

Ratings Central: Accurate, Automated,
Bayesian Table Tennis Ratings for Clubs,
Leagues, Tournaments, and Organizations

David J. Marcus

New England Symposium on Statistics in Sports
September 24, 2011

Overview

History

System Components and Flow

Who Is Using

Model and Algorithm

Event Processing and Reports

Priors

Confusions and Misconceptions

Reference

History

History

- ▶ Since the 1970s, USATT (USA Table Tennis) has had a rating system.
 - ▶ Superficial similarity to the Elo (chess) system
 - ▶ Ad hoc point update chart and adjustment formulas
 - ▶ All ratings treated as equally accurate, even though some players have played a lot and others have played little or not for a long time.
- ▶ Players of all levels (from champion to five-year old) compete in USATT-sanctioned tournaments and get ratings.
- ▶ As a member of the USATT Ratings Committee in 1997, I was asked if I could develop a better rating system.
 - ▶ Problems with the ratings included excessive volatility, inaccurate ratings, and players protecting their ratings by not playing.

History (cont.)

- ▶ By the time I had developed the new system in 1999, supportive USATT Board members had been replaced by people who would oppose anything the USATT Ratings Committee suggested.
 - ▶ USATT did not adopt the new system.
- ▶ In 2004, I and my colleague, Sean O'Neill, launched Ratings Central on our own.
 - ▶ Sean is five-time U.S. Men's Singles Champion.
 - ▶ It was Sean's idea to provide ratings to clubs.
 - ▶ I do the math, statistics, and programming.

System Components and Flow

Ratings Central Components

- ▶ Two Windows desktop apps for submitting events
 - ▶ Zermelo: Manages all aspects of running a table tennis tournament.
 - ▶ Cantor: Much simpler app whose sole purpose is to submit events to Ratings Central.
- ▶ The ratings processor app that runs on my PC at home.
- ▶ The website (www.ratingscentral.com) where match results and ratings are displayed.

System Flow

- ▶ Zermelo/Cantor submit events via email using a simple text format.
 - ▶ An *event* is a set of matches that are processed together, e.g., a tournament, a league night, all league matches in a state during a week.
- ▶ The ratings processor retrieves the email, processes the event, and uploads the new ratings to the website.
 - ▶ This typically takes a few minutes from the time the director submits the event.
- ▶ Corrections are handled by resubmitting the event.
 - ▶ The system automatically reprocesses all affected events.
- ▶ Several event directors bypass Zermelo/Cantor and submit directly into the system using their own software.
 - ▶ In particular, the Austrian table tennis associations do this.

Who Is Using

Who Is Using

- ▶ Currently, the system contains
 - ▶ 8,195 events
 - ▶ 37,874 players
 - ▶ 765,302 matches
 - ▶ 464 clubs
 - ▶ 125 event directors
- ▶ Many clubs in the U.S. submit leagues or tournaments that they run.
 - ▶ In the U.S., USATT is a “competitor”: USATT provides both tournament ratings (for a fee) and a separate league rating system (for free).
 - ▶ Some directors submit their events to both Ratings Central and USATT.

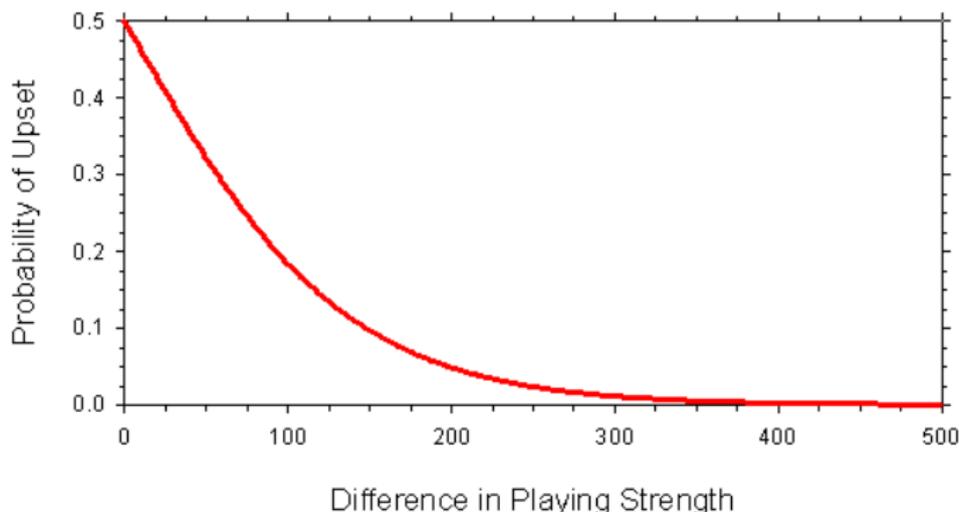
Who Is Using (cont.)

