

## I. IDENTIFICATION DATA

<b>Thesis title:</b>	<b>Problems of Maximum Likelihood Estimates.</b>
<b>Author's name:</b>	<b>Danil Alshaev.</b>
<b>Type of thesis :</b>	bachelor
<b>Faculty/Institute:</b>	Faculty of Electrical Engineering (FEE)
<b>Department:</b>	Department of Cybernetics.
<b>Thesis reviewer:</b>	Maxime Pietrantoni.
<b>Reviewer's department:</b>	CIIRC

## II. EVALUATION OF INDIVIDUAL CRITERIA

### Assignment

**challenging**

*How demanding was the assigned project?*

The assigned tasks are relatively challenging and require knowledge of estimation techniques and a good understanding of chess.

Overall, assigned tasks are open and leave some autonomy to the student to explore the topic in depth or not.

### Fulfilment of assignment

**fulfilled**

*How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.*

This thesis revolves around applying maximum likelihood estimators (MLE) to estimate score/rating in chess. Rating is estimated from a set of chess games and may use the averaged opponents rating where a closed form solution exists or use the set of opponent individual ratings which increases complexity and requires root finding algorithms. The student created small datasets from databases of chess games, applied both approaches, compared their validity/effectiveness (fulfilling item 2).

Through practical experiments and using confidence intervals, the student identifies statistical shortcomings associated with MLEs (no optimal properties with small samples) and shortcomings induced by the chess game itself (FIDE rating not reflecting a player's true strength, white/black discrepancy). It fulfills item 1. Examples where MLE yields accurate results are also included. Section 2.1 is too short to be used as supplementary material for education. Still, I would consider item 3 as fulfilled. Chapter 4 and 5 provide additional improvements and discussions.

### Methodology

**correct**

*Comment on the correctness of the approach and/or the solution methods.*

Modeling a chess game with two half contests using Bernoulli distribution is sound. MLE is correctly applied for sequences of such games, assumptions are clearly stated. To estimate the rating, the student derives a closed form solution when considering the aggregated opponents ratings and implements a root finding algorithm when considering multiple individual opponents ratings. Wilson confidence intervals are also rightly used to analyze the relevance in the context of small sample sizes. Chess related flaws in the rating estimation process are well identified, and the modifications made to the rating formula are coherent.

However, the validity of the estimator is only assessed by comparing the estimated rating with its confidence interval to the FIDE rating, which may not necessarily reveal the problems of MLEs. As pointed out in Sec. 3.5, comparing the estimated rating over a tournament to the FIDE rating mostly reveals if a player under-performed or over-performed in the tournament. Some other evaluation criterion may be necessary.

### Technical level

**B - very good.**

*Is the thesis technically sound? How well did the student employ expertise in the field of his/her field of study? Does the student explain clearly what he/she has done?*

On the technical level, the student uses MLE with associated confidence intervals and newton method.

Derivations are correct and the student explains his reasoning while detailing the steps. Throughout the thesis, these

techniques seem to be understood and mastered. Derivations revolving around the rating formula are correct, modifications brought up to the rating formula, while being based on intuition also make sense. Overall, I would have liked to see a more statistical perspective, eg. experiments related to properties of the MLE and the sample size as well as addressing more in depth the small dataset problem.

## Formal and language level, scope of thesis

**B - very good.**

*Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?*

The overall organization of the thesis is coherent and follows a logical progressive order: MLE applied to the chess rating estimation, identified shortcomings, further improvement and discussion. The English is grammatically correct without faults. The thesis is well-presented and covers the assignments while providing additional insights and improvements.

Overall, notations are well defined but for the following:

Section 2.1:  $x_i$  not defined -> realization of  $X_i$

Section 2.3.3:  $n$  is not defined -> number of games

Section 3.4.2:  $f$  not defined,  $R_p$  and  $R_A$  both refer to the same quantity?

## Selection of sources, citation correctness

**A - excellent.**

*Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?*

The study is an application paper and not a research paper, as such it doesn't require a related work section to position itself from existing work in the literature. Relevant prior work is cited for the technical parts of this thesis and chess related resource as well as pointers are also provided when necessary.

## Additional commentary and evaluation (optional)

*Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.*

Novelty and impact on the field are not relevant here as the thesis is not a research paper. Strength and weaknesses are summarized in the overall evaluation below.

## III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

*Summarize your opinion on the thesis and explain your final grading. Pose questions that should be answered during the presentation and defense of the student's work.*

*This thesis aimed at identifying shortcomings and limitations when applying maximum likelihood estimators to estimate chess score and ratings. This is an interesting and well written thesis, all associated assignments are fulfilled and technical aspects are mastered. The student demonstrated good understanding of chess and ability to translate it into mathematical formalism. However, some statistical aspects would need to be expanded upon and methodological shortcomings would need to be addressed (see previous comments). As such, I recommend the grade B – very good.*

### Open Questions:

What are the properties of MLEs?

Could you consider using a MAP estimator, what is the relation between ML and MAP estimators?

Does the IID assumption hold in a tournament?

The grade that I award for the thesis is B - very good.



## THESIS REVIEWER'S REPORT

Date: **03/06/2024**

Signature: