IMPROVING THE RANKING SYSTEM FOR WOMEN'S PROFESIONAL TENNIS

By

MAYA MILADINOVIC

Advisor

DR. ERICH FRIEDMAN

A senior research paper submitted in partial fulfillment of the requirement
for the degree of Bachelor of Science
in the Department of Mathematics and Computer Science
in the College of Arts and Science
at Stetson University
Deland, Florida

Spring Term 2008

TABLE OF CONTENTS

TABLE OF CONTENTS……………………………………………………………….…...2

LIST OF TABLES………………………………………………………………..….3

LIST OF FIGURES…………………………………………………………………4

ABSTRACT…………………………………………………………………………5

1. INTRODUCTION……………………………………………….……….......6

2. MODEL……………………………………………………..………..…..…..8

2.1. Normal Distribution…………………………………….……….……8

2.2 Likelihood Function ……………………………….…………...….…9

2.3 Maximizing Likelihood Estimation………………………..……..….10

2.4 Maximum Likelihood Estimation of Parameters…………...……………..…11

2.5 Method………………...………………………………………………..11

2.6 Software……………………………………………..…………………12

2.7 Limitations of the Model and Software……………………………...13

3. RESULTS…………………………………………………….………………14

3.1 New Ranking……………………………...……………..……......14

3.2 Comparison of Rankings……………………...……………...…………17

4. FUTURE WORK……………………………………………………...……18

APPENDIX I- Current Ranking System……………………………………………19

APPENDIX II- Mathematica Code………………………………………...……20

REFERENCES……………………………………………………..…………23

BIOGRAPHICAL SKETCH…………………………………………………..…24

LIST OF TABLES

1. The Results of Mathematica's ratings for a New Ranking System.......................14

2. Current Women's Professional Tennis Ranking System for Singles Play…..….19

LIST OF FIGURES

1. Likelihood Function Surface……………………………………………………..…10

2. Normal Distribution of players A and B ………………………………………………12

ABSTRACT

IMPROVING THE RANKING SYSTEM FOR WOMEN’S PROFESIONAL TENNIS

By

MAYA MILADINOVIC

December, 2008

Advisor: Dr. Erich Friedman
Department: Mathematics and Computer Science

The current ranking system for women’s professional tennis system (WTA) is based on how far a player advances in a tournament play. Players are ranked on the basis of their total points (round points) from different scales of points, depending on what type of tournament they have played in. The tier of the tournament is the major factor for calculating a player's round points.

We believe this ranking system is not the best method to rate a top player. Our goal in this project is to make an improvement in the women’s tennis ranking system that will be based on “quality points”, the points that are based on the opponents that a player beats in a tournament. The ranking improvement can be done by maximizing the likelihood function, a statistical method used to calculate the best way of fitting a mathematical model to our data. The probability function where a player beats an opponent can be written as $P(Z_1 > Z_2)$. Multiplying all the probability functions from our data, between players and results of their matches, will give us a likelihood function, and maximizing the L function will make the data “more likely”. The new ranking system better predicts a ranking for the players than the current system, because has fewer upsets that are based on the same data that we used for both systems.

CHAPTER 1

INTRODUCTION

The current ranking system for women's professional tennis reflects and is based on the performance in tournament play. This ranking system is a 52-week, cumulative system where the number of tournament results is restricted to 17 tournament results for singles play. The method that is used to rank the players is only based on how far a player advances in a tournament. Players are ranked on the bases of their total points, which are called the round points, and they must have at least three valid tournaments to appear on the WTA ranking list [A1].

The current ranking system is based on 7 different types of tournaments during the year, and each type is weighted differently depending on round of a tournament. The scale of significance is; Grand Slam, Year End Championship, tier I, tier II, tier IIIA, tier IIIB, and tier IV. Challenger and Satellite events are not regarded as part of the regular tier system, although they also earn the player points. In appendix I is a table of points earned in all types of tournaments. In the case of round points, there exist two cases: the gap between the Grand Slams and every other tournament, and the gap between tier II and tier III tournaments. The Slams are overstated for reasons not having directly to do with their difficulty, because they have large draws spread out over a long period, it is "easier" to win a typical match at a Grand Slam, where you are rested and have lower ranked opponents in the first few rounds because of a very large draw with many unseeded players, than at lesser tournaments, (Tier II and below) where your opponent is sure to be ranked higher and you have had less rest. The gap between tier II and tier III tournaments is too small, because there are tier II events where every player who earned direct entry was in the top 25, and in addition there are tier III events where no one is in the top 20.

