CHALMERS | GÖTEBORG UNIVERSITY

PREPRINT 2005:28

Which ball is the roundest? – a suggested tournament stability index

TORBJÖRN LUNDH

Department of Mathematical Sciences
Division of Mathematics
CHALMERS UNIVERSITY OF TECHNOLOGY
GÖTEBORG UNIVERSITY
Göteborg Sweden 2005

Preprint 2005:28

Which ball is the roundest?
– a suggested tournament stability index

Torbjörn Lundh

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Sciences
Division of Mathematics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg, Sweden

Göteborg, July 2005

Preprint 2005:28
ISSN 1652-9715

Matematiskt centrum
Göteborg 2005

Which ball is the roundest?
– a suggested tournament stability index

Torbjörn Lundh
Mathematical Sciences
Chalmers and Göteborg University, Sweden
torbjorn.lundh@chalmers.se

July 18, 2005

Abstract

All sports have components of randomness that cause the "best" individual or team not to win every game. According to many spectators this uncertainty is part of the charm when following a competition or a match. Have different sports more or less of this unpredictability? We suggest here a general measure, a tournament stability index, which could measure this randomness factor for different tournaments, and different sports. As an illustration we use exemplify this measure for basketball, squash, and soccer. Results will also be given on tournaments in (American) football, ice-hockey, and handball. Furthermore, we will state a couple of combinatorial optimization questions that turned up on the way.

1 Introduction

We have probably all heard sport commentators saying something like: "The ball is round and can go either way. That's why football is special." You can here change the word football to almost any other ball sport. Even to ice–hockey, if you allow yourself to a rather liberal view of the definition of a ball.

How to quantify this unpredictability? The idea we will use is very simple: How often will a "better" team lose against a "weaker" opponent in a tournament?

2 Tournament stability index

Suppose that there is a ranking list[1] ρ of a group of n teams or individuals that play a tournament. Let a and b be two teams (or individuals) that play each other in game i in

[1] I.e. ρ is a bijection from the set of n teams to $\{1,2,\ldots,n\}$.

the tournament. Assume that a is ranked higher than b, i.e. $\rho(a) > \rho(b)$. We put a value, v_i, on this game i according to the following scheme:

$$v_i = \begin{cases} 1 & \text{if } a \text{ wins} \\ -1 & \text{if } b \text{ wins} \\ 0 & \text{if there is a draw.} \end{cases} \tag{1}$$

This evaluation is used in the so called JWB[2] ranking system; see [8] for example.

We get the tournament index if we sum up all matches according to the scheme (1) above and divide the total by the number of matches played. That is, if a total of k matches in the tournament is played, the tournament index, $\mathcal{T}$, is defined as

$$\mathcal{T}(\rho) = \frac{\sum_{i=1}^{k} v_i}{k}. \tag{2}$$

We have immediately that $-1 \leq \mathcal{T}(\rho) \leq 1$ and that $\mathcal{T}(\rho)$ is close to 1 if the ranking ρ is "correct" and there is not much randomness in the game. On the other hand, if there is much randomness, $\mathcal{T}(\rho)$ will be close to zero. If on the other hand $\mathcal{T}(\rho)$ is close to -1, then $n - \rho + 1$ would be a good ranking, where n is the number of teams or players.

3 After–ranking

The index depends heavily on the ranking we choose, see for example the last paragraph in Section 5 . To get around that problem, we will use an after–ranking or quite simply, a result–list. Remember that we are interested in the stability of an already completed tournament, not to predict any future result, which is what rankings are usually supposed to do.

We will pick an after–ranking based on the number of games won, and if that number is equal for two or more teams (or individuals), the internal meetings will decide which team is ranked higher. If ρ_r is this result ranking, we denote $\mathcal{T}(\rho_r)$ by $\mathcal{T}_r$.

Using this after–ranking, we can expect a high tournament index; but perhaps a little surprisingly, we cannot always expect to get the highest possible index result by using this result–list ranking. That is, in some cases, there is an optimal ranking ρ_o such that $\mathcal{T}_o := \mathcal{T}(\rho_o) > \mathcal{T}_r$. We will come back to this peculiarity in Appendix A below, but for now, let us concentrate on $\mathcal{T}_r$.

