indicates the group and $i \in N_\lambda$ the individual within the group. Recall the group voting margins S_λ defined in (3), and set

$$\bar{S} := \sum_{\lambda=1}^{M} S_\lambda$$

to be the overall voting margin.

Given the group voting margin S_λ, we define the council vote of each group.

Definition 43. The council vote of each group $\lambda \in \mathbb{N}_M$ is

$$\chi_\lambda := \begin{cases} 1 & \text{if } S_\lambda > 0, \\ -1 & \text{otherwise.} \end{cases}$$

Since the groups may vary in size, it is natural to assign different voting weights w_λ to each representative, reflecting the relative sizes of their groups. In some situations, the groups can be formed in such a manner that they have similar sizes. For example, when electing a parliament such as the U.S. House of Representatives, the country is typically divided into districts with roughly equal populations, each receiving one seat. This approach is practical within a single country but becomes less feasible in other contexts. It would be impractical to reassemble countries or member states (like those in the United Nations or European Union) into districts of equal size or equally populated groups due to sovereignty concerns. Thus, the issue of how to assign voting weights to groups of different sizes cannot be avoided.

The problem of determining these optimal weights involves minimising the democracy deficit (see Definition 44 below), i.e. the deviation between a council vote and an idealised referendum across the entire population. This concept was first explored for binary voting by Felsenthal and Machover [12], and later analysed in various contexts by other authors (e.g., [17, 34, 18, 28, 31, 22, 21]. Informally, imposing the democracy deficit as a criterion, we endeavour to assign the voting weights in the council in such fashion that, on average, the council vote shall be as close as possible to a hypothetical referendum.

Other approaches to optimal weights are based on welfare considerations or the goal of equalising the influence of all voters within the overall population. The seminal work by Penrose [30], which introduced the square root law as a rule for assigning voting weights that equalises each voter's probability of being decisive in a two-tier system under the assumption of independent voting, exemplifies this approach. Further contributions to understanding optimal voting weights from welfare and influence perspectives can be found in [2, 24, 25].

In order to define the democracy deficit, we need a voting model that describes voting behaviour across the overall population. We will assume that the overall population behaves according to a Curie-Weiss model $\mathbb{P}_{\boldsymbol{\beta},\boldsymbol{N}}$ (see Definition 1) for some fixed $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_M) \geq 0$ and $\boldsymbol{N} = (N_1, \ldots, N_M) \in \mathbb{N}^M$. With a voting model in place, we can proceed to define the democracy deficit, which will serve as a criterion for the determination of the optimal voting weights each group receives in the council.

Definition 44. The democracy deficit for a set of voting weights $w_1, \ldots, w_M \in \mathbb{R}$ is given by

$$\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{N}} \left[\bar{S} - \sum_{\lambda=1}^{M} w_\lambda \chi_\lambda \right]^2 .$$

We will call any vector $(w_1, \ldots, w_M) \in \mathbb{R}$ of weights which minimises the democracy deficit 'optimal'.

Proposition 45. *For all $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_M) \geq 0$, the* optimal weights *are given by*

$$w_\lambda = \mathbb{E}_{\beta_\lambda, N_\lambda} |S_\lambda| , \quad \lambda \in \mathbb{N}_M .$$

Proof. This result was first proved in [17]. For the reader's convenience, we present a short proof here.

We find the minimum of the expression defining the democracy deficit by deriving $\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{N}} \left[S - \sum_{\lambda=1}^{M} w_\lambda \chi_\lambda \right]^2$ with respect to each w_λ, $\lambda \in \mathbb{N}_M$, and equating each derivative to 0:

$$\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{N}} \left[\left(\bar{S} - \sum_{\nu=1}^{M} w_\nu \chi_\nu \right) \chi_\lambda \right] = 0,$$

which is equivalent to

$$\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{N}} \, \bar{S} \, \chi_\lambda = \sum_{\nu=1}^{M} w_\nu \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{N}} \chi_\nu \chi_\lambda .$$

Due to Definitions 1 and 43, we have

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{N}}\, \bar{S}\, \chi_\lambda &= \sum_{\nu=1}^{M} \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{N}}\, S_\nu\, \chi_\lambda \\ &= \mathbb{E}_{\beta_\lambda,N_\lambda}\, S_\lambda\, \chi_\lambda \\ &= \mathbb{E}_{\beta_\lambda,N_\lambda} \left|S_\lambda\right|\end{aligned}$$

and

$$\begin{aligned}\sum_{\nu=1}^{M} w_\nu\, \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{N}} S_\nu \chi_\lambda &= w_\lambda\, \mathbb{E}_{\beta_\lambda,N_\lambda} \chi_\lambda^2 \\ &= w_\lambda.\end{aligned}$$

The optimality condition is therefore

$$w_\lambda = \mathbb{E}_{\beta_\lambda,N_\lambda} \left|S_\lambda\right|, \quad \lambda \in \mathbb{N}_M.$$

□

We will now show that the expectation $\mathbb{E}_{\beta_\lambda,N_\lambda} |S|$ of the voting margin of a single group is strictly increasing in the respective coupling constant β_λ (cf. Proposition 20). The following is an auxiliary lemma for this purpose.