No matter what the tier of the event, points are awarded based on wins. If a player loses an opening match, she receives only one round point, no matter whether she loses in the round of 128 or the Round of 16. The only exemptions are the Grand Slams; a first round loss is worth two points; at the chase Championships, where only sixteen players play, a first-round loss is worth 54 points. In the current system players need to “defend” their earned points, because they expire after a year. If a player wants to keep her point total, she must come up with new points to replace those which expire. If a player earns a lot of points in a given week last year, she must play the same tournament or another one and earn new points or else she loses them.

We are trying to improve the current ranking system and design a new one, which will be based on “quality points” that are based on the opponents that a player beats in a tournament. If a player beats the number 1 player in the world, she gets the quality points whether it is the first round or the last round of a tournament. For quality point calculations, the rankings and a seat of opponents that a player beats, is the major factor. These quality points will be rated equally no matter what type of tournament a player plays.

CHAPTER 2

MODEL

2.1 Normal Distribution

The normal distribution is pattern for the distribution of a set of data which follows a bell shaped curve. Normal distribution may be defined by two parameters, the mean (μ) and variance (σ^2). Standard deviation, being the square root of that quantity, therefore measures the dispersion of data about the mean. The normal distribution maximizes information entropy among all distributions with known mean and variance, which makes it the natural choice of underlying distribution for data summarized in terms of sample mean and variance. [A5] The normal distribution is the most widely used family of distributions in statistics and many statistical tests are based on the assumption of normality. In probability theory, normal distributions arise as the limiting distributions of several continuous and discrete families of distributions.

The continuous probability density function of the normal distribution is

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right),$$

where $\sigma > 0$ is the standard deviation, μ is the expected value, and

$$\varphi(x) = \varphi_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

is the density function of the "standard" normal distribution, i.e., the normal distribution with $\mu = 0$ and $\sigma = 1$.

It can be proved that the linear combination of normally distributed random variables is a normally distributed random variable.

Theorem: Let X be a random variable with $N(\mu_1, \sigma_1{}^2)$ distribution and let Y be a random variable with $N(\mu_2, \sigma_2{}^2)$ distribution, then random variable $Z = aX + bY$ has $N(a\mu_1 + b\mu_2, a^2\sigma_1{}^2 + b^2\sigma_2{}^2)$ distribution.

Proof: Suppose $X \sim N(\mu_1, \sigma_1{}^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ and $Z = aX + bY$.

Let $\varphi_x(t), \varphi_y(t)$ and $\varphi_z(t)$ be the characteristics functions of random variables X, Y and Z respectively. Then,

$$\varphi_x(t) = e^{\left(itt\mu_1 - \frac{t^2\sigma_1{}^2}{2}\right)} \text{ and } \varphi_y(t) = e^{\left(itt\mu_2 - \frac{t_2\sigma_2{}^2}{2}\right)}$$

Then by plugging into $Z = aX + bY$ we get,

$$\varphi_z(t) = \varphi_x(at)\varphi_y(bt) = e^{(it(a\mu_1 + b\mu_2)} - t^{2\left(\frac{a^2\sigma_1^2 + b^2\sigma_2^2}{2}\right)}$$

Since the characteristic function determines the distribution uniquely random variable Z has $N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$ distribution. QED [A5]

2.2 Likelihood Function

The likelihood of a set of data is the probability of obtaining that particular set of data, given the chosen probability distribution model. It indicates how likely a particular population is to produce an observed sample. In general, the likelihood function of the random sample is written as

$$L(\theta; x_1, x_2, ..., x_n) = \prod_{i=1}^{n} \mathrm{P}(X_i = x_i; \theta)$$

where a random sample $X_1, X_2, ..., X_n$ of some discrete random variable has probability distribution $\mathrm{P}(X = x; \theta)$, and θ represents the vector of parameters. [A2]