4 Normalized tournament index

A problem with using the tournament's result–list as a ranking for studying the stability of the same tournament is that the index will be biased. For example, even if all games were decided by coin flipping, we would of course get a clearly positive tournament index.

[2] Just Win Baby

Suppose n teams meet every other team m times in a tournament, where all games were settled by coin-tossing. Let $M(n, m)$ denote the expected value of the index $\mathcal{T}_r$ of such a random tournament. In Table 1 we give approximate values of $M(n, m)$ using a Monte Carlo method.

To make up for the internal bias we introduce by choosing an after–ranking, we define a normalized result tournament index, $\widehat{\mathcal{T}_r}$, by a translation and rescaling of $\mathcal{T}_r$ in the following way.

$$\widehat{\mathcal{T}_r} = \frac{\mathcal{T}_r - E(\mathcal{T}_r)}{1 - E(\mathcal{T}_r)},$$

where $\mathcal{T}_r$ is calculated as in (2) for n teams playing each other m times, giving a total of $k = n(n-1)m/2$ games played and using the result–ranking ρ_r. We note that $\widehat{\mathcal{T}_r} \leq 1$, and that the expected value of $\widehat{\mathcal{T}_r}$ would be zero, if the outcome of all games was decided randomly.

We can use the index $\widehat{\mathcal{T}_r}$ as a measure of the stability of a tournament, which could be compared to other tournaments of different sizes and for different sports. Let us now pick a few real world examples as an illustration.

5 NBA 1995–1996

Let us start with basketball and the NBA season 1995–1996. 29 teams played 82 games each except the play-off teams who played up to 103 games in total (Seattle Super Sonics). Wins and losses are conveniently presented in a matrix [5]. All teams met each other either two or four times before the play-off. In total there were 1189 games played. Using this data, we find $\mathcal{T} = 0.41$. In order to compute the normalized index, we need the expected value of an analogue tournament where all matches were decided randomly. In order to get a first approximation of this, we consider a tournament where all 29 teams met exactly three times each. We will discuss an alternative to this approximation in Section B where we obtain $M \approx 0.15$. Such a tournament would give a total of 1218 games which is a good approximation of 1189. We can then use Table 1 where $M(29, 3) \approx 0.14 \approx 0.15$ and can estimate

$$\widehat{\mathcal{T}_r} \approx \frac{0.41 - 0.15}{1 - 0.15} \approx 0.29.$$

For comparison, let us see what happens if we pick ordinary rankings ahead of the actual season. With the ranking ρ given in [10], based on the previous season, taking into consideration the actual points difference in each game, we get a tournament index of $\mathcal{T}(\rho) = 0.31$. But if we instead choose a ranking method described in [9] which weights the different games according to the strength of the opponent, we get instead $\mathcal{T}(\rho) = 0.073$. (We will come back to this approach in Section C.)

6 Squash

We will now exemplify the tournament index for an individual sport, namely squash, and more specifically the professional cups which are played around the world. The professional squash association, PSA, produces rankings of the players, see [1]. We pick the twenty highest ranked players from the list of 1st January 2002 and follow their results during the year 2001.

We record each game whenever two players from the list meet making a result matrix this way. In total we recorded 153 games this way. Using the result-ranking (which differs from the PSA January 2002 ranking) we get $\mathcal{T} \approx 0.71$. Normalizing this, we find that $\widehat{\mathcal{T}} \approx (0.71 - 0.33)/(1 - 0.33) \approx 0.56$, where we used the normalization factor 0.33 taken from a Monte Carlo method of thinning of random matrixes described below in Section B. Alternatively, we can use Table 1 to see that in a tournament with 20 players where each one meets once, we get a total of 190 games and $M(20, 1) = 0.34$. In comparison to $\widehat{\mathcal{T}} = 0.56$, we see that if we use the ranking from January 2001 and follow the 20 highest ranked players during 2001, we get $\mathcal{T} = 0.26$, and if we use the PSA ranking from January 2002 to follow the games during 2001, we get $\mathcal{T} = 0.50$.