Lemma 46. *Let X and Y be random variables that take values in a countable set $E \subset \mathbb{R}$. Let $c \in \mathbb{R}$ and define the sets $A \coloneqq E \cap (-\infty, c)$, $B \coloneqq E \cap [c, \infty)$. Assume that $A, B \neq \emptyset$ and*

$$\begin{aligned}\mathbb{P}\left\{X = x\right\} &> \mathbb{P}\left\{Y = x\right\}, \quad x \in A, \\ \mathbb{P}\left\{X = x\right\} &< \mathbb{P}\left\{Y = x\right\}, \quad x \in B.\end{aligned}$$

Then $\mathbb{E}\, Y > \mathbb{E}\, X$ holds.

Proof. We define the constant

$$t \coloneqq \mathbb{P}\left\{X \in A\right\} - \mathbb{P}\left\{Y \in A\right\}. \tag{25}$$

The constant t can also be expressed as

$$t = 1 - \mathbb{P}\left\{X \in B\right\} - \left(1 - \mathbb{P}\left\{Y \in B\right\}\right) = \mathbb{P}\left\{Y \in B\right\} - \mathbb{P}\left\{X \in B\right\}. \tag{26}$$

We write

$$\begin{aligned}\mathbb{E}\, Y - \mathbb{E}\, X &= \sum_{x \in E} x \left(\mathbb{P}\left\{Y = x\right\} - \mathbb{P}\left\{X = x\right\}\right) \\ &= \sum_{x \in A} x \left(\mathbb{P}\left\{Y = x\right\} - \mathbb{P}\left\{X = x\right\}\right) + \sum_{x \in B} x \left(\mathbb{P}\left\{Y = x\right\} - \mathbb{P}\left\{X = x\right\}\right). \qquad (27)\end{aligned}$$

For all $x \in A$, $x < c$ and $\mathbb{P}\left\{Y = x\right\} - \mathbb{P}\left\{X = x\right\} < 0$ hold. Thus, for the first summand in (27), we have the lower bound

$$\begin{aligned}\sum_{x \in A} x \left(\mathbb{P}\left\{Y = x\right\} - \mathbb{P}\left\{X = x\right\}\right) &> c \sum_{x \in A} \left(\mathbb{P}\left\{Y = x\right\} - \mathbb{P}\left\{X = x\right\}\right) \\ &= c \left(\mathbb{P}\left\{Y \in A\right\} - \mathbb{P}\left\{X \in A\right\}\right) \\ &= -ct, \qquad (28)\end{aligned}$$

where in the last step we used (25).

For all $x \in B$, $x \geq c$ and $\mathbb{P}\{Y = x\} - \mathbb{P}\{X = x\} > 0$ hold. Thus, for the second summand in (27), we have the bound

$$\begin{aligned}\sum_{x \in B} x\left(\mathbb{P}\{Y = x\} - \mathbb{P}\{X = x\}\right) &\geq c \sum_{x \in B}\left(\mathbb{P}\{Y = x\} - \mathbb{P}\{X = x\}\right) \\ &= c\left(\mathbb{P}\{Y \in B\} - \mathbb{P}\{X \in B\}\right) \\ &= ct, \end{aligned} \tag{29}$$

where in the last step we used (26).

Combining the lower bounds in (28) and (29) yields the claim due to (27). □

Proposition 47. *For fixed $N \in \mathbb{N}$, the function $\beta \in \mathbb{R} \mapsto \mathbb{E}_{\beta,N}|S| \in \mathbb{R}$ is strictly increasing and continuous.*

Proof. The continuity is immediate. We show that $\beta \mapsto \mathbb{E}_{\beta,N}|S|$ is strictly increasing. For this purpose, we calculate the derivative of $p_\beta(x)/Z_{\beta,N}$ for any $x \in \Omega_N$ employing Lemma 51. Recall that $p_\beta(x)$ equals $\exp\left(\frac{\beta}{2N}\left(\sum_{i=1}^N x_i\right)^2\right)$ and $s(x)$ stands for $\sum_{i=1}^N x_i$ for all $x \in \Omega_N$.

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\beta}\left(\frac{p_\beta(x)}{Z_{\beta,N}}\right) &= \frac{\frac{\mathrm{d}p_\beta(x)}{\mathrm{d}\beta} Z_{\beta,N} - p_\beta(x)\frac{\mathrm{d}Z_{\beta,N}}{\mathrm{d}\beta}}{Z_{\beta,N}^2} \\ &= \frac{\frac{s(x)^2}{2N} p_\beta(x)}{Z_{\beta,N}} - \frac{p_\beta(x)\frac{Z_{\beta,N}}{2N}\mathbb{E}_{\beta,N}S^2}{Z_{\beta,N}^2} \\ &= \frac{p_\beta(x)}{2N Z_{\beta,N}}\left(s(x)^2 - \mathbb{E}_{\beta,N}S^2\right). \end{aligned} \tag{30}$$

We note that the derivative is positive if and only if

$$s(x)^2 > \mathbb{E}_{\beta,N}S^2. \tag{31}$$

We define the set

$$G := \left\{\ell^2 \,|\, \ell \in \mathbb{N},\ \ell = N \bmod 2,\ m < \ell \leq N\right\},$$

and let

$$G = \left\{g_1, \ldots, g_{|G|}\right\}$$

be an enumeration of G in ascending order. By Proposition 20, $\beta \in \mathbb{R} \mapsto \mathbb{E}_{\beta,N}S^2 \in \left(\kappa, N^2\right)$ is continuous and strictly increasing. Hence, the function is injective. By Lemma 18, it is surjective.

We define the constants $b_i \in \mathbb{R} \cup \{\infty\}$, $i \in \mathbb{N}_{|G|}$, by the condition

$$\mathbb{E}_{b_i,N}S^2 = g_i.$$

Note that $(b_i)_{i \in \mathbb{N}_{|G|}}$ is a strictly increasing (since $(g_i)_{i \in \mathbb{N}_{|G|}}$ is strictly increasing) finite sequence, and $b_{|G|} = \infty$ due to $g_{|G|} = N^2$ and $\lim_{\beta \to \infty} \mathbb{E}_{\beta,N}S^2 = N^2$ by Lemma 18. Using these constants, we define the sets

$$\begin{aligned} B_1 &:= (-\infty, b_1), \\ B_i &:= [b_{i-1}, b_i), \quad i \in \mathbb{N}_{|G|} \setminus \{1\}. \end{aligned}$$

Due to the bijectivity of $\beta \in \mathbb{R} \mapsto \mathbb{E}_{\beta,N} S^2 \in \left(\kappa, N^2\right)$, $B_1, \ldots, B_{|G|}$ is a partition of $\mathbb{R}$.

With these preparations, we first show that for all $\beta_1 < \beta_2$ with $\beta_1, \beta_2 \in B_i$ for some $i \in \mathbb{N}_{|G|}$, $\mathbb{E}_{\beta_1,N} |S| < \mathbb{E}_{\beta_2,N} |S|$ holds. We define the following subsets of Ω_N:

$$A_r := \left\{ x \in \Omega_N \;\middle|\; s(x)^2 \leq \mathbb{E}_{\beta_r,N} S^2 \right\}, \quad r = 1, 2,$$

and write A^c for the complement of any subset A of Ω_N. By the definition of the set G and the sets B_j, $j \in \mathbb{N}_{|G|}$, the equality $A_1 = A_2$ is satisfied. We set

$$A := A_1.$$

We use the derivatives of $p_\beta(x)/Z_{\beta,N}$ in (30) and the positivity condition (31) for said derivatives.

Since $\beta \in \mathbb{R} \mapsto \mathbb{E}_{\beta,N} S^2 \in \left(m, N^2\right)$ is strictly increasing, the sign of the derivative of $p_\beta(\cdot)/Z_{\beta,N}$, for any $\beta \in (\beta_1, \beta_2)$ and any $x \in \Omega_N$ is

$$\begin{aligned}
\frac{p_\beta(x)}{2N Z_{\beta,N}} \left(s(x)^2 - \mathbb{E}_{\beta,N} S^2 \right) &< 0, \quad x \in A, \\
\frac{p_\beta(x)}{2N Z_{\beta,N}} \left(s(x)^2 - \mathbb{E}_{\beta,N} S^2 \right) &> 0, \quad x \in A^c.
\end{aligned}$$

These signs and the fundamental theorem of calculus (note that the derivatives of $p_\beta(x)/Z_{\beta,N}$ are continuous) yield for all $x \in A$

$$\frac{p_{\beta_2}(x)}{Z_{\beta_2,N}} = \frac{p_{\beta_1}(x)}{Z_{\beta_1,N}} + \int_{\beta_1}^{\beta_2} \frac{\mathrm{d}}{\mathrm{d}\beta} \left(\frac{p_\beta(x)}{Z_{\beta,N}} \right) \mathrm{d}\beta < \frac{p_{\beta_1}(x)}{Z_{\beta_1,N}}$$

and for all $x \in A^c$

$$\frac{p_{\beta_2}(x)}{Z_{\beta_2,N}} = \frac{p_{\beta_1}(x)}{Z_{\beta_1,N}} + \int_{\beta_1}^{\beta_2} \frac{\mathrm{d}}{\mathrm{d}\beta} \left(\frac{p_\beta(x)}{Z_{\beta,N}} \right) \mathrm{d}\beta > \frac{p_{\beta_1}(x)}{Z_{\beta_1,N}}.$$

We apply Lemma 46 with X being $|S|$ following the distribution $\mathbb{P}_{\beta_1,N} \circ |S|^{-1}$ and Y being $|S|$ following the distribution $\mathbb{P}_{\beta_2,N} \circ |S|^{-1}$. Then the statement $\mathbb{E}_{\beta_1,N} |S| < \mathbb{E}_{\beta_2,N} |S|$ follows by Lemma 46.