- ▶ Ratings Central provides official ratings/rankings for
 - ▶ Austrian Table Tennis Association, www.oettv.org
 - ▶ Lower Austrian Table Tennis Association, www.noettv.org (state in Austria)
 - ▶ Salzburg Table Tennis Association, www.sttv.co.at (state in Austria)
 - ▶ Table Tennis Queensland, www.tabletennisqld.org (state in Australia)
 - ▶ National Collegiate Table Tennis Association, www.nctta.org
- ▶ Three additional Austrian states plan to join. One additional Australian state is thinking of joining.
- ▶ I extract the ITTF (International Table Tennis Federation) Pro Tour and Junior Circuit events from the ITTF website and submit them to the system myself.
 - ▶ Colleagues sometimes submit other international leagues or tournaments.
 - ▶ You can compare our rankings of the top players with the ITTF's rankings.

Model and Algorithm

The Bayesian Model

- ▶ Each player has a *playing strength*, i.e., a number that quantifies how strong the player is.
- ▶ Define the *probability-of-upset function*
 $\pi(x) := 1 / (1 + e^{x/67})$:



- ▶ The probability that a player with playing strength s will beat a player with playing strength t is $\pi(t - s)$.

The Bayesian Model (cont.)

- ▶ A player's playing strength is not known, so model it as a random variable (the *player's law*) with a normal prior.
- ▶ The *temporal update* models a player's playing strength changing with time:
 - ▶ Add a zero-mean normal random walk to the player's law with a standard deviation of 70 rating points per year.

Intractability of Direct Calculation of Posterior

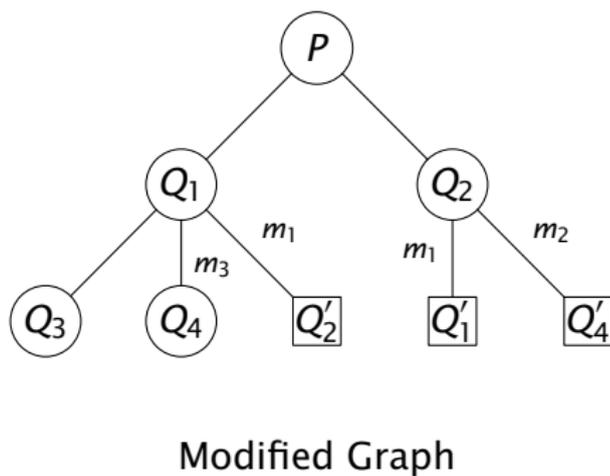
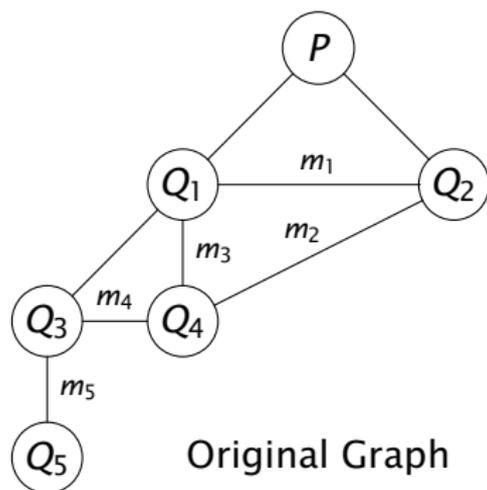
- ▶ Let N be the number of players in an event. Let L_j be the initial (i.e., start of event) law for player j .
- ▶ Let M be the number of matches in the event. Let $p(i)$ be the number of the player who wins the i th match. Let $q(i)$ be the number of the loser.
- ▶ Define

$$U(x_1) := \int_{\mathbf{R}^{N-1}} \prod_{i=1}^M \pi(x_{q(i)} - x_{p(i)}) dL_2(x_2) \cdots dL_N(x_N) L_1(x_1).$$

