Are women worse at chess?
Answering stupid questions using numerical methods.

Ariday Bordon (aridaybordon@gmail.com)

In this paper, it has been studied how differences in the size of two populations extracted from the same distribution lead to differences in its maximum values. These results have been compared to the public FIDE data (dated on November 2020) in order to explain to some extent the difference in ELO rating between sexes.

I. INTRODUCTION

The supposed superiority of men above women in the world of chess has been stated numerous times by some of the most recognised players of the last few decades. Famous are the comments of former world champion G. Kasparov in 1989 defending that the competitiveness and creativity needed in chess are incompatible with female genes (although in recent interviews have taken back the vast majority of his past claims) and, more recently, this controversy has been awakened by N. Short that suggested again how biological differences may be the cause of the underperformance of women on chess and other intellectual domains.

During this past years, all this controversy has originated an increased interest in the topic and some literature have emerged (e.g. C. F. Chabris & M.E. Glickman [1, 2] and R. Howard [3]). The first paper that explained the observed difference in ELO rating as an expected statistical deviation due to the difference in the size of male and female population is due to M. Bilalic et al. [4] who stated that at least a 94 % of the difference can be explained in these terms. Besides, a short article by W. Ji [5] showed that how for Indian players, the distribution of ELO for both sexes is similar and the difference in rating expected due to the men-women ratio.

Following this line, we will show that a large amount of the difference in the performance between sexes, if not all, can be explained satisfactorily simply using statistical arguments regarding the population of chess players and the ratio of women within that population.

It is worth to mention that using the ELO rating to make global comparations between players may be tempting but not accurate, especially at the weaker levels and in small populations, due to the appearance of local trends. To avoid this problem, an analysis per country has been carried.

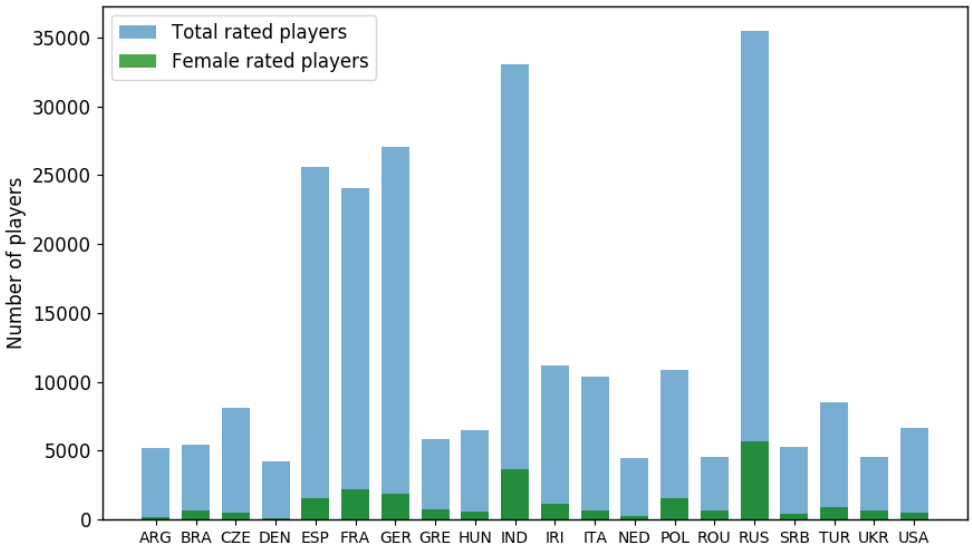
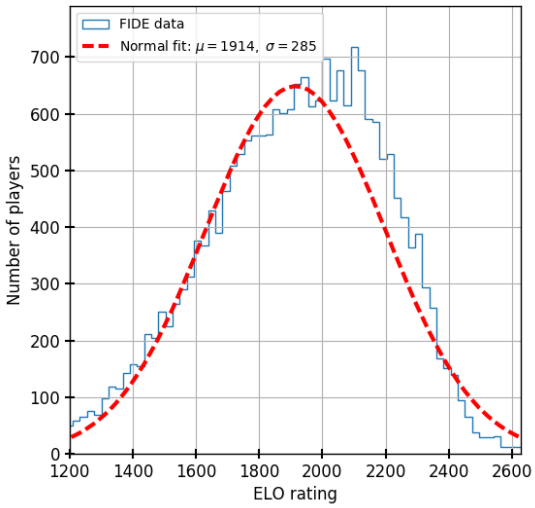


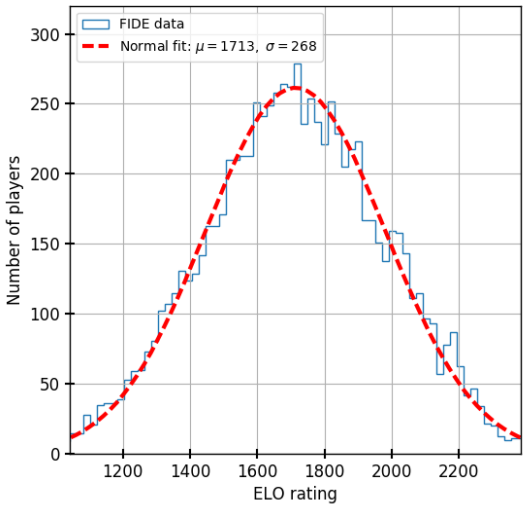
FIG. 1: Bar diagram of rated FIDE players for the first 20 countries with more players. There exist huge differences both in the total number of players and in the ratio of female players for different countries. The mean women-men ratio is around 10 %.