Now assume that $\beta_1, \beta_2 \in \mathbb{R}$, $\beta_1 < \beta_2$, and $\beta_1 \in B_{j_1}$, $\beta_2 \in B_{j_2}$ for some $j_1 < j_2 \in \mathbb{N}_{|G|}$. Since $\beta \mapsto \mathbb{E}_{\beta,N} |S|$ is continuous and strictly increasing on each B_i, $i \in \mathbb{N}_{|G|}$, we obtain for all $\beta \in B_i$

$$\mathbb{E}_{b_i,N} |S| = \lim_{\beta' \nearrow b_i} \mathbb{E}_{\beta',N} |S| > \mathbb{E}_{\beta,N} |S|$$

and

$$\mathbb{E}_{b_{i+1},N} |S| = g_{i+1} > g_i = \mathbb{E}_{b_i,N} |S|.$$

Taking into account $\beta_1 \in B_{j_1} = [b_{j_1-1}, b_{j_1})$, $\beta_2 \in [b_{j_2-1}, b_{j_2})$, and $j_2 - 1 \geq j_1$, we therefore have

$$\mathbb{E}_{\beta_2,N} |S| \geq \mathbb{E}_{b_{j_2-1},N} |S| \geq \mathbb{E}_{b_{j_1},N} |S| > \mathbb{E}_{\beta_1,N} |S|,$$

and we have proved that the function $\beta \mapsto \mathbb{E}_{\beta,N} |S|$ is strictly increasing. $\square$

By Propositions 45 and 47, we have

Corollary 48. *Let $\lambda, \mu \in \mathbb{N}_M$, $N_\lambda = N_\mu$, and $0 \leq \beta_\lambda < \beta_\mu$. Then the optimal weights w_λ and w_μ satisfy the inequality*

$$w_\lambda < w_\mu.$$

This implies that among groups of equal sizes, those groups with stronger interactions between voters, that is groups with a higher parameter β, will receive a larger weight in the council than groups with voters who interact more loosely with each other, i.e. groups that that have a lower parameter β. Put in a different way, under the CWM as a voting model, there are two avenues for groups to obtain a higher voting weight under the democracy deficit criterion: have a larger population or be more cohesive in the group votes.

Proposition 45 yields an estimator for the optimal weights of each group by substituting any estimator for the parameter β_λ. For example, an estimator for the optimal weights based on the estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}$ from Definition 7 can be defined as follows. Let for all $\lambda \in \mathbb{N}_M$, $\hat{\beta}_{N_\lambda}(\lambda) := \left(\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}\right)_\lambda$. Then, $\hat{\beta}_{N_\lambda}(\lambda) : \Omega_{N_\lambda}^n \to [-\infty, \infty]$ is an estimator for β_λ calculated for a sample of observations of voting in group λ.

Definition 49. Let $\lambda \in \mathbb{N}_M$, $\beta_\lambda \geq 0$. The estimator $\hat{w}_\lambda : \Omega_{N_\lambda}^n \to [0, \infty)$ for the optimal weight of group λ based on $\hat{\beta}_{N_\lambda}(\lambda)$ is defined by

$$\hat{w}_\lambda = \mathbb{E}_{\hat{\beta}_{N_\lambda}(\lambda), N_\lambda} |S_\lambda| \,.$$

The estimator $\hat{w}_\lambda$ inherits many of the properties of $\hat{\boldsymbol{\beta}}_{\boldsymbol{N}}$. This is the subject of the next theorem. Recall the good rate function $\boldsymbol{J}$ defined in (19), and the single-group rate function $J_\lambda : [-\infty, \infty] \to [0, \infty]$ in (18) for group λ. We define the good rate function $H_\lambda : \mathbb{R} \to [0, \infty]$ by

$$H_\lambda(y) := \inf \left\{ J_\lambda(\beta) \mid \beta \in [-\infty, \infty], \mathbb{E}_{\beta, N_\lambda} |S_\lambda| = y \right\}, \quad y \in \mathbb{R}.$$

Theorem 50. *Let $\lambda \in \mathbb{N}_M$ and $N_\lambda \in \mathbb{N}$. Let w_λ be the optimal weight from Proposition 45. Then the following statements hold:*

1. $\hat{w}_\lambda \xrightarrow[n \to \infty]{\mathrm{p}} w_\lambda$.
2. $\sqrt{n}\left(\hat{w}_\lambda - w_\lambda\right) \xrightarrow[n \to \infty]{\mathrm{d}} \mathcal{N}\left(0, v^2\right)$ *and*

$$v^2 = \frac{\left(\mathbb{E}_{\beta, N_\lambda} |S_\lambda|^3 - \mathbb{E}_{\beta, N_\lambda} |S_\lambda| \, \mathbb{E}_{\beta, N_\lambda} S_\lambda^2\right)^2}{\mathbb{V}_{\beta, N} S^2}.$$

3. *$\hat{w}_\lambda$ satisfies a large deviations principle with rate n and rate function H_λ. H_λ has a unique minimum at w_λ, and we have for each closed set $K \subset \mathbb{R}$ that does not contain w_λ, $\inf_{y \in K} H_\lambda(y) > 0$ and*

$$\mathbb{P}\left\{\hat{w}_N \in K\right\} \leq 2 \exp\left(-n \inf_{y \in K} H_\lambda(y)\right).$$

Proof. This theorem is proved in close analogy to Theorem 10. □

Appendix

We present a number of concepts and auxiliary results we use. Recall the expression $Z_{\boldsymbol{\beta}, \boldsymbol{N}}$ from (2).