- ▶ The posterior law for player 1 is $U / \int_{\mathbf{R}} dU$.
- ▶ For the values of N (up to 1000) and M (up to 3000) that we have, it is not feasible to calculate this directly.

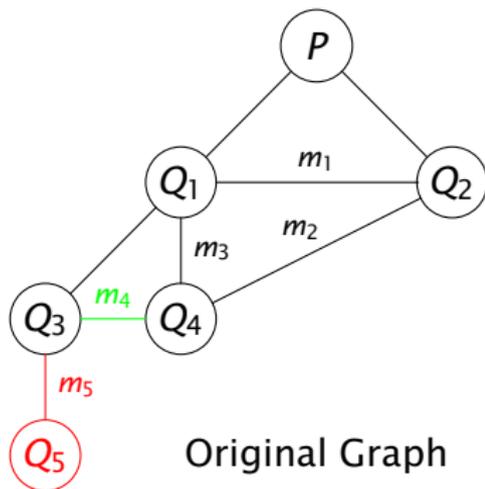
Tournament Surgery

- ▶ For an event, consider the graph where each player is a node and each match is an edge connecting two players.
- ▶ For each player P , construct a modified graph (as explained on the following slides) and use the modified graph to calculate P 's posterior law.

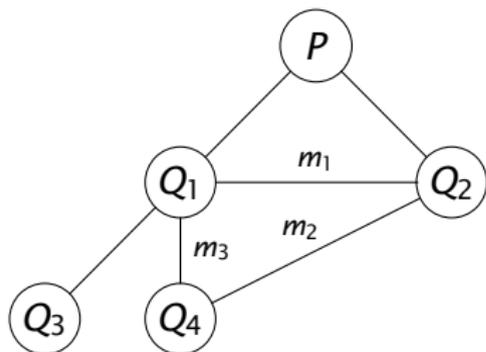


Tournament Surgery

Steps 1-2



Original Graph

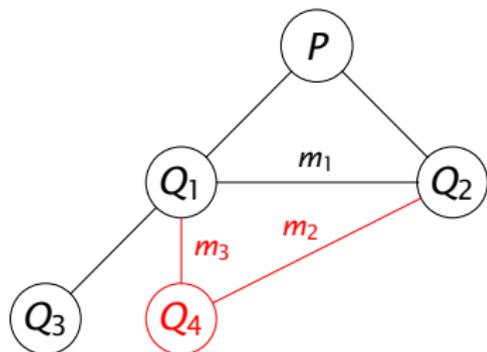


After Step 2

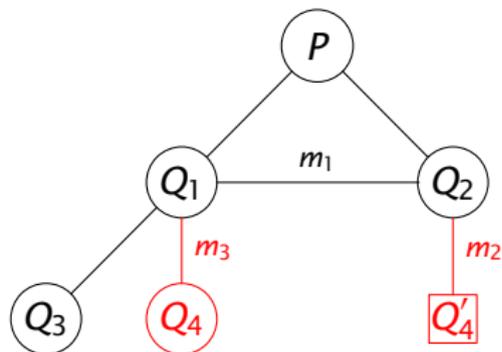
- ▶ Discard all edges that extend down from the second level of opponents (e.g., match m_5).
- ▶ Discard all edges that connect two nodes at the second level (e.g., m_4).