II. ANALYSIS

In this section, the method to study the differences in ELO will be detailed. For the sake of completeness, all the code used during this work may be found on our GitHub repository. Moreover, all the data have been taken from the official FIDE webpage (that can be found here). As in [5], the players born after the year 2000 have not been taking into consideration due to both the high variance in the rating of junior players and the inference that these players provoke in the ratio women-men. In order to avoid population with a small number of players that are subject to enormous variances, we have restringed all this analysis to the top 20 countries by number of players (see FIG. 1).



(a) Distribution of ELO ratings in Russia



(b) Distribution of ELO ratings in Italy

FIG. 2: ELO distribution for two different countries. To simulate the populations, the FIDE data have been fitted to a normal distribution.

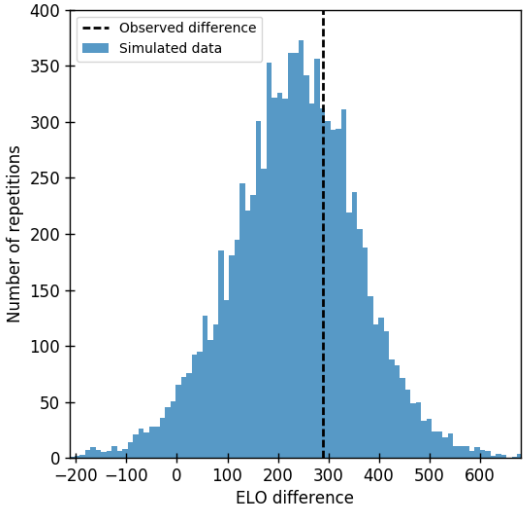


FIG. 3: Histogram of the simulated difference for 10000 repetitions beetween the best players in a men/women from Spain data. The observed difference have been plotted for comparison.

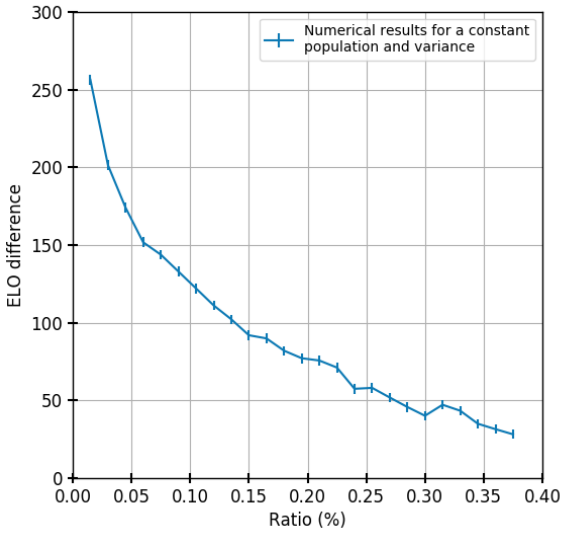
A major task during the elaboration of this paper has been to simulate different populations in order to obtain statistical information. To do so, the ELO distribution have been fitted to a normal distribution (see FIG. 2a and 2b) which allows us to describe any population in terms of three independent parameters, i.e. the size of the population and the mean and variance of the ELO distribution.

To be able to compare with the observed differences in ELO, the simulation have been repeated numerous times to form a histogram in which we compute the mean μ and the variance σ^2. If the observed difference in rating μ_{obs} enters within the range of expected values (i.e. if μ_{obs} is compatible with a standard 1σ deviation, $\mu \pm \sigma$), we will say that the difference can be explained as a simple statistical deviation.

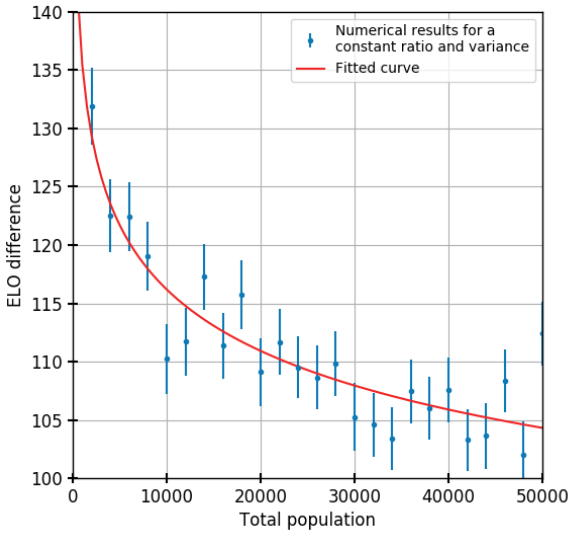
Finally, to compare the results from different countries, the observed difference has been scaled to the distribution obtained from the simulation (i.e. $\mu^N_{obs} = (\mu_{obs} - \mu)/\sigma^2$, where the superindex N refers to the normalized mean).