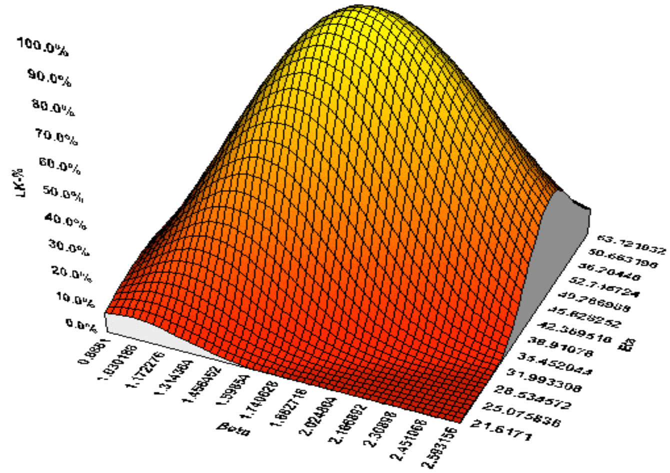

Figure 1: Likelihood function surface

This graphic gives an example of a likelihood function surface plot for a two-parameter Weibull distribution, a continuous probability distribution. The values of the parameters that maximize the likelihood function are called MLE estimates for the distribution's parameters.

2.3 Maximum Likelihood Estimation

Maximum likelihood estimation begins with writing a mathematical expression known as the likelihood function of the sample data. This expression contains the unknown model parameters. The values of these parameters that maximize the sample likelihood are known as the Maximum Likelihood Estimates (MLEs). [A3]

The value of θ that maximizes L is called the maximum likelihood estimator. Modeling real world data by estimating maximum likelihood gives a way of tuning the free parameters of the model to provide an optimum fit. In addition, for a fixed set of data and underlying probability model, maximum likelihood picks the values of the model parameters that make the data "more likely" than any other values of the parameters would make them. [B2]

2.4 Maximum Likelihood Estimation of Parameters

Suppose we have $x_1,...,x_n$ normally distributed independent variables with expectation μ and variance σ^2. Then the continuous joint probability function of n independent random variables is written by

$$f(x_1,...,x_n;\mu,\sigma) = \prod_{i=1}^{n}\varphi_{\mu,\sigma^2}(x_i) = \frac{1}{(\sigma\sqrt{2\pi})^n}\prod_{i=1}^{n}\exp\left(-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right)$$

In the method of maximum likelihood estimation, the values of μ and σ that maximize the likelihood function are taken as estimates of the population parameters μ and σ. In addition, the value of μ that maximizes the likelihood function with σ fixed does not depend on σ. Therefore, we can find that value of μ, then substitute it for μ in the likelihood function, and find the value of σ that maximizes the resulting expression. [A2]

2.5 Method

The mathematical method that we are using for calculating a new ranking system is based on maximizing a likelihood function L. The first step is to find a value of ratings of players that maximizes the L function, which is the probability that we get from our data with all the matches between the top 62 tennis players. The probability function where a player beats an opponent can be written as $P(Z_1 > Z_2)$. We can get this probability function from a normal distribution,

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

where x represents players that we used from the current ranking system, standard deviation of 1, and different means that best fit the data.

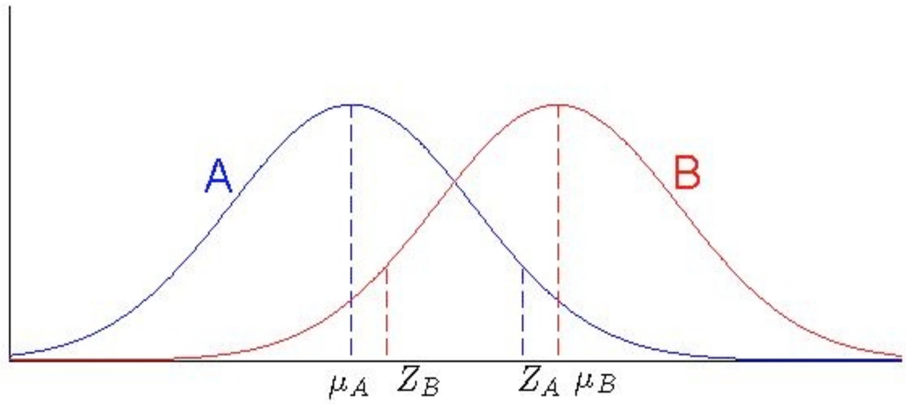

Figure 2: Normal distribution of players *A* and *B*

$$Z_A \sim N(\mu_A, 1)$$
$$Z_B \sim N(\mu_B, 1)$$
$$Z_A - Z_B \sim N(\mu_A - \mu_B, 2)$$

