ISSN 1512-1232

UDC 519.876.5

MODELING OF IN-PLAY TENNIS ODDS

Victor Khutsishvili

Archil Eliashvili Institute of Control Systems of Georgian Technical University, Tbilisi, Georgia
e-mail: otariko@yahoo.com

Abstract:
A computer model of in-play tennis odds automatic calculation has been developed. It uses the so-called root probabilities - the probabilities of winning the current point on own and opponent's serve. In the model, the roots are identified on the basis of reliable odds given before the match starts, and are adjusted in a certain way depending on the result of the current point. The illustrations of concrete calculations are given.

Keywords: *tennis match, in-play odds, modeling, root probabilities, cut-off, odds-probabilities transition.*

1. INTRODUCTION

The problem of calculating the odds of events accompanying the match of professional tennis players is interesting to bookmakers as well as their customers. It is undoubtedly relevant for both mentioned parties participating in a kind of competition for predicting the results of a tennis match. By popularity in the field of betting on sports events, tennis rightfully occupies an honorable second place after football. This fact was the motive of our research. A preliminary version of this article was published in Russian [1].

2. FORMULATION OF THE PROBLEM

From the view of the fact that the five-set matches are played by only men during the only 8 weeks of the year, except for the matches for the Davis Cup, we limited ourselves to modeling three-set matches. Specifically, we consider the following problem. The odds announced by the bookmaker on the victory/defeat of the first player in a pair of rivals, the total – the value with which the total number of played games is compared and the odds of this number will be smaller/larger than total – in all 5 numbers are given. In addition, the second group of odds can be specified – to the exact score of the sets and to win/defeat with a given handicap – in all 7 numbers. If the second group of odds is not specified, it should be determined, and also other odds of interesting events should be calculated, and not only before match start moment, but also during the game – in-play. The division of odds into groups is due to the fact that the markets for odds from the first group are more popular, a huge amount of statistical material has been accumulated for them and they have been studied thoroughly. In the odds from the second group bookmakers are less confident, so often the task is to determine the odds of the second group, based on the reliable odds of the first. Obviously, bookmakers are even less confident in the in-play odds. This is due to the fact that the volume of statistical data is dispersed in numerous variants of the scenario of the match. Probably, that's why about 90% of all tennis bets are fixed during the game. This fact makes the problem of operative calculation of in-play odds even more actual.

It should be accented, that our problem is not to determine the odds of the first group - the basic odds. For this it is necessary to have information about current ratings of the players, their

ISSN 1512-1232

previous results according to cover (grass, ground, hard) and so on. This information has been already processed by bookmakers and is reflected in the basic odds, announced by them. The other odds, including in-play ones, can be called the derivatives. So our problem is to determine the derivatives from the given basic odds.

3. Splitting the problem into two components

In view of the fact that the odds for the events are approximately inversely proportional to the probabilities of these events, it seems natural to break down the problem of modeling the odds into two components. This is the modeling of the probabilities of events in the process of a tennis match using explicit rules and formulas of probability theory and the modeling the transition from odds to probabilities and back. It should be noted that the relationship of the odds with the corresponding probabilities is nontrivial. It was investigated in the author's papers [2], [3] and the results obtained in them allow us not to be distracted in this article on the problem of qualitative modeling of the transition odds-probabilities and back, but to concentrate on modeling tennis probabilities.

Briefly, in [2] the new H-model of the transition from odds to probabilities and back is substantiated, its advantages over the classical E-model, in which the payout e – the product of the event odd k by its probability p – is independent of p, are listed. Based on the H-model, which the authors of [4] called Khutsishvili's theory, and which they successfully used in the same work, the phenomenon of the favorite-longshot bias, first noted in [5], is explained. As a result, the fundamental property of optimal odds was revealed, which opens the possibility of a qualified transition from the odds of events to their probabilities and back. It is presented as an axiom which states that for optimal odds the payout $e(p) = p \cdot k(p)$ is an increasing function of probability. It should be noted that the authors of the recently published paper [6] called the H-model a "power model" and showed that it is the best among known to them. In [3] the H-model is generalized and applied to various concrete situations. The generalization is due to the necessity of taking into account tax rates on bets. For example, in Georgia, the tax is 5% of bet's amount.