III. RESULTS

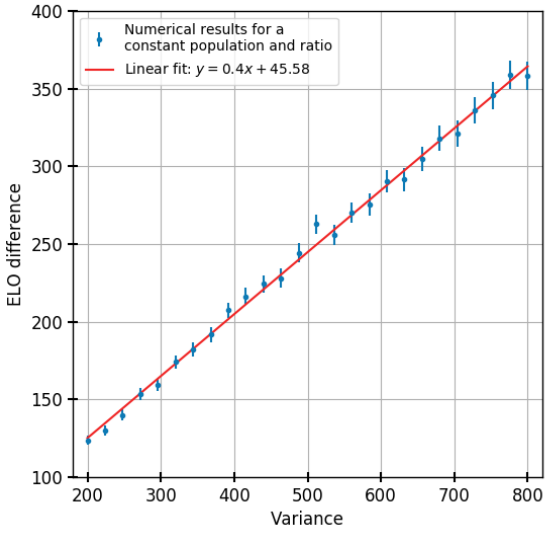
In FIG. 4, it has been plotted the dependence of the ELO gap as a function of the different para-



(a) ELO difference as a function of the women ratio



(b) ELO difference as a function of the total population



(c) ELO difference as a function of the variance of the distribution

FIG. 4: Rating difference for a simulated sample with mean $\mu = 1500$, variance $\sigma = 200$, total population $N = 5000$ and ratio women-men $r = 0.1$ as a function of the different parameters. As in [4], we obtain that the ELO gap decays with the logarithm of the population.

meters for a standard population (i.e. for a population of 5000 players in which a 10% of them are women). The existence of a difference in the ratio women-men leads naturally to an ELO difference between the two groups even is both populations come from the same distribution. The ELO deviation obtained using this procedure will be called the expected statistical deviation.

Finally in FIG. 5, the normalized comparison between the expected and observed difference have been plotted. Except for four countries, all the rest fall under the expected 1σ range of values. The mean normalized deviation is compatible with a null value, therefore there are no substantial differences beyond the expected statistical deviation.

IV. CONCLUSIONS

During the present work, a method to compare the differences in the performance of two different

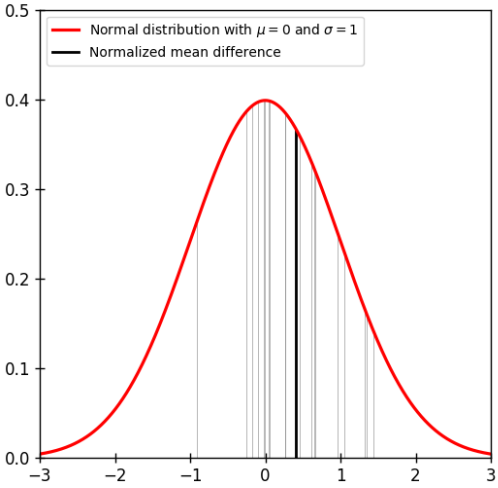


FIG. 5: Normalized difference between the expected and the actual performance of female players. The vertical grey lines show the normalized mean difference while the black one correspond to the mean of all the top 20 countries. The red line have been plotted for reference.

populations coming from the same normal distribution have been presented.

The dependence of this performance with respect the variance, ratio of women and size of the population for a fitted normal distribution with standard parameters have been analyzed recovering results obtained in previous works (see M. Bilalic et al. [5]).

Finally, all these expected statistical results have been compared to the official ELO rating provided by the FIDE database. Our computations show that the hypothesis of equality of gender (i.e. that both the ELO rating of men and women come from the same distribution) is compatible with our results.

Acknowledgements

I would like to acknowledge the outstanding effort of the Internacional Chess Federation in order of making available to the general public the data of all players with a FIDE rating. Its work makes possible studies like the one presented here.

[1] C. Chabris & M. Glickman, Psychol. Sci. 17, 1040–1046, *Sex Differences in Intellectual Performance Analysis of a Large Cohort of Competitive Chess Players*, (2006)

[2] M. Glickman, Appl. Stat. 48, 377-394, *Parameter estimation in large dynamic paired comparison experiments*, (1999)

[3] R. Howard, J. Biosoc. Sci. 37, 371–380, *Are gender differences in high achievement disappearing? A test in one intellectual domain*, (2005)

[4] M. Bilalic et al, Proc. Biol. Sci. 276(1659) 1161-5 *Why are (the best) women so good at chess? Participation rates and gender differences in intellectual domains*, (2009)

[5] W. Ji, *What gender gap in chess?*, (2020)