The probability function that a player A beats a player B, where A and B have a normal distribution $N(\mu_A, 1)$ and $N(\mu_B, 1)$, can be written as:

$$\begin{aligned} & P(A > B) \\ & = P(A - B > 0) \\ & = P\left(\frac{A - B - (\mu_A - \mu_B)}{\sqrt{2}} > \frac{0 - (\mu_A - \mu_B)}{\sqrt{2}}\right) \end{aligned}$$

where $A - B \sim N(\mu_A - \mu_B, 2)$, then the likelihood function is:

$$L(\mu_A, \mu_B, \mu_C, \ldots) = P\left(Z > \frac{\mu_B - \mu_A}{\sqrt{2}}\right) \cdot P\left(Z > \frac{\mu_C - \mu_B}{\sqrt{2}}\right) \cdot \ldots$$

This L function shows us the probability where player A beats a player B, and player B beats a player C, and so on. Multiplying all the probability functions, between the match results from our existing data, we attempt to maximize the likelihood function i.e. to find the values of $\mu_A, \mu_B, \ldots$ that best fit the data.

2.6 Software

For calculating a new women's professional ranking system, we used the Mathematica program. First of all, we started off by taking the top 62 players from the current ranking list, and we found all data, matches that the players played between each other from year 2002 to 2007. We wrote a function error that calculates the probability P(X>0), where $X \sim N(\mu,1)$. Furthermore, we list all the matches from our data with winners first and losers second. We set up one of the players to have a fixed μ value, because otherwise there would not be a unique solution, and then we make the list of players, giving them some initial rating. Subsequently, we write the likelihood function, and as we said before, we multiply all the probability functions where player A beats player B, and so on. Then, we maximize the L function using Mathematica's built-in FindMaximum command to get a new ranking system (Appendix II).

Limitations of the Model and Software

The software that we used for finding a new ranking system would not converge unless every player has won and lost at least one game. For some players that had only played one match in the data that we used, the system would send those players to –oo. By taking out those players we ended up with more players with only one win or one loss.

Mathematica had trouble approximating the rankings of 150 players, so we had to take out all the players that have played only one game. Therefore, we were left with 62 players, from the current ranking list that we used which have won and lost at least one game.

CHAPTER 3

RESULTS

3.1 New Ranking

Players	New Rating	New Ranking	Current Rating	Current Ranking
Justine Henin	51.27	1	6105	1
Jelena Jankovic	50.91	2	4095	3
Venus Williams	50.87	3	2781	6
Serena Williams	50.66	4	2297	9
Ana Ivanovic	50.30	5	4216	2
Nadia Petrova	50.26	6	2198	11
L. Davenport	50.15	7	930	30
Casey Dellacqua	50.12	8	542	53
Sybille Bammer	50.09	9	1434	19
S. Kuznetsova	50.08	10	3905	4
Na Li	49.99	11	820	33
Marion Bartoli	49.96	12	1915	12
Maria Sharapova	49.94	13	3601	5
M. Domachowska	49.93	14	415	78
Ana Chakvetadze	49.92	15	2665	7
Shahar Peer	49.87	16	1245	17
Nicole Vaidisova	49.87	17	1489	15
Alicia Molik	49.86	18	473	65
Elena Likhovtseva	49.86	19	433	75

Daniela Hantuchova	49.86	20	2305	8
Camille Pin	49.84	21	454	70
Flavia Pennetta	49.82	22	820	33
Domini Cibulkova	49.81	23	618	46
Virginia Pascual	49.81	24	500	63
M. Shaughnessy	49.80	25	529	55
Eleni Daniilidou	49.77	26	758	36
Laura Granville	49.67	27	443	74
C. Wozniacki	49.66	28	533	54
Martina Muller	49.65	29	517	61
Fran. Schiavone	49.64	30	1150	23
Emilie Loit	49.63	31	584	49
Aravane Rezai	49.63	32	527	56
Lourdes Lin	49.63	33	390	86
Olga Govortsova	49.61	34	638	45
Nathalie Dechy	49.61	35	492	64
Danira Safina	49.60	36	1458	16
Victoria Azarenka	49.58	37	1107	24
Gisela Dulko	49.55	38	718	38
Olga Poutchkova	49.54	39	358	97
Maria Kirilenko	49.53	40	910	32
Meng Yuan	49.50	41	387	87
Jill Craybas	49.49	42	424	77