4. Scheme of the algorithm

In most probabilistic models of a tennis match, the so-called root probabilities are introduced. This is the probability of winning a point by one of the rivals (for certainty the first) on his serve and on the serve of the second player, in short – on the serve and on the return. These two probabilities are especially different in men's tennis. The simplest model in the hierarchy of models suggests that the root probabilities remain unchanged during the match. Multiple usage of the rules of addition and multiplication of probabilities, as well as the formula of full probability, gives expressions (albeit very cumbersome) for calculating the probabilities of all possible events in a tennis match. We list some of these events. Winning the match, set, game, including the different handicaps, the exact score of sets, games, less/more than the different totals and so on. For a good model, the probabilities of events derived from the root ones must coincide with the probabilities of the same events, obtained from the corresponding given odds by the transition algorithm [2]. In fact, there may not be coincidences, but we get the equations with respect to root probabilities. In view of the fact that the number of equations exceeds the number of unknowns, it is impossible to satisfy all equations. Therefore, a certain function of the arising discrepancies is compiled, the subsequent minimization of which gives the required root probabilities. Having the roots, one can already get the probability of all the necessary events, including in the game process from any current score. Next, it is necessary to use the transition algorithm probabilities-odds [2]. This is the general scheme of the algorithm for obtaining any derivative odds that accompany the tennis match.

5. The adequacy of the model

ISSN 1512-1232

The question, of course, is – how much the described model is adequate to the reality. Numerical experiments have shown that the deviations of the algorithmic odds obtained from the given ones are significant, in particular, the odds for the exact scores of the sets 2:0 and 0:2 and for less than total are overstated. This means that the assumption that the root probabilities are unchanged is rough and should be discarded.

We make the following assumption about the correction of the roots in the match process. The values of the current roots of the player, who lost the current point, are trimmed by a small percentage and, correspondingly to this cut-off, the values of the current roots of the player, who won the current point, are increased. For example, if the opponents' active root probabilities before the point rally in the ordinary game were 0.6 and 0.4, then with a half-percentage cut-off after the rally, with a probability 0.6 we will have a ratio of the roots 0.602 - 0.398 and with a probability 0.4 the ratio 0.597 - 0.403. Note that the mathematical expectation of the server's root after the point rally is equal to its value before the rally. Indeed, $0.6 \times 0.602 + 0.4 \times 0.597 = 0.6$. The returner's root probability has the same property: $0.6 \times 0.398 + 0.4 \times 0.403 = 0.4$. Thus, the described method of cut-off has an important property of conservatism. The introduction of cut-off has greatly increased the level of the model's accordance to reality. In particular, the following odds were obtained for the exact scores in the men's match with equal basic odds on victory-defeat: 2:0 - 3.15, 2:1 - 4.20, 1:2 - 4.20, 0:2 - 3.15. These values are very realistic, what would be impossible without the correction of the root probabilities. Note that interesting results were obtained in [7], where the cut-off method was used to correct the probability of success in the Bernoulli trials.

Further, an attempt was made to improve the model by introduction of the cut-off percentage dependency on the importance of the current point, namely, after losing an important point, the percentage of cut-off was increased. However, this did not give a tangible effect. This can be explained by the fact that rallies of important points are not an isolated phenomenon - in order to achieve such a rally, it is necessary to pass a series of rallies of less important points, which eliminates the difference between important and less important points.

In general, the formulation of adequacy criteria is a separate problem, quite difficult for modeling of the changing chances of tennis players. This problem should be solved together with professional bookmakers, so one of the reasons for publishing this article is just to attract bookmakers' attention to our model.