Lemma 51. *The first derivative of $Z_{\boldsymbol{\beta},\boldsymbol{N}}$ with respect to β_λ, $\lambda \in \mathbb{N}_M$ is*

$$\frac{\mathrm{d}Z_{\boldsymbol{\beta},\boldsymbol{N}}}{\mathrm{d}\beta_\lambda} = \frac{Z_{\boldsymbol{\beta},\boldsymbol{N}}}{2N_\lambda}\,\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{N}} S_\lambda^2.$$

Proof. The derivative of the partition function $Z_{\boldsymbol{\beta},\boldsymbol{N}}$ with respect to β_λ is

$$\begin{aligned}
\frac{\mathrm{d}Z_{\boldsymbol{\beta},\boldsymbol{N}}}{\mathrm{d}\beta_\lambda} &= \sum_{x \in \Omega_{N_1+\cdots+N_M}} \frac{\mathrm{d}}{\mathrm{d}\beta_\lambda}\left[\exp\left(\frac{1}{2}\sum_{\lambda=1}^{M}\frac{\beta_\lambda}{N_\lambda}\left(\sum_{i=1}^{N_\lambda} x_{\lambda i}\right)^2\right)\right] \\
&= \sum_{x \in \Omega_{N_1+\cdots+N_M}} \frac{1}{2N_\lambda}\left(\sum_{i=1}^{N_\lambda} x_{\lambda i}\right)^2 \exp\left(\frac{1}{2}\sum_{\lambda=1}^{M}\frac{\beta_\lambda}{N_\lambda}\left(\sum_{i=1}^{N_\lambda} x_{\lambda i}\right)^2\right) \\
&= \frac{Z_{\boldsymbol{\beta},\boldsymbol{N}}}{2N_\lambda}\,\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{N}} S_\lambda^2.
\end{aligned}$$

□

Theorem 52 (Slutsky)**.** *Let $(Y_n)_{n\in\mathbb{N}}$ and $(Z_n)_{n\in\mathbb{N}}$ be sequences of random variables, Y a random variable, and $a \in \mathbb{R}$ a constant such that $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$ and $Z_n \xrightarrow[n\to\infty]{\mathrm{p}} a$. Then*

$$Y_n + Z_n \xrightarrow[n\to\infty]{\mathrm{d}} Y + a \quad \text{and} \quad Y_n Z_n \xrightarrow[n\to\infty]{\mathrm{d}} aY.$$

Proof. These statements are Theorems 11.3 and 11.4 in [14]. □

Theorem 53 (Continuous Mapping)**.** *Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of random variables and Y a random variable, each of them taking values in some subset $A \subset \mathbb{R}$, such that $Y_n \xrightarrow[n\to\infty]{\mathrm{p}} Y$, and let $g : A \to \mathbb{R}$ be a continuous function. Then*

$$g(Y_n) \xrightarrow[n\to\infty]{\mathrm{p}} g(Y).$$

Proof. See Theorem 2.3 in [32]. □

Theorem 54 (Delta Method)**.** *Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of random variables such that $\mathbb{E}\,Y_n = \mu \in \mathbb{R}$ for all $n \in \mathbb{N}$ and $\sqrt{n}\,(Y_n - \mu) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \sigma^2\right)$ for a constant $\sigma > 0$. Let $f : D \to \mathbb{R}$ be a continuously differentiable function with domain $D \subset \mathbb{R}$ such that $Y_n \in D$ for all $n \in \mathbb{N}$. Assume $f'(\mu) \neq 0$. Then*

$$\sqrt{n}\,(f(Y_n) - f(\mu)) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, (f'(\mu))^2\,\sigma^2\right)$$

is satisfied.

Proof. Since this result is not found as frequently in textbooks, we present a proof here for the convenience of the readers.

We Taylor expand the function f around the point μ:

$$f(Y_n) = f(\mu) + (Y_n - \mu)\, f'(\xi_n)$$

for some ξ_n which lies between μ and Y_n. We can rewrite the above as

$$\sqrt{n}\left(f\left(Y_n\right)-f\left(\mu\right)\right)=\sqrt{n}\left(Y_n-\mu\right)f'\left(\xi_n\right). \tag{32}$$

Due to the assumption $\sqrt{n}\left(Y_n-\mu\right)\xrightarrow[n\to\infty]{\mathrm{d}}\mathcal{N}\left(0,\sigma^2\right)$,

$$\left|\xi_n-Y_n\right|\leq\left|\mu-Y_n\right|\xrightarrow[n\to\infty]{\mathrm{p}}0,$$

and as f' is continuous, Theorem 53 implies

$$f'\left(\xi_n\right)\xrightarrow[n\to\infty]{\mathrm{p}}f'\left(\mu\right).$$

The last display, $\sqrt{n}\left(Y_n-\mu\right)\xrightarrow[n\to\infty]{\mathrm{d}}\mathcal{N}\left(0,\sigma^2\right)$, (32), and Theorem 52 together yield

$$\sqrt{n}\left(f\left(Y_n\right)-f\left(\mu\right)\right)\xrightarrow[n\to\infty]{\mathrm{d}}\mathcal{N}\left(0,\left(f'\left(\mu\right)\right)^2\sigma^2\right).$$

□

Recall Notation 6 for the expressions $[-\infty,\infty]$ and $[0,\infty]$.

