ORIGINAL ARTICLE

Journal Section

Urnings: A new method for tracking dynamically changing parameters in paired comparison systems

Maria Bolsinova[1] | Gunter Maris[2] | Abe D. Hofman[3,4] | Han L. J. van der Maas[3] | Matthieu J. S. Brinkhuis[5]

[1]Tilburg University, Methodology and Statistics, Tilburg, 5000 LE, The Netherlands

[2]Metior Consulting, The Netherlands

[3]University of Amsterdam, Psychological Methods, Amsterdam, 1018 WS, The Netherlands

[4]Prowise, Amsterdam, 1011 VL, The Netherlands

[5]Utrecht University, Information and Computing Sciences, Utrecht, 3584 CC, The Netherlands

Correspondence
Maria Bolsinova PhD, Department of Methodology and Statistics, Faculty of Social Sciences, Tilburg University, PO Box 90153, 5000 LE, Tilburg, The Netherlands
Email: m.a.bolsinova@uvt.nl

Funding information
The work of Maria Bolsinova was partially funded by the NAEd/Spencer Foundation Fellowship

We introduce a new rating system for tracking the development of parameters based on a stream of observations that can be viewed as paired comparisons. Rating systems are applied in competitive games, adaptive learning systems, and platforms for product and service reviews. We model each observation as an outcome of a game of chance that depends on the parameters of interest (e.g., the outcome of a chess game depends on the abilities of the two players). Determining the probabilities of the different game outcomes is conceptualized as an urn problem, where a rating is represented by a probability (i.e., proportion of balls in the urn). This setup allows for evaluating the standard errors of the ratings and performing statistical inferences about the development of, and relations between, parameters. Theoretical properties of the system in terms of the invariant distributions of the ratings and their convergence are derived. The properties of the rating system are illustrated with simulated examples and its potential for answering research questions is illustrated using data from competitive chess, a movie review system, and an adaptive learning system for math.

KEYWORDS
Paired comparisons, rating system, statistical inference, tracking

1 | INTRODUCTION

Tracking the evolution of a parameter that changes with time, based on a stream of observations that are independent but non identically distributed, is a problem encountered in many situations. Let us sketch three different contexts in which this is the case. First, in competitive sports the outcome of every match between players is dependent on the difference between the strength of the players (Elo, 1978; Glickman, 2001). It is of interest to track the change in the strength of the players over time. Second, online platforms ask users to provide reviews of products or services, for example using a number of stars. Based on these reviews, the quality (likeability) of the products can be tracked over time. However, how a product is evaluated depends not only on its quality, but also on the user (i.e., some users are more likely to give positive reviews than others), which might also change with time. Third, in online learning systems, learners respond to a series of exercises (Brinkhuis et al., 2018; Klinkenberg et al., 2011). Here the parameters of interest, that determine the probability of success of a learner on an exercise, are the ability of the learner and the difficulty of the exercise. Especially ability is expected to change over time.

Although these applications are rather different, they all can be modeled using a model with the same simple structure, which is referred to as the Bradley-Terry-Luce model for paired comparisons (Bradley and Terry, 1952; Luce, 1959) or the Rasch model for educational testing and preference data (Rasch, 1960). The probability of observing one of the two possible outcomes is modeled with a logistic function of the difference between the properties of the involved objects, denoted by θ_i and θ_j:

$$\Pr(X_{ij} = 1) = \frac{\exp(\theta_i - \theta_j)}{1 + \exp(\theta_i - \theta_j)}. \tag{1}$$

In the context of sports, $X_{ij} = 1(0)$ when player i (j) won the game, and both parameters refer to the momentary playing strength of the players. In the context of reviews of products, $X_{ij} = 1(0)$ when product i receives a positive (negative) review from user j[1], the parameter θ_i reflects the tendency of product i to receive positive reviews (i.e., quality or likeability of a product) and the parameter θ_j reflects the tendency of user j to give negative reviews (i.e., strictness of the user). In online learning systems, $X_{ij} = 1(0)$ when the learner i solves the exercise j correctly (incorrectly), the parameter θ_i refers to the skill of the learner, and the parameter θ_j to the difficulty of the exercise. Since the parameters change over time, their values are time-specific. However, for notational convenience we will not use time-specific subscripts.