Comparing with the NBA example, where we got $\widehat{\mathcal{T}} = 0.29$, it seems that the professional top-squash during 2001 was more "stable" than the NBA season 1995–1996.

7 Premier League 2000–2001

Let us now turn to soccer, or more accurately football since we will take a look at the English Premier League results during the season 2000–2001. We collect our data from [4]. Here, there were 20 teams playing each other 2 times each. That gives us a total of 380 games. Using our result-ranking, we get $\mathcal{T} = 0.24$, and from Table 1 we get $M(20, 2) = 0.207$. Hence

$$\widehat{\mathcal{T}} \approx \frac{0.24 - 0.207}{1 - 0.207} \approx 0.048.$$

This result seems to indicate that professional soccer is much more random than both basketball and squash. At least for these three tournaments studied.

Let us look at another soccer tournament to see if we get a similar result. But before we do so, let us give an alternative description of what $\widehat{\mathcal{T}} = 0.049$ means.

8 An alternative description of small tournament indexes

One might wonder if a normalized tournament index of 0.048 indicates a situation close to complete randomness? Let us express this in another way by letting X be the the tournament index we get from an experiment where a we toss a coin to decide each game in a tournament with 20 teams, meeting each other 2 times each. The expected value of X

is defined $M(20,2)$. By Monte Carlo simulations we can estimate the standard deviation of X.We can also, thanks to a Lilliefors' test, approximate the distribution of X with a normal distribution. Now we can compute $P[X > 0.24] \approx 0.16$. That is, if we were to repeat the coin tossing experiments, we would in 16% of the trials, get an index $\mathcal{T}$ higher than what we got in the soccer example. If we did the same thing for the basketball season above, we would get an analogue frequency less than 0.0001%.

9 More soccer

Let us compare the Premier League result with another European professional soccer tournament, the German Bundesliga. We pick up the data from [2] and treat it in a similar way as above. This will give us $\mathcal{T} = 0.24$ and from the Table 1, we have $M(18,2) = 0.22$, which gives us a normalized tournament index $\widehat{\mathcal{T}} \approx 0.027$. That is even less than its English version! If we make coin tossing tournaments of the same size, we will in 31 % of the trials get a higher index.

10 Discussion

Our few tournament results listed in Table 3, might indicate that for example squash and basketball seem to be more stable sports than soccer, in the sense that the "better" player or team more often wins, compared to soccer, at least on a professional level. However, more tournament results need of course to be studied before one could make a more solid statement on this.

But what if soccer is indeed more random in its nature. What could then be the causes? One obvious reason is that rather few goals are scored so that many games end in a draw which will decrease the index. But this is not the only reasons as we shall now see. Let us again look at the Bundesliga season and disregard all drawn games as if they were never played. This will give us a total of 237 games instead of 306. This will give a lower M, see Section B, and a higher normalized index $\widehat{\mathcal{T}} = 0.082$, which is still rather low.

Another reason might be that the level of the top soccer players is extremely high and even. There are very few natural talents in that sport that are not taken care of at an early stage. Many children play with a soccer in some form all over the world, but not all have ever seen a squash ball.

By measuring more basketball, squash, soccer, and other tournaments, one would ask if there might be some universal numbers of the randomness for the different sports. How do professional series differ from amateur tournaments? Maybe there is an interval where the tournament index should lie to become an attractive public sport? Maybe this interval differs from person to person? How often do we want "David to defeat Goliath"?