6. IMPLEMENTATION, ILLUSTRATIONS

The model is implemented in the C++ programming language. The first group of odds mentioned at the beginning of the article is given at the input. The model gives the odds for all kinds of events by the time the match begins. Among them there are also quite exotic for non-specialists, and in the screenshots below they can be ignored. Next, the system requires entering the name of the player who serves the first, which becomes known after the draw. The first player in the pair has the name z, and the second – x and the input is done by pressing the corresponding keys. This first input does not have any noticeable effect on the odds, except for some. For example, if the most probable exact score of the set when z serves the first was 6:3, then with the x starting the set, the most probable score is 6:4. Further, the input of z or x already means the input of the winner's name of the current point. The consequence of this input is the recalculation and change in some degree of odds in all the markets we observe. Next, the name of the winner of the next point is entered, a new current score and new current odds are displayed on the screen. And so on until the end of the match. To determine the root probabilities before the match start takes less than two minutes, the subsequent calculations are made practically instantly. To keep within 2 minutes, while minimizing the discrepancy function, was achieved by using the change of variables. Instead of the original roots, we use their half-sum, which is responsible for the power of the first tennis player, and their half-difference, which is responsible for how "male" is the considered match. The effect of these new variables on the discrepancy function is largely independent, which, while minimizing, gives a great saving in computer time.

ISSN 1512-1232

Below are the concrete content of the initial data file and screenshots for the two possible current states of the considered tennis match. More information about usage of the program code is available at http://213.157.215.237/liveodds/.