Three properties of these applications are of importance when developing methods to track evolving parameters. First, in each of the contexts we sketched, the goal of tracking the parameters of interest is not only tracking their development over time, but also using their current values to select which observation to record next, such as to create a optimal buying, playing or learning experience. In the context of product reviews, one wants to recommended a product to the user's liking. In competitive games or sports one wants to pair players of about equal strength such that the game is challenging for the both of them. In the context of learning, the learning experience is more engaging and efficient if exercises match the learner's ability (Jansen et al., 2013). We will refer to this personalisation as adaptive matchmaking. Second, the applications that we are considering are large-scale systems in which the number of observations and the number of parameters that have to be tracked are very high. Below, as illustrations, we will use a data set of 2 million chess games involving more than 60 thousand players, data of an online movie review system with more than 20 million movie evaluations of more than 25 thousand movies from more than 100 thousand users, and data of an online learning system which generates more than a million responses per day in multiple domains.

[1] Instead of dichotomous scores (positive vs. negative), reviews are often provided on an ordinal scale, for example as star ratings from 1 to 5. For simplicity, here we consider dichotomous scores, but a similar approach can be used for ordinal scores as shown in the appendix, Section B.

Furthermore, the parameters need to be updated on-the-fly as the observations are coming in. Third, there is typically no theory on how the parameters of interest change over time. Different parameters might also develop differently. For example, there are large individual differences in how people learn. Moreover, the dynamics of the parameters may include complex feedback loops based on the information in the system. For example, interventions may be taken by a teacher based on the information from the learning system about a particular learner which would change the development of her ability. Analogously, a product might be modified if it obtains a lot of negative reviews, and a player might change tactics or take an new coach when her progress in playing strength levels off.

One approach to tracking parameter change is that of rating systems. Among rating systems the most prominent is developed by Arpad Elo, who developed a method for tracking idiosyncratic changes in chess playing proficiency over long time periods in which individual players played at irregular intervals and frequency (Elo, 1978; Batchelder and Bershad, 1979). The Elo rating system is a simple, transparent, computationally efficient, and self-correcting algorithm. In the Elo system for chess, each player's rating is updated after every game based on the difference between the actual outcome of the game (i.e., whether player A won against player B, ignoring draws for now) and the expected outcome of the game (i.e., the probability that player A wins against player B), where the probability that one player wins against another player is based on the Bradley-Terry-Luce model for paired comparisons. In this way, winning against a more proficient player leads to a larger increase of one's rating than winning against a less proficient player. The system is still used by many chess federations, such as the World Chess Organisation (FIDE), and has been adopted (with some modifications) in a wide range of different areas: from rating of different sports and games (Hvattum and Arntzen, 2010; Mangan and Collins, 2016) to educational measurement (Klinkenberg et al., 2011; Pelánek, 2016), behavioral ecology (Neumann et al., 2011; Albers and Vries, 2001), and infometrics (Lehmann and Wohlrabe, 2017).

The Elo rating system generates a Markov chain for every parameter of interest, meaning that conditional on the current rating the new rating is independent of all the previous ratings. However, a problematic feature is that it cannot be proven that this Markov chain has an invariant distribution, and the properties of this distribution are unknown (Brinkhuis and Maris, 2009, 2019). This makes it difficult to study the statistical properties of the ratings. More concretely, the standard errors and reliability of the ratings in the system are unknown, such that one cannot perform proper statistical inference using the ratings. For example, it is not possible to test whether the user-evaluated quality of the product has changed after changing the manufacturing process or test whether the learner's skills have grown. Furthermore, if decisions need to be taken based on the ratings, it is important to take their uncertainty into account. For example, should learning and assessment within an online system be combined, at least the psychometric properties of the ratings (i.e., reliability) need to be established.