	m					
n	1	2	3	4	5	6
2	1.000	0.501	0.498	0.372	0.377	0.316
3	0.842	0.473	0.417	0.348	0.316	0.284
4	0.791	0.458	0.381	0.314	0.286	0.256
5	0.713	0.409	0.343	0.283	0.259	0.228
6	0.642	0.381	0.318	0.262	0.236	0.210
7	0.596	0.352	0.294	0.243	0.219	0.193
8	0.552	0.331	0.276	0.225	0.206	0.182
9	0.521	0.310	0.261	0.215	0.193	0.172
10	0.489	0.295	0.246	0.204	0.183	0.163
11	0.467	0.281	0.234	0.194	0.175	0.156
12	0.445	0.270	0.225	0.184	0.169	0.149
13	0.426	0.259	0.216	0.178	0.161	0.144
14	0.409	0.250	0.206	0.171	0.156	0.139
15	0.393	0.240	0.201	0.166	0.149	0.133
16	0.380	0.233	0.195	0.160	0.145	0.129
17	0.368	0.224	0.187	0.155	0.141	0.125
18	0.356	0.218	0.183	0.151	0.137	0.122
19	0.345	0.211	0.178	0.147	0.133	0.119
20	0.336	0.207	0.173	0.144	0.130	0.115
21	0.327	0.201	0.169	0.140	0.127	0.112
22	0.319	0.198	0.166	0.136	0.123	0.110
23	0.310	0.193	0.161	0.133	0.121	0.107
24	0.305	0.187	0.158	0.130	0.118	0.105
25	0.298	0.185	0.154	0.127	0.116	0.103
26	0.290	0.180	0.151	0.126	0.113	0.101
27	0.286	0.176	0.149	0.122	0.111	0.098
28	0.280	0.173	0.145	0.120	0.109	0.097
29	0.274	0.169	0.143	0.118	0.107	0.095
30	0.269	0.167	0.141	0.116	0.106	0.094

Table 1: Approximations of expected tournament indexes for completely random games, $M(n, m)$, where the third decimal should only be viewed as an indication. To illustrate this, note that Equations (5) or (6) gives the exact values for the first line which should then really read 1.000 0.500 0.500 0.375 0.375 0.3125. We have used 5000 random matrixes in the Monte Carlo simulation for each pair n, m.

	m					
n	1	2	3	4	5	6
2	1.00	0.49	0.49	0.38	0.38	0.31
3	0.83	0.49	0.43	0.35	0.34	0.29
4	0.79	0.46	0.42	0.33	0.32	0.28
5	0.74	0.44	0.39	0.32	0.30	0.26
6	0.70	0.42	0.37	0.31	0.29	0.25
7	0.65	0.39	0.35	0.29	0.27	0.24
8	0.61	0.38	0.33	0.27	0.26	0.22
9	0.59	0.36	0.31	0.26	0.24	0.21
10	0.56	0.34	0.30	0.24	0.23	0.20
11	0.53	0.33	0.29	0.23	0.22	0.19
12	0.51	0.30	0.28	0.22	0.21	0.18
13	0.49	0.30	0.27	0.21	0.20	0.17
14	0.47	0.28	0.26	0.20	0.20	0.17
15	0.45	0.27	0.25	0.20	0.19	0.16
16	0.44	0.26	0.24	0.19	0.18	0.16
17	0.42	0.25	0.23	0.19	0.18	0.15
18	0.41	0.24	0.22	0.18	0.17	0.14
19	0.40	0.24	0.22	0.17	0.17	0.14
20	0.39	0.23	0.21	0.17	0.17	0.14
21	0.37	0.22	0.21	0.16	0.16	0.14
22	0.37	0.21	0.20	0.16	0.15	0.13
23	0.36	0.21	0.20	0.15	0.15	0.13
24	0.35	0.21	0.19	0.15	0.15	0.12
26	0.34	0.21	0.19	0.15	0.14	0.12
27	0.33	0.20	0.18	0.14	0.14	0.12
28	0.33	0.19	0.18	0.14	0.14	0.12
29	0.32	0.19	0.18	0.14	0.13	0.11
30	0.32	0.18	0.17	0.14	0.13	0.11
31	0.31	0.18	0.17	0.13	0.13	0.11

Table 2: Crude approximations upper estimations of expected optimized tournament indexes for random games, $M_o(n, m)$. We have here used 500 random matrixes in the Monte Carlo simulation for each pair n, m, where as usual m stands for the number of games each team play each other team, and n for the number of teams.