1.57 2.25 1.85 21.5 1.85 / 1 2 < Total > /

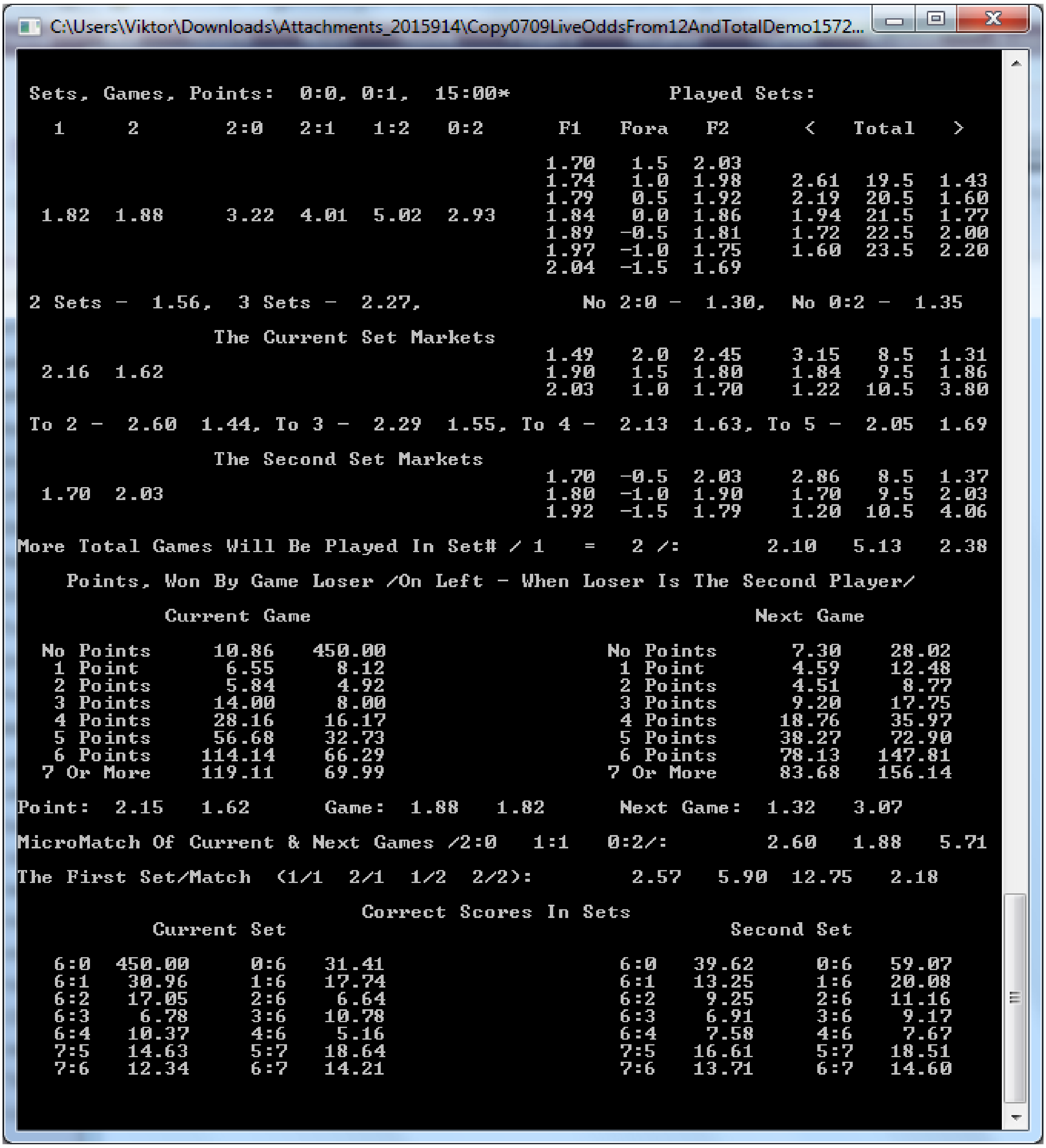

Fig. 1. The odds after losing the favorite of the first game on its own serve

The symbol * indicates the server. “To 3” means first to win 3 games in the current set. As it is seen from the first screenshot, the odd for the favorite’s winning is increased from 1.57 to 1.82, and the other odds are changed, too. Numbers from the second screenshot also allow one to feel the dynamics of the match.