Alternatives to the Elo system have been proposed that allow for taking uncertainty about the ratings into account. Two popular examples are Glicko (Glickman, 2001) and TrueSkill (Herbrich et al., 2006; Minka et al., 2018). Glicko provides an updating algorithm for a Gaussian state-space model, and TrueSkill is based on Gaussian density filtering. Other, yet related, approaches are proposed under the name of Dynamic Difficulty Adjustment (DDA) using hidden Markov models (see Zohaib (2018) for a review). All these methods can be considered within a broader class of methods for filtering problems, that is for estimating states of a dynamical system based on noisy observations. Filters typically consist of two submodels: a measurement model that describes the relationship between the noisy measurements and the parameters of interest, and a dynamic model that describes how the parameters change over time. A notable example of such methods is the Kalman filter (Kalman, 1960; Welch and Bishop, 1995) that allows for the recursive computation of the posterior densities of the parameters. In this method, the statistical properties of the parameter estimates are known, which allows statistical tests for change and group differences, for example. However, specific assumptions have to hold, such as normality and linearity. More flexible methods, such as particle

filters (Arulampalam et al., 2002), relax these assumptions, but require more intensive computations, which might be not suitable for large-scale applications where real-time updating of a large number of parameters is required. Furthermore, particle filters similar to other filtering methods require the specification of the dynamic model for the change of the true values of the parameters of interest over time. While different flexible growth models could be specified, for the types of applications that we are considering there is not a lot of theory available for formulating such dynamic models. In some cases one of the main purposes of collecting the data and tracking the parameters is to study the patterns of development and to formulate and evaluate such theories (such as in the case of models of learning). Therefore, having to specify a dynamic model beforehand is undesirable in these contexts.

We have considered two approaches to tracking the evolution of parameters over time: Elo rating system and filters. The first one is computationally efficient and flexible in terms of following the development of the parameters without having to specify the dynamic model for this development, but does not easily provide measures of uncertainty about the parameters. The second approach provides parameter estimates with known statistical properties, but requires one to specify the model for growth. In this paper we propose an alternative rating system which, on the one hand, is similar to the Elo rating system in that it is a simple, scalable and self-correcting algorithm for estimating dynamically changing parameters, without having to specify a dynamic model, and, on the other hand, provides ratings with known standard errors when no change occurs for some time.

The rest of the paper is organised as follows. In Section 2 we describe our proposed rating system. In Section 3 we present the results of simulations demonstrating the statistical properties of the rating system and the results of three different applications of the system.

2 | METHODS

If we re-parameterize the Bradley-Terry-Luce/Rasch model by taking the expit transformation of the parameters, it can be expressed as follows:

$$\Pr(X_{ij} = 1 | \pi_i, \pi_j) = \frac{\pi_i(1-\pi_j)}{\pi_i(1-\pi_j) + (1-\pi_i)\pi_j}, \tag{2}$$

where $\pi_i = \frac{\exp(\theta_i)}{1+\exp(\theta_i)}$ and $\pi_j = \frac{\exp(\theta_j)}{1+\exp(\theta_j)}$. For the purposes of developing our rating system, we will work with the transformed parameters π_i and π_j throughout the paper. That is, the parameters of interest are probabilities between 0 and 1. Note, that the parameters π_i and π_j are not uniquely identified since the outcome depends only on the difference of their logits. Therefore, some identifying constraint is needed, and different constraints that lead to different estimates of the parameters are possible.

Under the model each observation X_{ij} can be conceptualised as an outcome of the following process:

repeat
  $Y_i \sim Bernoulli(\pi_i)$
  $Y_j \sim Bernoulli(\pi_j)$
until $Y_i \neq Y_j$
return $X_{ij} = Y_i = 1 - Y_j$

This process can be thought of as a simple game of chance where each of the two players draws a ball from an (infinite) urn (Johnson and Kotz, 1977) containing a proportion π_i (π_j) of green balls (the others being red) until they have drawn balls of different color. This conceptualisation of the observations allows us to set up a rating system in which the actual game of chance is mimicked by a game of chance with urns of finite size and the compositions of these urns

are used to track the parameters of interest.