Sport	**Country**	**Season**	**Tournament**	$\widehat{\mathcal{T}_r}$	$\widehat{\mathcal{T}_o}$	**CTI**	$\widehat{\mathcal{W}_o}$
Soccer	England	00/01	Premier League	4.8	17	16	22
Soccer	England	02/03	Premier League	13	24	0.42	23
Soccer	England	03/04	Premier League	11	27	1.2	24
Soccer	Germany	00/01	Bundesliga	3	13	31	18
Soccer	Germany	02/03	Bundesliga	-1.0	16	57	19
Soccer	Germany	03/04	Bundesliga	11	25	3.1	28
Soccer	Germany	02/03	2 Bundesliga	6	21	16	24
Soccer	Germany	03/04	2 Bundesliga	-7.0	5.0	88	13
Soccer	France	03/04	Division 1	10	22	1.8	22
Soccer	Spain	03/04	Division 1	11	26	1.2	18
Soccer	Italy	03/04	Seria A	21	47	0.006	34
Am. Football	USA	01/02	NFC	17.5	·	11	
Am. Football	USA	01/02	AFC	14	·	16	
Ice-hockey	North America	02	NHL eastern conf.	-19	·	58	
Ice-hockey	North America	02	NHL western conf.	6.8	·	26	
Ice-hockey	Switzerland	03/04	National league	19	26	0.008	22
Ice-hockey	Germany	03/04	DEL	9.8	5.8	1.6	15
Handball	Germany	03/04	Bundesliga	47	·	< 0.0001	
Basketball	USA	96/97	NBA	29	·	< 0.0001	
Squash	Intern.	01	PSA	56	·	0.0035	

Table 3: This is a comparison between different tournaments from different sports, countries,and years. *Please note that the values of the indexes have been multiplied by 100 for easier display.* We list rather *crude approximations* of four suggested indexes, the normalized touring index with respect to the result list $\widehat{\mathcal{T}_r}$, the normalized touring index with respect to the optimal ranking $\widehat{\mathcal{T}_o}$, the *Coin Tossing Index*, **CTI**, which is the expected probability for a random tournament to have a higher $\widehat{\mathcal{T}_r}$ than obtain in the tournament in question, and finally, a rough estimate of the normalized Weighted Ranking Index, $\widehat{\mathcal{W}_o}$, see Section C below.

In the following appendix, we list a few more or less technical questions that turned up during our study above.

A Optimal ranking

We can represent a tournament with a $n \times n$–matrix A with elements $a_{ij} \geq 0$ denoting the number of victories team i has against team j among the n teams in the tournament. Since the teams do not meet themselves, the diagonal will be zero. Let us view the ranking ρ as a permutation of $(1, 2, 3, \dots, n)$. Then we have that

$$\mathcal{T}(\rho) = \frac{1}{k} \sum_{i=1}^{n} \sum_{j=1}^{n} sgn\big(\rho(i) - \rho(j)\big) a_{ij}, \tag{3}$$

where $k = \sum_i \sum_j a_{ij}$, i.e. the total number of decided games, and $sgn(\cdot)$ is the sign function. Then the optimal ranking is the ranking/permutation ρ that gives the largest index, i.e.

$$\mathcal{T}_o := \max_{\rho} \mathcal{T}(\rho). \tag{4}$$

We will now give two examples where the usual result list, i.e. a team with more victories will be ranked higher than one with fewer wins, does not give the highest index. That is examples where $\mathcal{T}_o > \mathcal{T}_r$.

A.1 Example 1 — three teams meeting each other 100 times.

Suppose such a tournament gives the following result matrix:

$$A = \begin{pmatrix} 0 & 41 & 52 \\ 59 & 0 & 53 \\ 48 & 47 & 0 \end{pmatrix}.$$

The usual result list, or as we also called it, after–ranking, will be $(2, 3, 1)$ since team number 2 has 112 wins in total, team 3 has 95 wins, and team 1 has 93 wins. Using this ranking we get an index $\mathcal{T} = 1/15$. But if we instead pick the ranking $(2, 1, 3)$ we get $\mathcal{T} = 7/75 > 1/15$. We see that we can increase the index by switching places of teams that have almost the same number of total victories. The reason in this case is that team 1 has 52 wins and 48 losses against team 3.

In the following example, we limit ourselves to tournaments with just one match per pair. We then have to increase the number of teams to five in order to find an example where the result list will not give the optimal ranking.

A.2 Example 2 — five teams meeting each other only once.