Tournament Surgery

Step 3



Before Step 3

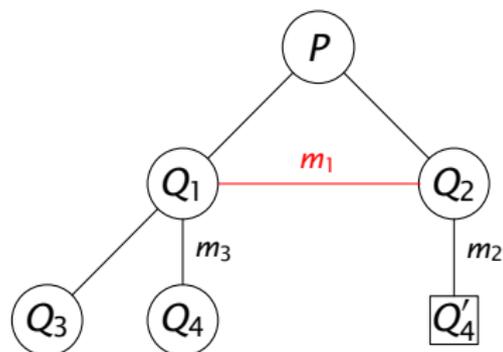


After Step 3

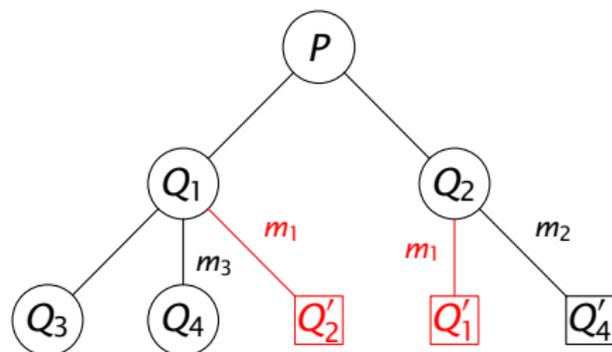
- ▶ If a node at the second level (e.g., Q_4) connects to two or more nodes at the first level, add twins of the node (e.g., Q'_4) and connect each of the nodes at the first level to exactly one of the twins.
- ▶ A *twin* is a new player who has the same initial law as the original player.

Tournament Surgery

Final Step



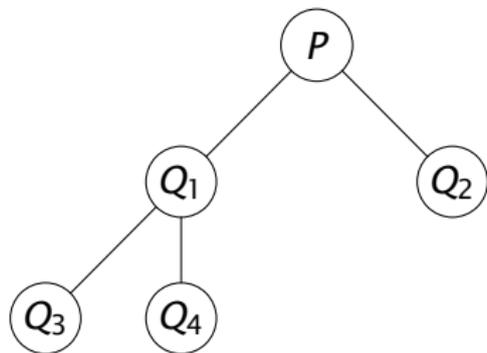
Before Final Step



Final Graph

- ▶ For each edge connecting two nodes at the first level of opponents (e.g., m_1), cut the edge and insert two new nodes on the newly created ends.
- ▶ The new nodes (e.g., Q_2' , Q_1') are twins of the original nodes.

Tractability of Posterior after Surgery



- ▶ Assume each player beats the ones immediately below them in the graph above. Let L_0 be the law of P and L_i the law of Q_i . Then to calculate the posterior for P , calculate

$$\int_{\mathbf{R}^4} \pi(x_1 - x_0) \pi(x_2 - x_0) \pi(x_3 - x_1) \pi(x_4 - x_1) \prod_{i=1}^4 dL_i(x_i) L_0(x_0)$$
$$= \int dL_4(x_4) dL_3(x_3) dL_1(x_1) dL_2(x_2) L_0(x_0).$$

Algorithm Intuition

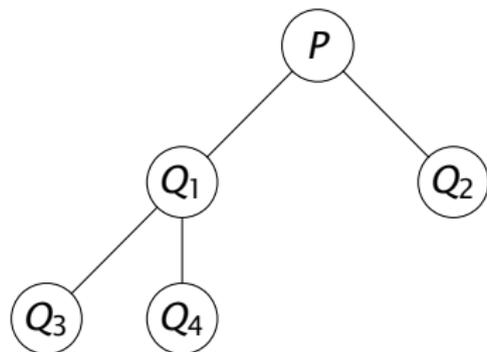
- ▶ When I go to a tournament and play somebody, I'm usually interested in how good my opponent is (especially if I just lost to them).
- ▶ I can look up their rating at the start of the tournament, but their rating may be out of date or they may be playing better or worse than their rating would indicate.
- ▶ So, I go to the posted draw sheets and find the other matches my opponent has played in the current tournament and see how they've done in those matches.
- ▶ These are the same matches that the rating system looks at for each player.

