The International Correspondence Chess Federation

Ratings Commissioner
Gerhard Binder

Sauerlandstrasse 8
D-70794 Filderstadt
Germany
GBinder@iccf.com

Datum 04.07.2016

Report for the Congress at Bremen 2016

Dear friends,

since last year's Congress four ratinglists were published on time including special ratings for Chess960. Next I will work on the list 2016/4 which will be published on September, 15 and will be valid from 1st of October, business as usual.

In my report at Cardiff I announced the intention for a general review of the rating system, especially to consider too high expectancies in our high level tournaments, maybe with the help of a professional statistican. In addition the Qualification Commissioner and the Service Director urgently wished to replace the look-up tables 3 and 4 by a continuous formula (already proposed by Prof. Elo) which calculates the expected score p(D) as follows:

$p(D) = 1 / (1+ 10^{(-D/400)})$ and vice versa $D(p) = \log(p / (1-p)) * 400$

This is a very good approximation to the results of the look-up tables and can easyly be adapted by changing the value of 400.

I started with an investigation to find the real distribution of past results. Therefore I exported all results since 1999 excluding unsuitable results like double default, games with less than 11 moves and those games where one of the players (or both) had no startrating.

I classified and counted the results and scores according to our tables 3 and 4, lets say into groups with the same expectation value and divided into intervals of five years.

467761	Results remaining	234710	draws =	50,18%
134719	Postal games			
51658	Email games			
281384	Server games			
71053	Games started before year 2000	25821	draws =	36,34%
79845	Games started between 2000 and 2005	36886	draws =	46,20%
143828	Games started between 2006 and 2010	74544	draws =	51,83%
173035	Games started after 2010	97459	draws =	56,32%

The first surprising discovery was that there is no significant difference between the time-intervals. Of course the number of draws increased vehemently but the avarage score remained nearly unchanged. Therefore I continued with the sum of 1999 – 2015.

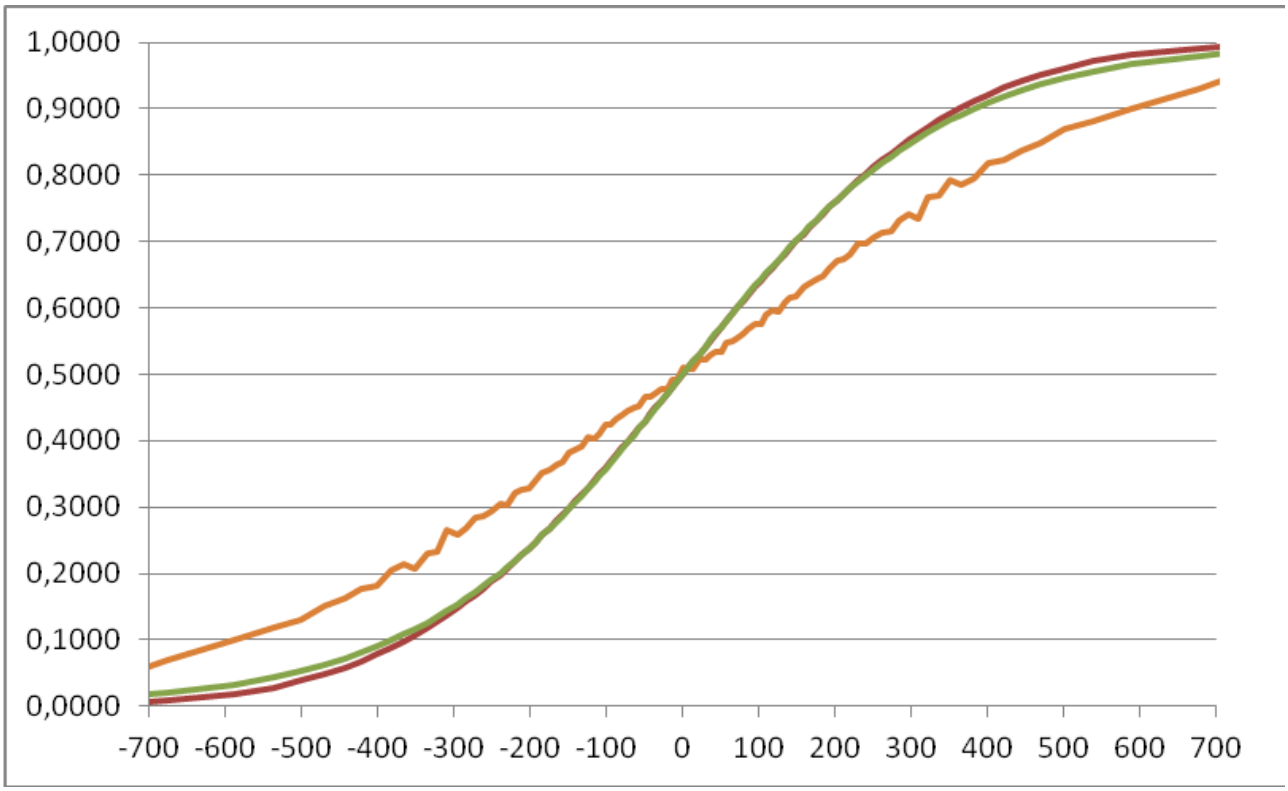
These data are shown in the charts below (orange curve)