Tamira Paszek	49.47	43	686	41
Virginie Razzano	49.47	44	980	27
K. Bondarenko	49.45	45	645	43
Julia Vaculenko	49.45	46	760	35
Timea Bacsinszky	49.41	47	367	93
Kaia Kanepi	49.37	48	454	69
E. Makarova	49.35	49	413	79
Elena Vesnina	49.32	50	556	51
Patty Schnyder	49.30	51	1152	14
P. Parmentier	49.29	52	542	53
Sofia Arvidsson	49.28	53	465	67
Samantha Stosur	49.23	54	375	89
Elena Dementieva	49.22	55	1642	13
Vera Dushevina	49.12	56	696	40
Sania Mirza	49.04	57	933	29
Laura Granville	48.93	58	367	91
Virginia Ruano	48.81	59	429	76
K. Srebotnik	48.80	60	910	28
Ai Sugiyama	48.57	61	677	42
Milagros Sequera	48.28	62	343	99

Table 1: The results of Mathematica's ratings for a new ranking system

3.2 Prediction of a new ranking system

To compare our new ranking system to the current system, we calculated how many upsets exist in both. An upset occurs in the match between players A and B when a player B, who is unseeded player, beats a player A, who is a seeded player. For the new ranking system we got 77 upsets, and for the current ranking system we got 82 upsets, which means that our ranking system better predicts a ranking for the players based on the same data that we used for both systems. The results of upsets can be seen in Mathematica code in appendix II.

We can see a big difference between our ranking system and the current one, and that is because the current ranking system is based only on one year tournament play, and points expire after a year. If a player wants to maintain her point total, she must come up with new points to replace those which expire, which is called defending points. If you earned a lot of points in a given week last year, you must play the same tournament and earn new points, or else you lose them. In our ranking system, we used data of matches from 2002 to 2007, without point defending, just based on wins and losses.

CHAPTER 4

FUTURE WORK

There are many other things that could be included in the new ranking system. The score of the match should be one of the major factors in rating the players. It should matter if a player A ranked 30 on the list, beats an opponent B ranked 3, with a score 7/6 6/7 7/6, then A should get more quality points for beating B. In addition, if a player B loses to a player A in straight sets 6/1 6/0, B should lose some quality points. We could also try to predict not only the winner but the probability of them winning by finding the function that will predict the probability of player A beating a player B in some number of matches. In addition, we can use more methods besides the number of upsets to predict which ranking system is "better". We could also apply our mathematical method to doubles play in women's professional tennis.

APPENDIX I

Current women's professional tennis ranking system for single play

Description	Winner	Finalist	Semifinalis t	QF	R16	R32	R64	R128
Grand Slams	1000	700	450	250	140	90	60	2
SE Champs	750	525	335	185	105	-	-	-
Tier $3milion	500	350	225	115	70	45	30	1
Tier $2milion	465	325	210	115	65	40	25	1
Tier $1.5 mi.	430	300	195	110	60	35	1	-
Tier 650,000	300	215	140	75	40	1	-	-
Tier 600,000	275	190	125	70	35	20	1	-
Tier 225,000	165	115	75	40	20	1	-	-
Tier 175,000	140	100	65	35	20	10	1	-
Tier 145,000	115	80	50	30	15	1	-	-
ITF 100,000	75	55	40	20	10	1	-	-
ITF 75,000	65	45	29	16	8	1	-	-
ITF 50,000	45	32	20	12	6	1	-	-
ITF 25,000	25	17	12	7	4	1	-	-
ITF10,000	6	4	3	2	1	-	-	-

Table 2: Points awarded for various tournaments.