Event Processing and Reports

Event Processing

- ▶ Steps that the rating system goes through when processing an event:
 - ▶ Assign a law to each unrated player.
 - ▶ Retrieve the law of each rated player from the database and apply the temporal update.
 - ▶ For each player:
 - ▶ Calculate an adjusted law for each of the player's opponents.
 - ▶ Update the player's law for each of the player's matches using the adjusted law for the player's opponent.
- ▶ The *adjusted law* is the opponent's law updated for all of the opponent's matches except for the matches with the current player.
 - ▶ The adjusted law depends on both the player and the opponent.
 - ▶ Same opponent will have different adjusted laws when different players are being processed.

Numerical Example



- ▶ Assume all players start out as 1800 ± 100 , i.e., mean of 1800, standard deviation of 100.
- ▶ Q_1 beats Q_4 : Q_1 's updated rating is 1844 ± 90 .
- ▶ Q_1 beats Q_3 : Q_1 's updated rating is 1873 ± 83 . This is Q_1 's adjusted rating for P .
- ▶ P beats Q_1 : P 's updated rating is 1862 ± 88 .
- ▶ P beats Q_2 : P 's updated and final rating is 1888 ± 82 .
- ▶ To calculate final ratings for Q_2 , Q_3 , and Q_4 , first need to do surgery.

Implementation

- ▶ Calculations are done by replacing continuous laws with discrete laws on $\{0, 10, \dots, 3500\}$.
- ▶ Because integrands are functions of playing-strength difference, integrals are convolutions and may be rapidly calculated via the FFT.

Reports

- ▶ For each event, the website displays a summary report and a detailed report.

Sample Summary Report

ID	Name	Initial Rating	Point Change	Final Rating
5766	Bulatao, Jose G.	1797±58	-4	1793±52
5568	Cembura, Julianne	1500±450	-539	961±246
7355	Ching, Joe T.	1984±38	+2	1986±36
6655	Chiu, David	2050±66	+20	2070±49
5925	Collamore, Gil	1121±95	-126	995±59
5184	Conley, Denny	1463±38	+19	1482±34
5044	Cortesi, Tony	1139±90	-58	1081±54

- ▶ The numbers after the plus/minus signs are the standard deviations of the laws.

Sample Detailed Report

Boulard, Claude			Rating Change $1701 \pm 53 + 90 = 1791 \pm 41$		
Wins			Losses		
Point Change	Opponent's Rating	Opponent	Point Change	Opponent's Rating	Opponent
+30	1915 ± 50	Chen, Wei Teng	0*	1812 ± 34	Bhatia, Sonu
+20*	1812 ± 34	Bhatia, Sonu	-1	2016 ± 48	Maitra, Subhajt
+12*	1785 ± 61	Landsman, Alex	0	2189 ± 40	Wang, Yin
+12*	1785 ± 61	Landsman, Alex			
+10	1750 ± 34	Marczak, Slawomir			
+5	1629 ± 52	Jordan, Kip			
+2	1587 ± 67	Warrier, Sunil			
0	1366 ± 53	Sharma, Rajeev			

Sample Detailed Report (cont.)

Landsman, Alex			Rating Change $1776 \pm 64 - 35 = 1741 \pm 54$		
Wins			Losses		
Point Change	Opponent's Rating	Opponent	Point Change	Opponent's Rating	Opponent
+8	1630 ± 49	Baird, Jim	-21*	1761 ± 43	Boulard, Claude
			-21*	1761 ± 43	Boulard, Claude
			0	2170 ± 29	Chui, Lim Ming

Marczak, Slawomir			Rating Change $1752 \pm 37 - 7 = 1745 \pm 33$		
Wins			Losses		
Point Change	Opponent's Rating	Opponent	Point Change	Opponent's Rating	Opponent
+9	1753 ± 50	Baylies, Michael	-7	1771 ± 42	Boulard, Claude
+3	1625 ± 52	Jordan, Kip	-7	1798 ± 33	Bhatia, Sonu
			-5	1811 ± 42	Massarsky, Lev
			0	2015 ± 48	Maitra, Subhajit
			0	2189 ± 40	Wang, Yin

