Report on the manuscript ECP2205-014

Probabilistic Voting Models with Varying Speeds of Correlation Decay
by Gabor Toth

This paper studies the asymptotic behavior of the voting margins of a multi-group population by means of a sequence of de Finetti measures. Depending on the speed of convergence of the latter, a variety of behaviors is observed. In particular, a subcritical, a critical and a supercritical regimes are identified. If the speed of convergence is sufficiently slow (subcritical & critical regimes), the details of the sequence of de Finetti measures affect the limit. On the contrary, if the speed of convergence is fast (supercritical regime), the limit is Gaussian and universal.
I found the idea of introducing multi-group probabilistic voting models interesting but current developments around it are not sufficient enough to be published. I will just point out a few things.

- Section 1 provides an overview on voting systems and reviews a couple of models adapted from statistical physics: the collective bias model (CBM) and the Curie-Weiss model (CWM). Moreover, the definition of the voting measure analyzed in the manuscript is given therein. Section 2 contains new results established by the author. Section 3 is devoted to the proofs. The techniques used are rather classical and not really that innovative.
- There is a crucial aspect that requires clarification: how to link properly $n \in \mathbb{N}$ with a subdivision $(n_1, \dots, n_M)$, with $n = \sum_{j=1}^{M} n_j$, so that it is clear what it means to take the limit as $n \to +\infty$ when working at the level of the subdivision. The proof of Theorem 5 might have an issue, as it is not clear to me how to deal with the sequences of sets $(A_n)_{n\in\mathbb{N}}$ and $(B_n)_{n\in\mathbb{N}}$, when moving from n to n.
- I find misleading the continuous referring to the CWM, as it does not fit the set up of the present manuscript. To improve clarity of the presentation, I suggest the author to collect all the results about the CWM in a dedicated introductory subsection where to discuss also the main differences with respect to the voting measures he is going to study.
- I find a little dissatisfying the fact that no example of a sequence of de Finetti measures satisfying the assumptions of the main theorems is given and that no direct application to voting systems is presented.

In summary, I feel in its current version the manuscript does not reach the standard of *Electronic Communications in Probability*. To follow a list of remarks.

Abstract

- L.7 – behaviours
- L.8 – limit, under fast convergence,
- L.9 – convergence

1 Introduction

- P.1 L.1 – After the very first sentence, references are needed.
- P.2 L.11 – ~~where~~ in the limit as
- P.2 L.34 – ~~Instead~~ Indeed
- P.2 L.48 – correlations
- P.3 L.16 – against of the proposal
- P.3 L.19 – The reference is probably wrong. I was not able to find out what you claim.

- P.3 Eq.(1.3) – Recall that the probability measure P_m is the same as in the previous example.
- P.3 LL.30-37 – The paragraph seems incomplete. In the first sentence, you mention three possible regimes for the CWM, but you in fact discuss only one of them.
- P.4 LL.9-11 – When taking the limit, do you keep the proportion of individuals belonging to each group fixed?
- P.4 L.16 – What do you mean precisely when you say "contrary to the static de Finetti measure μ"?
- P.4 Def.1 – As n is both the size of the whole population and the index of the sequences of probability measures $(\mathbb{P}_n)_{n\in\mathbb{N}}$ and $(\mu_n)_{n\in\mathbb{N}}$, you should either comment on how you relate any fixed index $n \in \mathbb{N}$ to a subdivision with groups of sizes $(n_1, \ldots, n_M)$, as it is not uniquely determined, if matters, or specify why it does not matter. This is relevant also in the sequel, when you define sequences of sets indexed by $n \in \mathbb{N}$.

  Why do you introduce $\bar{m}$? It is not used in the paper.
- P.4 L.27 – population~~s~~
- P.4 LL.28-31 – I do not see the point in making this comment.
- P.4 L.31 – The full stop is misplaced; move it after the round bracket.
- P.4 L.35 – distribution~~s~~

2 Results

- P.7 Thm.5 – Recall what ϕ is.

3 Proofs

- P.8 L.5 – $E_m \exp(i(t_1 S_{n,1}/\sqrt{n_1} + \cdots + t_M S_{n,M}/\sqrt{n_M}))$
- P.8 L.9 – Since you denote by i the imaginary unit, it is better to use j as summation index.

  Replace m_1 with m_λ.
- P.8 L.16 – In the definition of the set A_n, the subscript λ of the variable t is missing.
- P.8 L.17 – $\varepsilon_{n,\lambda} = o(1/\sqrt{n_\lambda})$
- P.8 L.18 – You should explain how it is possible to get the set B_n as range of $b_{n,\lambda}$, for any fixed $n \in \mathbb{N}$. How could you get rid of the remainder term? Why are you allowed to write exp?
- P.9 L.1 – $\exp(i\varepsilon_{n,\lambda} s_\lambda t_\lambda \sqrt{n_\lambda})$
- P.9 L.3 – ~~With~~ By
- P.9 L.13 – $E_{\varepsilon_n \circ s} \exp\left(it_\lambda \frac{S_{n,\lambda}}{\gamma_{n,\lambda}}\right)$
- P.10 L.13 – Reference [28] is not accessible to all readers. It would be better to refer to a Probability book written in English.
- P.10 L.15 – $\prod_{\lambda=1}^{M}$
- P.10 L.16 – $\int_{\mathbb{R}^M}$
- P.10 L.26 – How is the set A_n (resp. B_n) increased (resp. decreased) depending on n? As the subdivision $(n_1, \ldots, n_M)$ is not unique, these sets might vary wildly from n to n. It is not very clear to me that there are no issues in this proof. The estimate at P.11 LL.15,16 can be nonsense.
- P.11 L.1 – E_s
- P.11 L.21 – How should the symbol k appearing in the $O(\cdot)$ factor be interpreted? What is the role of δ in the proof?