APPENDIX II

Mathematica Code for calculating a probability functions between players

```
error[x_]:=1-Erf[-Infinity,x/Sqrt[2]]/2

games={{HH,OP},{EM,AS},{HH,EM},{DS,KB},{HH,DS},{BB,SE},{VZ,
EL},{JJ,OG},{AC,CW},{SB,MS},{AB,DD},{VD,EM},{II,AR},{BB,LS}
,{SW,VZ},{SB,ED},{VW,AB},{II,VD},{HH,DS},{SW,BB},{JJ,SB},{V
W,II},{HH,SW},{VW,JJ},{HH,VW},{NP,RR},{MK,KS},{VA,DC},{SK,C
P},{SM,LG},{TP,FS},{SH,CD},{JV,MK},{CH,SM},{TP,PS},{SP,NV},
{AR,SH},{SK,VA},{CH,TP},{SP,AR},{CH,SP},{SK,CH},{HH,SK},{MS
,VZ},{OP,MS},{DD,VD},{DC,SK},{CP,ED},{LG,SM},{CD,SH},{AM,MS
},{SW,HA},{SH,LL},{EL,HP},{CH,AB},{TG,AR},{NP,VRP},{TG,NP},
{VA,CH},{SH,EL},{SW,II},{AM,SW},{DS,VR},{SH,LI},{TG,CH},{SH
,TG},{AM,DS},{SH,AM},{SH,HH},{HH,JJ},{JJ,ED},{HH,LD},{HH,SP
},{LD,PS},{ED,AR},{JJ,SK},{JJ,NV},{AR,MK},{ED,VZ},{PS,BB},{
LD,KS},{SP,FS},{HH,AS},{AS,TG},{SP,MM},{FS,SM},{BB,KS},{AR,
LS},{VRP,AR},{LI,CH},{VR,SW},{MM,BB},{VR,TG},{SH,NP},{SH,SM
},{NP,NV},{VW,SW},{SW,FS},{VW,HA},{NP,LG},{SM,BB},{ED,LD},{
LD,NN},{NN,ED},{ED,PS},{SW,PS},{SW,TG},{PS,HA},{SK,LD},{SK,
NP},{LD,VW},{NP,HH},{FS,AS},{HH,DS},{VW,AM},{SH,JJ},{JJ,SW}
,{SH,HH},{SW,NV},{JJ,CD},{SH,ED},{HH,SWH},{SW,VA},{NV,AS},{
CD,AM},{JJ,VR},{SH,EV},{ED,SP},{SH,AR},{HH,FS},{SW,MY},{NV,
AM},{CD,PS},{AM,YS},{JJ,EG},{SH,LD},{EV,JC},{ED,AMG},{AR,TG
},{SWH,SB},{HH,OP},{II,HA},{HA,AR},{II,VW},{AR,NP},{HA,MK},
{II,CW},{VW,MD},{AR,SK},{NP,EM},{HA,VRP},{MK,CH},{II,KS},{M
D,NL},{VW,SM},{AR,PP},{EM,YS},{YS,EG},{NL,EV},{NP,AK},{AK,N
P},{HA,AC},{OG,VRP},{MK,AM},{LL,TG},{KS,AR},{CW,AB},{MD,SA}
,{SM,TB},{TB,SM},{VW,CP},{CP,MK},{MY,CH},{AR,VR},{SH,RR},{T
P,RR},{VZ,TP},{II,AR},{CH,LG},{LDL,TG},{KK,VR},{AM,KK},{VR,
AB},{AH,VR},{HA,EL},{LI,LDL},{SH,TG},{VZ,II},{HA,AH},{NL,DS
},{SH,VZ},{CH,PS},{SH,CH},{SW,NA},{SW,LS},{LS,AM},{NV,ED},{
SP,SK},{NV,KS},{SK,MK},{SP,TG},{JJ,VA},{SW,NP},{NP,GD},{VA,
BB},{JJ,VRP},{MK,JV},{HH,JJ},{JJ,NV},{HH,SW},{NV,TG},{JJ,BB
},{SW,DS},{NV,SS},{BB,ED},{JJ,VW},{SW,MK},{MK,BB},{DS,FS},{
SB,LI},{HH,MS},{SH,II},{SH,CH},{II,SK},{SH,PS},{CH,LS},{SK,
SP},{II,AMG},{PS,KK},{CH,AS},{LS,AM},{SP,KS},{AMG,HA},{SK,N
V},{SK,DS},{NV,VW},{FS,SK},{DS,SH},{VW,PS},{NV,AM},{AM,JJ},
{AR,NV},{PS,JV},{DS,KS},{FS,FP},{SK,LI},{FP,LI},{LI,SH},{AM
,VD},{JJ,BB},{AR,AS},{BB,HH},{BB,MK},{MK,CH},{HH,SW},{HH,PS
},{SW,HA},{BB,JJ},{MK,LG},{BB,SP},{JJ,LS},{HA,KS},{PS,AB},{
AB,AM},{JJ,LS},{II,NV},{VW,SK},{VW,SH},{SK,TP},{NV,AM},{NP,
II},{SH,AS},{TP,ED},{SK,AR},{AM,MS},{NV,VA},{NP,VRP},{II,AR
},{AM,SH},{SH,ED},{SH,FP},{AM,II},{ED,EL},{FP,SP},{JJ,VW},{
AMG,SH},{SP,CH},{II,DS},{LI,NV},{HH,HA},{LI,SK},{HH,CH},{CH
,EV},{HA,KS},{LD,AM},{LD,SK},{AM,EL}};
```

```
players = Union[Flatten[games]]

rank = Table[0,{i,Length[players]},{j,2}];
For[i=1,I <=Length[players],i++,
 rank[[i]] = {players[[i]], 50-0.01*i }
 ];
```