Suppose the tournament matrix will be

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

then one, of three possible, result–list rankings will be $(4, 1, 2, 3, 5)$ giving the index $\mathcal{T} = 2/5$. (The reason why there is more than one possible result is that team 1, 2, and 3 all have two wins each and one internal win among each other, i.e. 1 won over 2 which in turn won over 3 who won over 1.) However, the ranking $(4, 3, 1, 5, 2)$ gives a higher index $\mathcal{T} = 3/5$. Note that team 5's only victory was against team 2.

A.3 How to find the optimal ranking?

Question 1 *Is the problem of finding the optimal ranking in (4) NP-complete?*

Due to the similarities to well known NP-complete problems such as *the (directed) optimal linear arrangement*, c.f. [3, p. 200]; *the quadratic assignment problem*, c.f. [7] and [3, p. 218], the author would be very surprised if the answer to that question would be no, even for the case where the team just meet each other once. See also other related problems in the three volumes of [6].

A.4 The expected value of the optimal index $\mathcal{T}_o$ for a random tournament

Intuitively, one might argue that $M_o(n, m)$ will decrease when the number of matches, m, increases since the difference between the artificial teams will be levelled out when there are more coin tosses. Similarly, we might expect $M_o(n, m)$ to decrease when the number of teams, n, increases, since it will be harder to find a clear ranking when more teams are involved.

In the simple case where we just have two teams we can give a closed expression for the expected random tournament index.

$$M_o(2, m) = \frac{(m-1)!}{2^m} \sum_{i=0}^{m} \frac{|m-2i|}{i!(m-i)!}. \tag{5}$$

Note here that if the number of matches m is an even number, then $M_o(2, m+1) = M_o(2, m)$.

To the authors grateful satisfaction, this formula (5) was later simplified by Sven-Erick Alm and Allan Gut of Uppsala University, to the following form.

$$M_o(2, m) = \frac{1}{2^{m-1}} \text{Bin}(m-1, [\frac{m-1}{2}]), \tag{6}$$

where $[\cdot]$ stands for the integer part.

For three teams, it is easy to see that $M_o(3, 1) = \frac{5}{6}$. In Figure 1 we illustrate $M_o(n, 1)$ and suggest an approximative function.

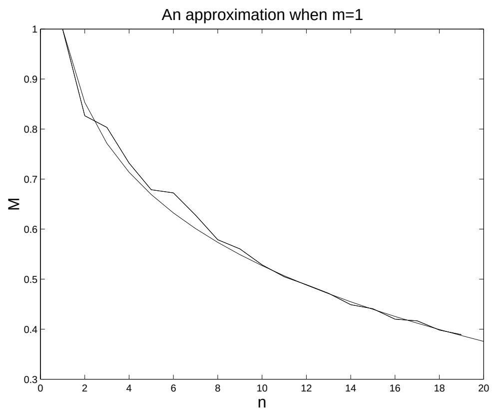

Figure 1: A numerical indication with the help of Monte Carlo simulations of the function $M_o(n, 1)$. The smooth curve comes from an extrapolation of exact values of $M_o(n, 1)$ for n from 2 to 7 with the suggested function $M_o(n, 1) \approx n^{-\frac{1}{5}} \exp(\frac{1-n}{50})$. The other curve is a result from a crude algorithm that searches for the optimal tournament index, M_o.

Question 2 *Is it possible to find closed expression for the expected index for the optimal ranking $M_o(n, m)$ for higher combinations of n and m than those given?*

Using a simple algorithm that tries to find the best ranking to compute the index, see Figure 1, we can compute an approximation of the optimal normalized tournament index, $\widehat{\mathcal{T}_o}$, for the examples above. In the NBA example, we get $\widehat{\mathcal{T}_o} \approx (0.42 - 0.17)/(1 - 0.17) \approx 0.30$. For the Premier League season we have $\widehat{\mathcal{T}_o} \approx (0.2553 - 0.23)/(1 - 0.23) \approx 0.033$. The squash example gives us $\widehat{\mathcal{T}_o} \approx (0.73 - 0.39)/(1 - 0.39) \approx 0.55$. In these three cases we do not get any dramatic change of the index, and we might assume that the same is true if we were able to find the best ranking, not just an approximation of it.

B A thinned matrix or an accumulated random matrix?