Comments on Detailed Report

- ▶ The rating system processes multiple matches between the same two players as a unit.
 - ▶ Total point change is distributed among the matches between the two players.
 - ▶ Indicated by an asterisk after the point change value.
 - ▶ E.g., Claude Boulard gained 20 points total for his one win and one loss to Sonu Bhatia and gained 24 points total for his two wins over Alex Landsman.
- ▶ The rating system uses different adjusted laws for the same opponent when processing different players.
 - ▶ E.g., Claude Boulard's adjusted rating is 1761 ± 43 when he played Alex Landsman, but 1771 ± 42 when he played Slawomir Marczak.
- ▶ Points gained by the winner of a match will hardly ever equal the number of points lost by the loser of a match.
 - ▶ E.g., Claude Boulard gained a total of 24 points for his two wins over Alex Landsman, but Alex lost 42 points for the same two matches.

Point Change per Match

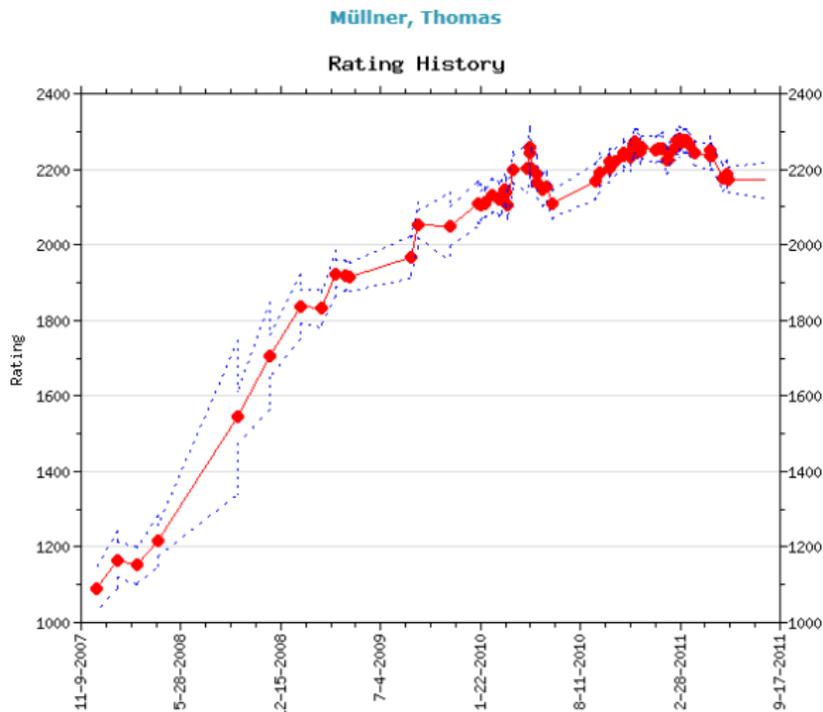
- ▶ The point change per match depends on the order that the rating system processes the matches, which is neither recorded nor shown.
- ▶ However, the total point change for the player for the event does not depend on the order that the rating system processes the matches.
 - ▶ General property of calculating a posterior or conditional probability in stages or iterated integrals.
- ▶ The dependence of the point change per match on the processing order makes intuitive sense:
 - ▶ Suppose we see a 1900 player beat a 2100 player. We will significantly increase our estimate of the rating of the 1900 player.
 - ▶ Next, suppose we see the same player beat another 2100 player. We will again increase our estimate of the player's rating, but not by as much as we did before.
- ▶ This occasionally confuses players, but they usually accept the explanation.

Website Features

- ▶ The Ratings Central website is copiously hyperlinked letting you jump
 - ▶ From a player to a summary report of an event they played in,
 - ▶ Then to their matches in the detailed report for an event,
 - ▶ Then to their opponent's matches in the same event,
 - ▶ Then to the entire event/rating history of the player or opponent.
- ▶ The website lets you search for players, events, and clubs by numerous attributes including name, ID, club, organization, state, country, rating, standard deviation, date played, and age.
 - ▶ Players may be sorted by name or rating/ranking.
 - ▶ Lists of players can show ratings as of a date in the past.

Website Features (cont.)

- ▶ The website can display the rating history for any player.



- ▶ You can click on a dot in the graph to go to the player's results for the event.