To track the parameter of interest, π_i, we will use a grid of discrete proportions $\frac{R_i}{n}$, $R_i \in [0:n]$, where the value of n determines the granularity of the grid. R_i can be thought of the number of green balls in the tracking urn of size n (with others being red). We call R_i "urnings", and refer to the proportions $\frac{R_i}{n}$ as scaled urnings, as they are transformed to be on the same scale as parameters π_i. For each i, R_i will be updated after each observation $X_{i\cdot}$ in such a way that the invariant distribution of R_i is binomial with parameters π_i and n (conditional on the total sum of R_i in the system), when there is no change. In the rest of this section we will first derive a rating update with desired properties and then determine how the urnings converge to their invariant distribution.

2.1 | Derivation of the Urnings algorithm

To construct an update that guarantees that distributions of R_i are binomial conditional on their sum, we mimic the data generating process governed by the true values of the parameters by a simulated process governed by the urning values:

repeat
  $Y_i^* \sim Bernoulli\left(\frac{R_i}{n}\right)$
  $Y_j^* \sim Bernoulli\left(\frac{R_j}{n}\right)$
until $Y_i^* \neq Y_j^*$
return $X_{ij}^* = Y_i^* = 1 - Y_j^*$

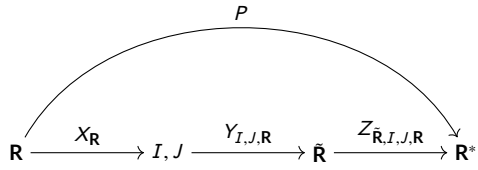
The main ingredient of the update is based on the difference between the observed and the simulated outcome (i.e., the balls drawn in the mock-up game are replaced by the balls drawn in the real game):

$$\tilde{R}_i = R_i + X_{ij} - X_{ij}^*; \tag{3}$$

$$\tilde{R}_j = R_j + (1 - X_{ij}) - (1 - X_{ij}^*). \tag{4}$$

Such an update is similar to the update of Elo (1978) where the update includes a difference between the observed and the *expected* outcome. However, in contrast to the Elo rating system (Brinkhuis and Maris, 2009), it is straightforward to derive the invariant distribution of the urnings as we show in subsequent sections. Note that the update conserves the sum of R_i and R_j and that this property is not unique to our algorithm. In the Elo rating system it is known as the economic notion of rating points (Batchelder et al., 1992).

Let us by $\mathbf{R} = \{R_1, \ldots, R_N\}$ denote the urnings of N players. The update generates a Markov chain $P : \mathbf{R} \xrightarrow{P} \mathbf{R}^*$ which is defined on the space $S = \{\mathbf{r} = \{r_1, \ldots, r_N\} \in (0, \ldots, n)^N, \sum_k r_k = r_+\}$, where $r_+ \in [1 : (nN - 1)]$ is the chosen value for the total number of green balls in the tracking urns of all players. At each stage the update $\mathbf{R} \xrightarrow{P} \mathbf{R}^*$ can be described by the following scheme:



where $\mathbf{R}, \tilde{\mathbf{R}}, \mathbf{R}^* \in S$, and $(I, J) \in \{1, \ldots, N\}^2, J > I$. $X_{\mathbf{R}}$ chooses players for the game, $Y_{I,J,\mathbf{R}}$ proposes a new state $\tilde{\mathbf{R}}$

as in Equations 3 and 4 with $\tilde{R}_k = R_k, \forall k \neq I, J$, and $Z_{\tilde{\mathbf{R}},I,J,\mathbf{R}}$ either accepts or rejects the new state[2] (i.e., $\mathbf{R}^* = \tilde{\mathbf{R}}$ or $\mathbf{R}$). The underscore terms mean that the law of that variable is independent of the whole past conditioned to those terms. The actual outcome of the game between the players (X_{ij}) is observed in correspondence with $Y_{I,J,\mathbf{R}}$ and is independent of the process seen so far (i.e., it depends only on π_i and π_j).

We start from the case of only two players and derive the invariant distribution of the urnings when all proposals $\tilde{\mathbf{R}}$ are accepted. We show that in this case there is a dependence between the urnings of the two players which can be removed by accepting the proposals with a specific probability. Then we move to the case of multiple players. Here, the probability of a unique pair of players to be selected $p_X(i,j \mid \mathbf{r})$ is specified. We first consider random selection of players for each game and then move to the case of adaptive matchmaking in which an additional term is added to the acceptance probability to remove the effect of adaptivity.