We had both in the basketball and the squash tournaments problems with the fact that not all teams, or individuals, played each other symmetrically equally many games. We solved the problem of finding a good M value needed in the normalization by generating a full random matrix that we thinned out by removing results at random positions until the accurate total number of games was left. We could then compute an index for that thinned random result matrix. Repeating this process gives us an estimate of the normalizing factor.

A variant of this would be if we instead started with an empty matrix and added wins, i.e. ones, at a random position in the matrix. Repeating this process we would get a estimate of the random index for tournaments where m teams play a total of t games between randomly picked pairs, and where all games were decided randomly.

Question 3 *What will be the difference between the thinning and the accumulation approaches?*

We illustrate this question in Figure 2.

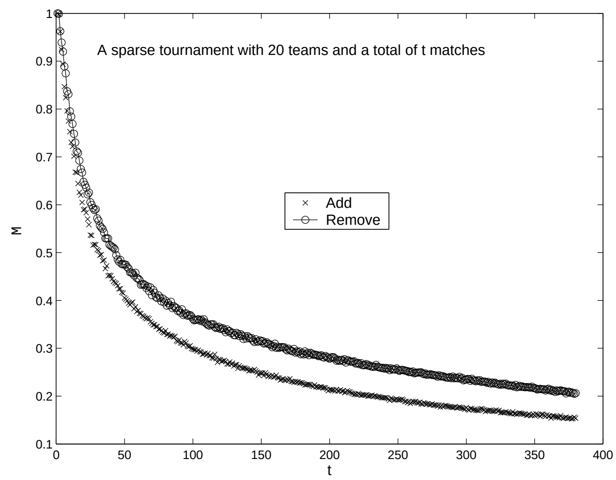

Figure 2: A numerical indication of the difference between the estimated random tournament index where 20 teams play a total of t games. We used the thinning of a full random result matrix with $n = 20$ and $m = 2$ and compared it when we accumulated a random matrix instead. From the picture it looks as if the estimated index from the accumulation process gives a lower value. Is that a common feature? If so, what could a heuristic explanation be?

In the depicted example in Figure 2, we started in the thinning case with a tournament matrix where $m = 2$. We might have picked a different m.

Question 4 *In a tournament where not all teams play the same number of decided games. Can we say anything in general how the resulting estimation of the tournament index will depend on m for "thinned tournaments"?*

This question is illustrated in an example in Figure 3. From the picture it seems as if the expected index will decrease as m increases.

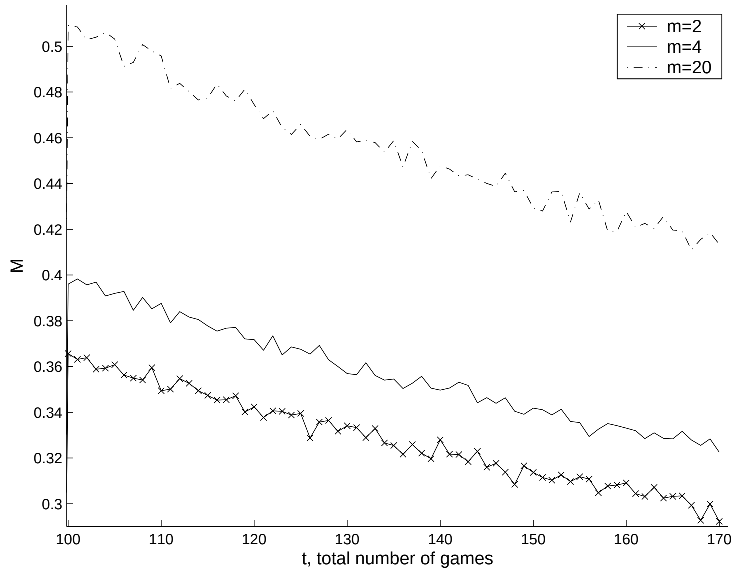

Figure 3: As in Figure 2 above, $n = 20$. The three graphs represent estimations of the expected random tournament index for different values of m of the starting matrix that will be thinned out until only the total sum of all the elements in the matrix is t. We have picked $m = 2$, $m = 4$ and $m = 20$.