The list of all 62 players from the current system that we used:

```
{AB,AC,AM,AR,AS,BB,CD,CH,CP,CW,DC,DD,DS,ED,EM,EV,FP,FS,GD,H
A,HH,II,JC,JJ,JV,KB,KK,KS,LD,LDL,LG,LI,LL,LS,MD,MK,MM,MS,MY
,NL,NN,NP,NV,OG,OP,PP,PS,SA,SB,SH,SK,SM,SP,SS,SW,TB,TG,TP,V
A,VD,VR,VRP,VW,VZ}
```

The list of all 62 players with initial guesses of μ that we think best fits the data:

```
{{AB,49.99},{AC,49.98},{AM,49.97},{AR,49.96},{AS,49.95},{BB
,49.94},{CD,49.93},{CH,49.92},{CP,49.91},{CW,49.9},{DC,49.8
9},{DD,49.88},{DS,49.87},{ED,49.86},{EM,49.85},{EV,49.84},{
FP,49.83},{FS,49.82},{GD,49.81},{HA,49.8},{HH,49.79},{II,49
.78},{JC,49.77},{JJ,49.76},{JV,49.75},{KB,49.74},{KK,49.73}
,{KS,49.72},{LD,49.71},{LDL,49.7},{LG,49.69},{LI,49.68},{LL
,49.67},{LS,49.66},{MD,49.65},{MK,49.64},{MM,49.63},{MS,49.
62},{MY,49.61},{NL,49.6},{NN,49.59},{NP,49.58},{NV,49.57},{
OG,49.56},{OP,49.55},{PP,49.54},{PS,49.53},{SA,49.52},{SB,4
9.51},{SH,49.5},{SK,49.49},{SM,49.48},{SP,49.47},{SS,49.46}
,{SW,49.45},{TB,49.44},{TG,49.43},{TP,49.42},{VA,49.41},{VD
,49.4},{VR,49.39},{VRP,49.38},{VW,49.37},{VZ,49.36}}
```

Results:

```
L=Apply[Times,Map[error[(#[[2]]-
#[[1]])/Sqrt[2]]&,(games/.SH→50),1]];

results=FindMaximum[10^100
L,rank,PrecisionGoal→3,AccuracyGoal→3]
```

{1.85558×10^34,{AB→49.1195,AC→49.9217,AM→49.8638,AR→49.6342
,AS→48.5707,BB→49.9567,CD→50.128,CH→49.5548,CP→49.8452,C
W→49.6623,DC→49.8145,DD→49.7782,DS→49.6029,ED→49.2276,EM
→49.3597,EV→49.32,FP→49.8202,FS→49.6446,GD→49.5504,HA→4
9.8533,HH→51.2601,II→50.3025,JC→49.4982,JJ→50.9177,JV→49
.4522,KB→49.4545,KK→49.3757,KS→48.2838,LD→50.1547,LDL→49
.6304,LG→48.9379,LI→49.9904,LL→49.6397,LS→49.0291,MD→49.
9257,MK→49.5327,MM→49.6562,MS→48.8075,MY→49.6765,NL→49.8
631,NN→49.6107,NP→50.2689,NV→49.8743,OG→49.6124,OP→49.54

```
19,PP→49.2918,PS→49.3035,SA→49.2844,SB→50.0976,SH→49.5,S
K→50.0822,SM→49.0431,SP→49.8775,SS→49.236,SW→50.6661,TB
→49.4119,TG→48.6723,TP→49.4793,VA→49.5891,VD→49.1292,VR
→49.4721,VRP→48.8133,VW→50.8712,VZ→49.2152}}
```