Priors

Prior Elicitation

- ▶ Priors come from the event directors.
- ▶ There are two types of priors for an event: event and player.
 - ▶ Event directors must set the event prior.
 - ▶ They may also set a prior for individual players, but are not required to.
- ▶ Setting a prior means specifying a mean and standard deviation.
- ▶ For a given unrated player, the system uses the player prior if it is set, otherwise, it uses the event prior.
- ▶ Following slides give excerpts from the instructions given to the event directors.
 - ▶ Full instructions are at www.ratingscentral.com/UnratedPlayers.php

Player Priors

- ▶ Excerpts from the instructions for player priors:
 - ▶ The prior standard deviation for a player measures how sure you are that you know that player's playing strength.
 - ▶ Here are some very rough guidelines: If you know an unrated player extremely well (e.g., they play at your club every week), then you might use a prior standard deviation of 50–75. . . .

Event Priors

- ▶ Excerpts from the instructions for event priors:
 - ▶ It is best to interpret the event prior mean and standard deviation as describing the range of unrated players at your event.
 - ▶ For example, if you think the unrated players range from 800 to 1400, then you would use the average of these two values (i.e., 1100) as the mean and the difference of these two values divided by four (i.e., 150) as the standard deviation.
 - ▶ ... you should interpret the range as being plus or minus two standard deviations, not three.
- ▶ Initially, I tried to interpret the range as being plus or minus three standard deviations.
 - ▶ So, told directors to divide the difference by six (giving 100 in the example above).
 - ▶ This always produced standard deviations that were too small.
 - ▶ People can't imagine three-sigma events.

Leagues in Lower Austria

- ▶ Lower Austria has 1822 players on 414 teams playing in leagues organized by playing level (division) and region. They also have many tournaments and additional leagues organized by age.
- ▶ Each team plays 16–22 times a year and fields 3–4 players for each team match.
- ▶ When a team plays, each player plays 2–4 matches.
 - ▶ Rather different from the U.S. where players generally play more matches than that in a single event.
- ▶ They use a prior for each league division in each region (40 different combinations). They also have separate priors for juniors by age and gender.
- ▶ The priors for the main leagues are determined once a year from the statistics of the rated players in each league division/region.

Confusions and Misconceptions

My Rating is Lower

- ▶ In 1999, when we originally announced the proposed new system to the USATT membership, we posted on the Web all the tournament results in the U.S. for 5½ years (15,549 players, 330,079 matches) with both the USATT ratings and the ratings calculated by the new system.
- ▶ Generally, if a player's rating in the new system was higher than in the USATT system, the player liked the new system. If it was lower, they did not like it.
- ▶ Getting tired of illogical complaints, we decided to raise all the ratings in the new system by 100 points.
 - ▶ This significantly decreased the number of complaints.
 - ▶ Ratings are relative, so the change made no real difference.

Standard Deviation Measures Consistency

- ▶ Sometimes people think the standard deviation measures how consistent a player is.
- ▶ It doesn't. There is nothing in the model that measures consistency.
- ▶ People aren't familiar with using probability to model uncertainty.

Won't Be Accepted Because Inscrutable

- ▶ Back when we were trying to get USATT to adopt the system, opponents argued that players wanted a system where they could check the rating calculations themselves.
 - ▶ Actually, it is much harder than players believe to check the USATT rating calculations because the system is more complicated than players think it is and players do not know what ratings will be given to unrated players and what adjustments the system will make.
- ▶ Once we got people actually using the new system, none of the players complained about not being able to check the calculations.
 - ▶ Since the ratings the system produces agree with what the players think they should be rated (based on their results) and thus seem fair, players don't feel the need to calculate the values.
 - ▶ Ratings are posted very soon after the event ends, so players don't have to wait days or weeks to find out their new rating.

Reference

Reference



Marcus, D. J. (2001),

New Table-Tennis Rating System.

Journal of the Royal Statistical Society: Series D (The Statistician), 50: 191–208. doi: 10.1111/1467-9884.00271

www.ratingscentral.com/Doc/NewTTRS.pdf