Definition 55. Let $\left(P_n\right)_{n\in\mathbb{N}}$ be a sequence of probability measures on a metric space $\mathcal{X}$, let $\left(a_n\right)_{n\in\mathbb{N}}$ be a sequence of positive numbers with $a_n\xrightarrow[n\to\infty]{}\infty$, and let $I:\mathcal{X}\to[0,\infty]$ be a function. If I is lower semi-continuous, i.e. its level sets $\{x\in\mathcal{X}\,|\,I(x)\leq\alpha\}$ are closed for each $\alpha\in[0,\infty)$, we call I a rate function. If the level sets are compact in $\mathcal{X}$ for each $\alpha\in[0,\infty)$, we call I a good rate function. If I is a good rate function, and the two conditions

1. $\limsup_{n\to\infty}\frac{1}{a_n}\ln P_nK\leq-\inf_{x\in K}I(x)$ for each closed set $K\subset\mathcal{X}$,
2. $\liminf_{n\to\infty}\frac{1}{a_n}\ln P_nG\geq-\inf_{x\in G}I(x)$ for each open set $G\subset\mathcal{X}$

hold, then we say that the sequence $\left(P_n\right)_{n\in\mathbb{N}}$ satisfies a large deviations principle with rate a_n and rate function I. If $\left(Y_n\right)_{n\in\mathbb{N}}$ is a sequence of random variables taking values in $\mathcal{X}$ such that, for each $n\in\mathbb{N}$, Y_n follows the distribution P_n, we will also say that $\left(Y_n\right)_{n\in\mathbb{N}}$ satisfies a large deviations principle with rate a_n and rate function I.

In our applications of large deviations principles, the metric space $\mathcal{X}$ will be $\mathbb{R}$ or $[-\infty,\infty]$.

We present a standard convergence result concerning a statistic employed in the estimation of the parameter β. This can be used to then demonstrate convergence to β for the estimators presented in this article. Recall Definition 24 of the entropy function of a distribution.

Proposition 56. *Let $n,N\in\mathbb{N}$ and let $R:\Omega_N^n\to\mathbb{R}$ be a statistic of the form*

$$R\left(x^{(1)},\ldots,x^{(n)}\right):=\frac{1}{n}\sum_{t=1}^{n}f\left(x^{(t)}\right),\quad\left(x^{(1)},\ldots,x^{(n)}\right)\in\Omega_N^n,$$

for some function $f:\Omega_N\to\mathbb{R}$. Let X be a random vector on Ω_N with Curie-Weiss distribution according to Definition 1. Let $\mu:=\mathbb{E}_{\beta,N}f(X)$, $\sigma^2:=\mathbb{V}_{\beta,N}f(X)$, and $\Lambda^*_{f(X)}$ the entropy function of $f(X)$.

Then the following three statements hold:

1. *A law of large numbers holds for the sequence* $R\left(x^{(1)},\ldots,x^{(n)}\right)$*:*

$$R\left(x^{(1)},\ldots,x^{(n)}\right)\xrightarrow[n\to\infty]{\mathrm{p}}\mu.$$

2. *A central limit theorem holds for the sequence* $\sqrt{n}\left(R\left(x^{(1)},\ldots,x^{(n)}\right)-\mu\right)$*:*

$$\sqrt{n}\left(R\left(x^{(1)},\ldots,x^{(n)}\right)-\mu\right)\xrightarrow[n\to\infty]{\mathrm{d}}\mathcal{N}\left(0,\sigma^2\right).$$

3. *A large deviations principle holds for the sequence* $R\left(x^{(1)},\ldots,x^{(n)}\right)$ *with rate* n *and rate function* $I:\mathbb{R}\to[0,\infty)\cup\{\infty\}$*,*

$$I\left(x\right):=\Lambda^*_{f(X)}\left(x\right),\quad x\in\mathbb{R}.$$

4. *If* I *has a unique global minimum at* $x_0\in\mathbb{R}$*, then for any closed set* $K\subset\mathbb{R}$ *that does not contain* x_0 *we have* $\delta:=\inf_{x\in K}I\left(x\right)>0$*, and*

$$\mathbb{P}\left\{R\in K\right\}\leq 2\exp\left(-\delta n\right)$$

holds for all $n\in\mathbb{N}$*.*

Proof. Since R is defined as a sum of i.i.d. random variables with existing variance, the first three results can be found in many books about probability theory, and we therefore omit their proof. The last statement is somewhat less commonly found, hence we prove it here.

The random variable $f\left(X\right)$ is bounded and not almost surely constant, so Lemma 26 applies to $\Lambda^*_{f(X)}$.