New Upsets:

```
upsets=Length[Select[games/.results[[2]],#[[1]]<#[[2]]&]]
77
Length[games]
272
```

Current System:

```
oldranking={{HH→1},{II→2},{SK→3},{JJ→4},{SH→5},{VW→6},{
AC→7},{HA→8},{BB→9},{SW→10},{NP→11},{TG→12},{ED→13},{PS
→14},{NV→15},{DS→16},{SP→17},{AM→18},{SB→19},{AR→21},{A
B→22},{FS→23},{VA→24},{VZ→25},{VR→26},{MK→27},{KS→28},{
SM→29},{LS→31},{MK→32},{FP→33},{LI→34},{JV→35},{MS→37},
{GD→38},{TG→39},{VD→40},{TP→41},{AS→42},{KB→43},{LD→44}
,{OG→45},{DC→46},{AM→47},{KK→48},{EV→51},{PP→53},{CW→54
},{MS→55},{AR→56},{CD→57},{SA→67},{CP→71},{VRP→76},{JC→
77},{MD→78},{EM→79},{AR→80},{LDL→86},{MY→87},{SS→89},{L
G→91},{TB→93},{OP→97}}
```

Current upsets:

```
oldupsets=Length[Select[games/.Flatten[oldranking],#[[1]]>#
[[2]]&]]
82
```

REFERENCES

A. OFFLINE

[A1] Robert Waltz, *Women's Tennis Ranking FAQ*, 2001 Edition

[A2] Thomas A. Severini, *Likelihood Methods in Statistics*, Oxford Statistical Science Series, Oxford University Press, USA

[A3] A. W. F. Edwards (1972). Likelihood: *An account of the statistical concept of likelihood and its application to scientific inference*, Cambridge University Press. Reprinted in 1992, expanded edition, Johns Hopkins University Press.

[A4] A paper on the history of Maximum Likelihood: Aldrich, John (1997). *"R.A. Fisher and the making of maximum likelihood 1912-1922"*. *Statistical Science* 12 (3): 162-176. doi:10.1214/ss/1030037906.

[A5] Pavle Mladenovic, book on: *Probability and Statistics,* Faculty of Mathematics, Belgrade University, 2005

[A6] Donald A. Berry, Bernard W. Lindgren, Statistics : *Theory and Methods*, Belmont, California : Duxbury Press, 1996.

B. ONLINE

[B1] Professional Tennis Network Beta; Sony Ericsson WTA Tour, *How the WTA Rankings Work*; http://www.sonyericssonwtatour.com/2/rankings/singles_numeric.asp 2007

[B2] NIST/SEMATECH, *e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/ March, 2008.

BIOGRAPHICAL SKETCH

Maya Miladinovic is a senior at Stetson University. She is an international student, from Serbia, Belgrade (formerly Yugoslavia). She came to Stetson University because of a tennis scholarship, which is her big passion besides mathematics. She started playing tennis since she was 6 years old, and she was a "huge" according to her coach. She played many tournaments in and around her country. Because of her financial situation, she did not continue as a professional tennis player, although she was pretty close to getting on the WTA ranking list. At the end of 2003, she thought her tennis career was over, and she was preparing to enter a Mathematical University in Belgrade.

Her friend Mili Milovanovic, who is also from Serbia, and her coach Sasha Schmid called her from Stetson University, and invited her if she would like to join them on the team, because they needed one more player for the team.

She likes drawing very much, and she has about 10 big drawings that ended up on exhibit last summer in Serbia. Another of her passions is traveling, especially when her family and friends are involved. Her husband Alex, who is also Serbian, and she got married last year in Belgrade, and they both live in Orlando.

Maya is very close to her family and her brother Dusan, who is her best friend. It has been hard not seeing her family for about 10 months every year since 2004. They are all coming to her graduation, and she is very excited about it.