Let us finally mention one last question in this section. In the squash example we had a tournament over a year that was composed by a series of cups. That gave the consequence that the best player[3] also played the most games.

Question 5 *If in a tournament, the "best" teams play more games as during a cup, how will that effect the normalized index?*

C A weighted ranking

The scheme (1) we have used so far for evaluate the outcome of a match is blunt in the sense that it punish the score with -1 indifferently if for example the highest ranked team is beaten by the lowest ranked, as if it would have been beaten by the second highest ranked team.

A way to get around this feature is to introduce a weighted ranking, x, in the following way. Let $x = (x_1, x_2, \ldots, x_n)$, where all $x_i \in [0, 1]$. Our new evaluation scheme of a given tournament $A = \{a_{ij}\}$, will be

$$\mathcal{W}(x) = \frac{1}{k} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j) a_{ij}, \tag{7}$$

[3] Peter Nicol

where $k = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}$. Compare this formulation with (3).

Now, let

$$\mathcal{W}_o = \max_{x \in [0,1]^n} \mathcal{W}(x). \tag{8}$$

We denote the optimal weighted ranking by x_o, i.e. $\mathcal{W}(x_o) = \mathcal{W}_o$, and where $x_o \in [0,1]^n$.

Question 6 *Is the problem to compute the optimal weighted ranking in (8) is NP-complete?*

We could use $\mathcal{W}_o$ as an alternative stability index for tournaments, after it has been normalized that is to $\widehat{\mathcal{W}_o}$.

We give a very rough estimate of this index in the last column in Table 3 above. Our naive first strategy to estimate this to use several of the candidates we used in our search for a numerical estimation of $\widehat{\mathcal{T}_o}$. We use this rankings and the lemma below to get candidates x with only zeroes and ones as their components, the ones at the highest ranked positions. All these candidate vectors were then evaluated in Equation (8).

The following immediate result can be a tool in the investigation of Question 6.

Proposition 7 *The optimal weighted ranking is in a (not necessary unique) corner in the unit hyper-cube, i.e.* $x_o \in \{0,1\}^n$.

Proof. The partial derivative with respect to x_i of Equation (8) gives us immediately that the extremal value of $\mathcal{W}(x)$ has to be attained when x is in a corner in the unit hyper-cube. $\square$

Let us look closer at the Premier League series 2000-20003, where the game matrix A is ordered by the final result ranking. In this case we get a candidate for x_o with

$$x = (1,1,1,1,1,1,1,1,0,1,1,0,0,0,0,0,0,0,0,0) \quad \text{giving us } \mathcal{W}_o \geq 0.2546.$$

We can then approximate $\widehat{\mathcal{W}_o} \approx 0.22$.

Question 8 *Is there a more suitable way of weighing the ranking, than the two alternatives we have studied?*

References

[1] R. Beck et al. *SquashTalk*, Cupresults and ranking in squash, www.squashtalk.com/

[2] Bundesliga 2000/01, www.ligaverwaltung.de/html_example/matrix.htm

[3] M.R. Garey and D.S. Johnson, *Computers and Intractability, A Guide to the Theory of NP-Completeness*, W.H. Freeman and company, San Fransisco, 1979.

[4] Premier League England 2000/01, www.ligaverwaltung.de/html_example/e/matrix.htm

[5] L.D. Oliver, *My '96 Numbers,* Journal of Basketball studies, see the statistics at www.rawbw.com/∼deano/

[6] A. Schrijver, *Combiatorial Optimization*, Algorithms and Combinatorics 24, Springer-Verlag, 2003.

[7] S. Sahni and T. Gonzalez, *P-Complete Approximation Problems*, Journal of the Association for Computing Machinery, Vol. 23, No. 3, July 1976, pp 555-565.

[8] S.P. Sorensen, *An overview of some methods for ranking sports teams,* preprint from 1999 at www.phys.utk.edu/sorensen/ranking/

[9] www.phys.utk.edu/sorensen/cfr/cfr/NBA/1994/NBA_1994_Ranking_Schedule.html

[10] www.phys.utk.edu/sorensen/cfr/cfr/NBA/1994/NBA_1994_Ranking_sCWS.html