Set

$$\delta:=\inf\left\{\Lambda^*_{f(X)}\left(x\right)\mid x\in K\right\}.\tag{33}$$

As $\mathbb{E}_{\beta,N}f\left(X\right)=\mu\in K^c$ and K^c is open, there is some $\eta>0$ such that the open ball $B_\eta\left(\mu\right)$ with radius η and centre μ is a subset of K^c. We choose $\varepsilon_r:=\sup\left\{\eta>0\mid(\mu,\mu+\eta)\subset K^c\right\}$ and $\varepsilon_l:=\sup\left\{\eta>0\mid(\mu-\eta,\mu)\subset K^c\right\}$. Then $G:=(\mu-\varepsilon_l,\mu+\varepsilon_r)\subset K^c$. By statements 4 and 5 of Lemma 26,

$$\delta=\inf\left\{\Lambda^*_{f(X)}\left(x\right)\mid x\in G^c\right\}=\min\left\{\Lambda^*_{f(X)}\left(\mu-\varepsilon_l\right),\Lambda^*_{f(X)}\left(\mu+\varepsilon_r\right)\right\}>0.$$

We write

$$\mathbb{P}\left\{R\in K\right\}\leq\mathbb{P}\left\{R\in G^c\right\}=\mathbb{P}\left\{R\leq\mu-\varepsilon_l\right\}+\mathbb{P}\left\{R\geq\mu+\varepsilon_r\right\}.$$

We use Markov's inequality to obtain for all $t\leq 0$

$$\begin{aligned}\mathbb{P}\left\{R\leq\mu-\varepsilon_l\right\}&\leq\mathbb{P}\left\{\exp\left(nt\left(R-(\mu-\varepsilon_l)\right)\right)\geq 1\right\}\leq\mathbb{E}\exp\left(nt\left(R-(\mu-\varepsilon_l)\right)\right)\\&=\exp\left(-nt\left(\mu-\varepsilon_l\right)\right)\prod_{r=1}^{n}\mathbb{E}\exp\left(tf\left(X^{(r)}\right)\right)=\exp\left(-nt\left(\mu-\varepsilon_l\right)\right)\left[\mathbb{E}\exp\left(tf\left(X\right)\right)\right]^n\\&=\exp\left(-nt\left(\mu-\varepsilon_l\right)\right)\exp\left(n\Lambda_{f(X)}\left(t\right)\right)=\exp\left(-n\left(\left(\mu-\varepsilon_l\right)t-\Lambda_{f(X)}\left(t\right)\right)\right).\end{aligned}$$

As this holds for all $t\leq 0$ and we have $\mu-\varepsilon_l<\mu=\mathbb{E}_{\beta,N}f\left(X\right)$, we arrive at

$$\mathbb{P}\left\{R\leq\mu-\varepsilon_l\right\}\leq\exp\left(-n\Lambda^*_{f(X)}\left(\mu-\varepsilon_l\right)\right).\tag{34}$$

Similarly, we calculate the upper bound

$$\mathbb{P}\left\{R \geq \mu + \varepsilon_r\right\} \leq \exp\left(-n\Lambda^*_{f(X)}\left(\mu + \varepsilon_r\right)\right). \tag{35}$$

Combining (34) and (35) yields

$$\begin{aligned}\mathbb{P}\left\{R \in K\right\} &\leq \exp\left(-n\Lambda^*_{f(X)}\left(\mu - \varepsilon_l\right)\right) + \exp\left(-n\Lambda^*_{f(X)}\left(\mu + \varepsilon_r\right)\right)\\ &\leq 2\exp\left(-\delta n\right).\end{aligned}$$

□

Theorem 57 (Contraction Principle)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be metric spaces. Let $\left(P_n\right)_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathcal{X}$, and let $f : D \to \mathcal{Y}$ be a continuous function with its domain $D \subset \mathcal{X}$ containing the support of P_n for each $n \in \mathbb{N}$. Let $\left(a_n\right)_{n\in\mathbb{N}}$ a sequence of positive numbers with $a_n \xrightarrow[n\to\infty]{} \infty$, and $I : \mathcal{X} \to [0,\infty]$ a good rate function. We define $J : \mathcal{Y} \to [0,\infty]$ by*

$$J\left(y\right) \coloneqq \inf\left\{I\left(x\right) \mid x \in D, f\left(x\right) = y\right\}, \quad y \in \mathcal{Y}.$$

Then J is a good rate function. If $\left(P_n\right)_{n\in\mathbb{N}}$ satisfies a large deviations principle with rate a_n and rate function I, then the sequence of push forward measures $\left(P_n \circ f^{-1}\right)_{n\in\mathbb{N}}$ on $\mathcal{Y}$ satisfies a large deviations principle with rate a_n and rate function J.

Proof. This result can be found, e.g., in [8, Theorem 4.2.1]. □

Acknowledgements

M. B. is a Fellow and G. T. is a Candidate of the Sistema Nacional de Investigadoras e Investigadores. G. T. was supported by a Secihti (formerly Conahcyt) postdoctoral fellowship.

References

[1] Miguel Ballesteros, Ramsés H. Mena, José Luis Pérez, and Gabor Toth. Detection of an arbitrary number of communities in a block spin ising model. 2023. arXiv:2311.18112.

[2] Claus Beisbart and Luc Bovens. Welfarist evaluations of decision rules for boards of representatives. *Soc. Choice Welf.*, 29(4):581–608, 2007.