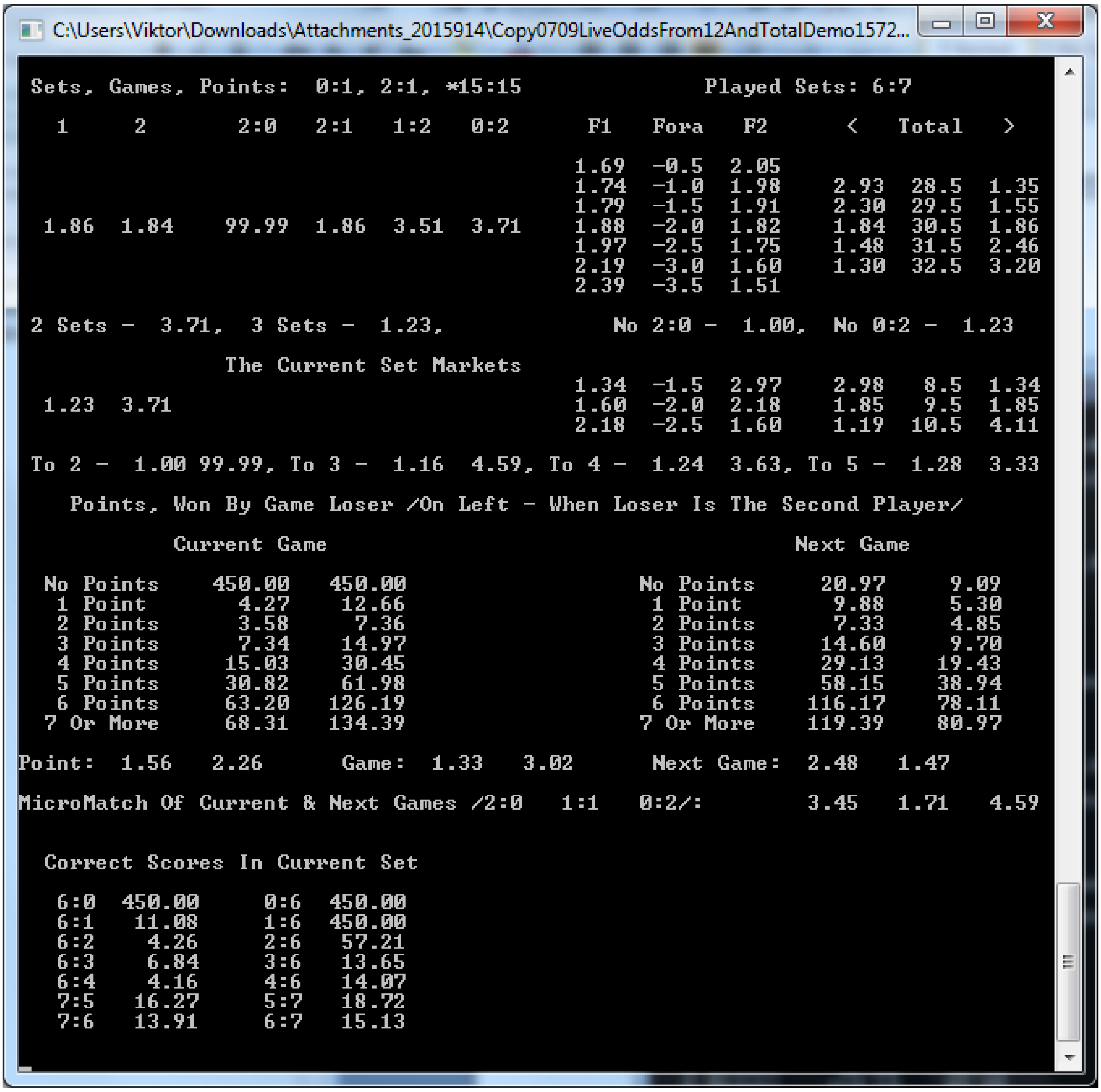

Fig. 2. The odds after losing the favorite of the first set and the break in the second

7. Conclusion

The results of the research demonstrate the effectiveness of both components of the presented model. As a purely probabilistic-tennis, and author's model of the transition odds - probabilities and back.

References

1. Хуцишвили В.В. Моделирование букмекерских коэффициентов в процессе теннисного матча. Международная научная конференция «Информационные и компьютерные технологии, моделирование, управление», посвящённая 85-летию со дня рождения академика И. В. Прангишвили. Тбилиси: «Технический Университет», 2015, стр. 477-480.

ISSN 1512-1232

2. Хуцишвили В.В. Фундаментальное свойство оптимальных букмекерских коэффициентов. Сборник трудов Института Систем Управления Арчила Элиашвили Грузинского Технического Университета, 2014, № 18, стр. 136-141.
3. Хуцишвили В.В. Связь между вероятностями событий и коэффициентами букмекеров. Сборник трудов Института Систем Управления Арчила Элиашвили Грузинского Технического Университета, 2014, № 18, стр. 142-146.
4. Vovk V., Zhdanov F. Prediction with expert advice for the Brier game. Journal of Machine Learning Research, 2009, vol. 10, pp. 2445-2771.
5. Griffith R.M. Odds adjustments by American horse-race bettors. American Journal of Psychology, 1949, vol. 62, pp. 290-294.
6. Clarke S., Kovalchik S., Ingram M. Adjusting Bookmaker's Odds to Allow for Overround. American Journal of Sports Science, 2017, vol. 5, No. 6, pp. 45-49. (doi: 10.11648/j.ajss.20170506.12).
7. Хуцишвили В.В. //О необходимости, способе и следствиях учёта апостериорной информации в схеме испытаний Бернулли// GESJ: Computer Sciences and Telecommunications. 2016, No. 1(47), pages 42-47. (http://gesj.internet-academy.org.ge/download.php?id=2651.pdf).

Article received: 2018-06-05