Next step was to compare this „real“ distribution with the normal distribution and the standard deviation 200 * SQRT(2) which Elo was using - approximation formula P = 1/(1 + 10^(-D/400)). The first chart below shows this comparison. It becomes evident that our result expectation is and was always too high – the feeling of our players. As FIDE is still happy as far as I know with this distribution then this graph shows a difference between OTB and CC.

The curves of the normal and real distribution look very similar – what means that our results correspond inded with the normal distribution. Only the value of the standard deviation differs. I tried to find a value instead of 200 which fits our real distribution better. With a value of 320 we get a nearly identical picture. This can be seen in the second chart (green curve). This requires to change the continuous formula to P = 1/(1 + 10^(-D/640)).

The following chart shows the distribution of the avarge score depending on the rating difference.

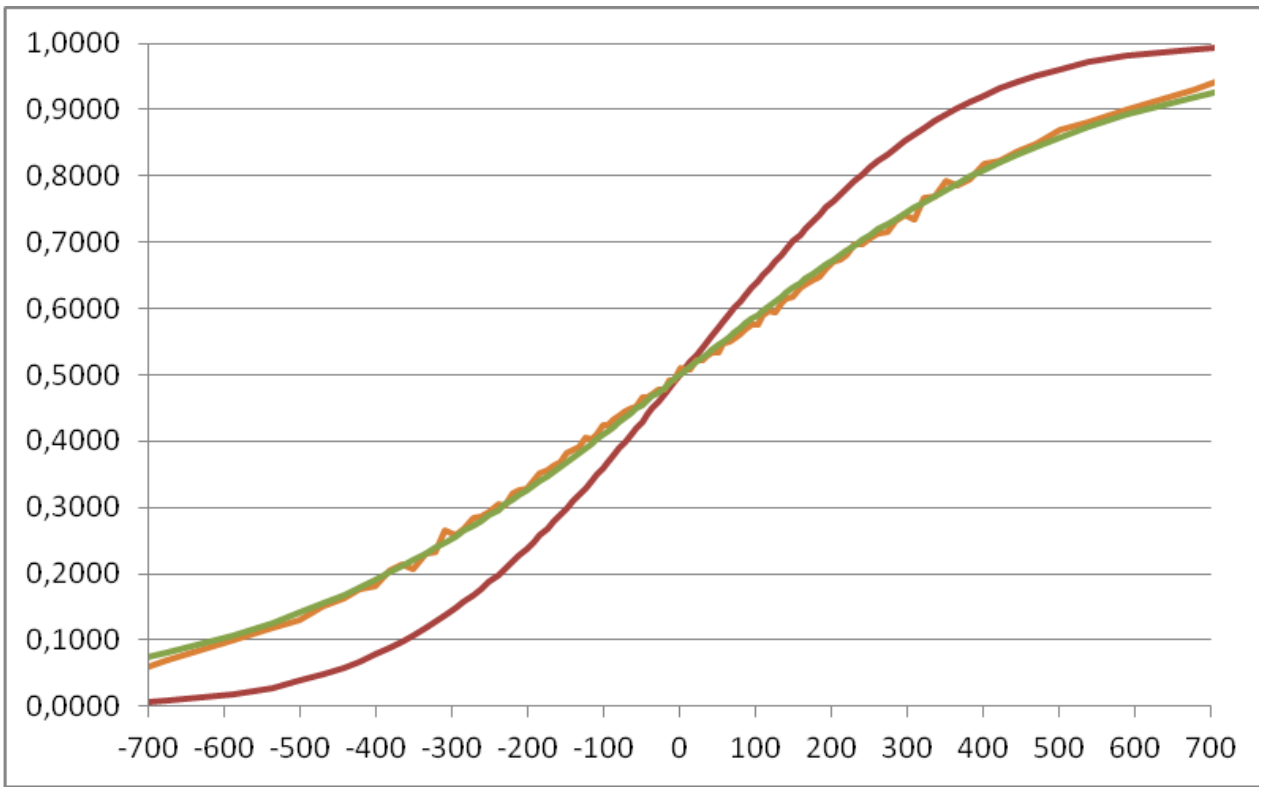
The orange curve shows the real results of all 467.761 results.
The red curve shows the expected scores using our current table 4.
The green curve shows th expected scores using the continuous formula proposed by Prof. Elo.