[3] Quentin Berthet, Philippe Rigollet, and Piyush Srivastava. Exact recovery in the ising blockmodel. *Ann. Statist.*, 47(4):1805–1834, 2019.

[4] William A. Brock and Steven N. Durlauf. Discrete choice with social interactions. *Rev. Econ. Stud.*, 68(2):235–260, 2001.

[5] Sourav Chatterjee. Estimation in spin glasses: A first step. *Ann. Stat.*, 35(5):1931–1946, 2007.

[6] Wei-Kuo Chen, Arnab Sen, and Qiang Wu. Joint parameter estimations for spin glasses. 2024. arXiv:2406.10760.

[7] Pierluigi Contucci and Stefano Ghirlanda. Modelling society with statistical mechanics: An application to cultural contact and immigration. *Qual. Quant.*, 41:569–578, 2007.

[8] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications.* Springer New York, 2 edition, 1998.

[9] William F. Donoghue. *Distributions and Fourier Transforms.* Academic Press, 1969.

[10] Richard Ellis. *Entropy, Large Deviations, and Statistical Mechanics.* Whiley, 1985.

[11] Dan S. Felsenthal and Moshé Machover. *The Measurement of Voting Power. Theory and Practice, Problems and Paradoxes.* Edward Elgar Publishing, 1998.

[12] Dan S. Felsenthal and Moshé Machover. Minimizing the mean majority deficit: The second square-root rule. *Math. Soc. Sci.*, 37(1):25–37, 1999.

[13] Ignacio Gallo, Adriano Barra, and Pierluigi Contucci. Parameter evaluation of a simple mean-field model of social interaction. *Math. Models Methods Appl. Sci.*, 19(supp01):1427–1439, 2009.

[14] Allan Gut. *Probability: A Graduate Course.* Springer New York, 2 edition, 2013.

[15] Ernst Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31:253–258, 1925.

[16] Pieter Willem Kasteleijn. *Statistical Problems in Ferromagnetism, Antiferromagnetism and Adsorption.* 1956.

[17] Werner Kirsch. On penrose's square-root law and beyond. *Homo oecon.*, 24(3/4):357–380, 2007.

[18] Werner Kirsch and Jessica Langner. *The Fate of the Square Root Law for Correlated Voting*, chapter Voting Power and Procedures, pages 147–158. Springer Cham, 2014.

[19] Werner Kirsch and Gabor Toth. Two groups in a Curie-Weiss model. *Math. Phys. Anal. Geom.*, 23(17), 2020.

[20] Werner Kirsch and Gabor Toth. Two groups in a Curie-Weiss model with heterogeneous coupling. *J. Theor. Probab.*, 33:2001–2026, 2020.

[21] Werner Kirsch and Gabor Toth. Optimal weights in a two-tier voting system with mean-field voters. arXiv:2111.08636, November 2021.

[22] Werner Kirsch and Gabor Toth. Collective bias models in two-tier voting systems and the democracy deficit. *Math. Soc. Sci.*, 119:118–137, 2022.

[23] Werner Kirsch and Gabor Toth. Limit theorems for multi-group Curie-Weiss models via the method of moments. *Math. Phys. Anal. Geom.*, 25(24), 2022.

[24] Yukio Koriyama, Antonin Macé, Rafael Treibich, and Jean-François Laslier. Optimal apportionment. *J. Political Econ.*, 121(3):584–608, 2013.

[25] Sascha Kurz, Nicola Maaser, and Stefan Napel. On the democratic weights of nations. *J. Political Econ.*, 125(5):1599–1634, 2017.

[26] Matthias Löwe and Kristina Schubert. Exact recovery in block spin Ising models at the critical line. *Electron. J. Stat.*, 14:1796–1815, 2020.

[27] Matthias Löwe, Kristina Schubert, and Franck Vermet. Multi-group binary choice with social interaction and a random communication structure – a random graph approach. *Physica A*, 556, 2020.

[28] Nicola Maaser and Stefan Napel. A note on the direct democracy deficit in two-tier voting. *Math. Soc. Sci.*, 63(2):174–180, 2012.

[29] Lars Onsager. Crystal statistics I. a two-dimensional model with an order-disorder transition. *Phys. Rev.*, 65(3-4):117–149, 1944.

[30] Lionel Penrose. The elementary statistics of majority voting. *Journal of the Royal Statistical Society*, 109(1):53–57, 1946.

[31] Gabor Toth. *Correlated Voting in Multipopulation Models, Two-Tier Voting Systems, and the Democracy Deficit.* PhD Thesis, FernUniversität in Hagen, April 2020. doi:10.18445/20200505-103735-0.

[32] Adrianus Willem Van der Vaart. *Asymptotic Statistics.* Cambridge University Press, 1 edition, 1998.

[33] Pierre Weiss. L'hypothèse du champ moléculaire et la propriété ferromagnétique. *Journal de Physique Théorique et Appliquée*, 6(1):661–690, 1907.

[34] Karol Zyczkowski and Wojciech Slomczynski. *Square Root Voting System, Optimal Threshold and* π, chapter Voting Power and Procedures, pages 127–146. Springer Cham., 2014.