That means that the winning expectancy for the higher rated player is calculated significantly higher that the real results and vice versa lower for the lower rated player.

The following chart shows the distribution of the avarge score depending on the rating difference, but now using the adapted continuous formula using the value 640 instead of 400.

The orange curve shows the real results of all 467.761 results.
The red curve shows the expected scores using our current table 4.
The green curve shows the expected scores using the adapted continuous formula.



The green curve matches very well the distribution of the real results.

In our current rules the rating difference in a game is limited to 350. If we use the adapted continuous formula this value should be extended to 1.6*350 = 560.

So far the original goal of my project was reached.

I discussed the method to adapt the continuous formula over some months with Prof. Glickman. First he was very sceptical and refused the idea. At the end he agreed after I sent him the following table (WCCC final 27). Dronov became World Champion and lost 9 rating points in this tournament. Must he play better than a World Champion to save his rating ?

						old formula		adapted formula	
ICCFNR	rating	Sheet	score	result	D	expected	ΔW	expected	ΔW
940089	2594	X½½½½½½½½½½½½½½½½	8,0	0,5	86	0,6213	-0,1213	0,5767	-0,0767
180109	2609	½X½½½½½½½½½½½½½½1½	8,5	0,5	71	0,6008	-0,1008	0,5635	-0,0635
110025	2541	½½X½½½0½½½½½½½½½½0	7,0	0,5	139	0,6900	-0,1900	0,6225	-0,1225
81027	2652	½½½X½½½½½½½1½½½½½1	9,0	0,5	28	0,5402	-0,0402	0,5252	-0,0252
70305	2569	½½½½X½½½½½½½0½½½0	7,0	1,0	111	0,6545	0,3455	0,5985	0,4015
20531	2551	½½½½½X½½½½½½½½½½½½	8,0	0,5	129	0,6776	-0,1776	0,6140	-0,1140
81204	2630	½½1½½½X1½½0½½½½1½	9,0	0,5	50	0,5715	-0,0715	0,5449	-0,0449
400062	2546	½½½½½½0X½½½00½½1½	7,0	1,0	134	0,6838	0,3162	0,6182	0,3818
71079	2521	½½½½½½½½X½½½½½½½0	7,5	0,5	159	0,7141	-0,2141	0,6392	-0,1392
370913	2542	½½½½½½½½½X½½½½½½½	8,0	0,5	138	0,6888	-0,1888	0,6216	-0,1216
140130	2549	½½½0½½1½½½X½½½½1½	8,5	0,5	131	0,6801	-0,1801	0,6157	-0,1157
100376	2602	½½½½½½½1½½½X½½½½½½	8,5	0,5	78	0,6104	-0,1104	0,5697	-0,0697
Dronov	2680	½½½½1½½1½½½½X½½1½	9,5						
279003	2628	½½½½½½½½½½½½X½½½	8,0	0,5	52	0,5743	-0,0743	0,5466	-0,0466
10464	2570	½½½½½½½½½½½½½X½½	8,0	0,5	110	0,6532	-0,1532	0,5977	-0,0977
81652	2520	½0½½½00½½0½0½½X½	5,5	1,0	160	0,7153	0,2847	0,6401	0,3599
83246	2542	½½101½½½1½½½½½½½X	9,0	0,5	138	0,6888	-0,1888	0,6216	-0,1216
					Sum	ΔW	-0,8645	ΔW	-0,0158
						ΔR	-9	ΔR	0

The table is theoretical using only the startratings and does not correspond with the real evaluation which lasted over more than one period what has impact on D.

Mr. Glickman made some other proposals, especially the calculation of unfixed ratings and for new players. Our current method described he as „clumsy“. I would agree with the proposed iterative method but there is currently no detailed description and no software available to test this. It could be a task for 2017 – if we do not decide to use the Glicko-System in 2018.

For more details see the reports of the Service Director and the Qualification Commissioner and the proposals 2016-017 and 2016-022.

Other ongoing tasks

- Improvement of the withdrawal wizzard and the Robo-TD
- New norm calculation

Looking forward to seeing you in Bremen !

Amici sumus

Gerhard Binder
ICCF Ratings Commissioner