2.1.1 | Two players playing each other

When $N = 2$, $p_X(1,2 \mid \mathbf{r}) = 1$. If the urnings are updated directly based on the difference between the observed and simulated outcomes, then $\Pr(\mathbf{R}^* = \tilde{\mathbf{R}} \mid \mathbf{R}, I = 1, J = 2, \tilde{\mathbf{R}}) = 1$. The properties of the generated Markov chain are given in Theorem 1.

Theorem 1 *When two players repeatedly play against each other and their urnings are directly updated based on the difference between the observed and simulated outcome, then P is an ergodic Markov chain with the following limiting distribution:*

$$\Pr(R_1 = r_1, R_2 = r_2) = \frac{[r_1(n-r_2) + (n-r_1)r_2]\binom{n}{r_1}\pi_1^{r_1}(1-\pi_1)^{n-r_1}\binom{n}{r_2}\pi_2^{r_2}(1-\pi_2)^{n-r_2}}{Z} \mathcal{I}_{r_1+r_2=r_+}, \tag{5}$$

where $\mathcal{I}$ is the indicator function and Z is the normalising constant that sums over all possible values for the urnings that sum to r_+:

$$Z = \sum_{s=\min(0,r_+-n)}^{\min(r_+,n)} [s(n-r_+ + s) + (n-s)(r_+ - s)]\binom{n}{s}\pi_i^s(1-\pi_i)^{n-s}\binom{n}{r_+ - s}\pi_j^{r_+-s}(1-\pi_j)^{n-r_++s} \tag{6}$$

The proof is presented in the appendix, in section A.1.

We found that the urnings of two players have a known invariant distribution when they are updated based on the difference between the observed and the simulated outcomes (Theorem 1). Inspecting this invariant distribution, we see that is has an undesirable property, as R_1 and R_2 are dependent on each other not only through their sum being constant but also through the presence of the factor $[r_1(n-r_2) + (n-r_1)r_2]$ in the invariant distribution in Equation 5. We modify the urning update such that, the only dependence between the urnings of the two players comes from their total sum being constant. Since $[r_1(n-r_2) + (n-r_1)r_2]$ depends on the quantities that are known and not on the unknown true values of the parameters, we can use a Metropolis-Hastings step to achieve our goal. Instead of directly accepting the proposal in $Z_{\tilde{\mathbf{R}},I,J,\mathbf{R}}$, we accept the proposed value $(\check{r}_1, \check{r}_2)$ with probability

$$\min\left(1, \frac{r_i(n-r_j) + (n-r_i)r_j}{\check{r}_i(n-\check{r}_j) + (n-\check{r}_i)\check{r}_j}\right), \tag{7}$$

[2] Sections 2.1.1 and 2.1.2 elaborate on this acceptance/rejection step.

with $i = 1$ and $j = 2$. The properties of the Markov chain with this modified update are given in Theorem 2.

Theorem 2 *When two players are playing against each other, and the proposed value based on the difference between the observed and simulated outcome is accepted with probability given in Equation 7, then P is a reversible ergodic Markov chain with the limiting stationary distribution that satisfies detailed balance and is given by*

$$p_{\pi_1,\pi_2}(r_1, r_2) = \frac{\binom{n}{r_1}\pi_1^{r_1}(1-\pi_1)^{n-r_1}\binom{n}{r_2}\pi_2^{r_2}(1-\pi_2)^{n-r_2}}{Z^*}\mathcal{I}_{r_1+r_2=r_+}, \tag{8}$$

where the denominator Z^* is the normalizing constant that sums the numerator over all values of (r_1, r_2) adding up to r_+, that satisfies detailed balance.

The proof is presented in the appendix, in section A.2.

2.1.2 | Tournament with N players

Let us consider a situation in which N players are competing against each other and the players for each game are selected randomly from the pool. Here, $p_X(i, j \mid \mathbf{r}) = \frac{2}{N(N-1)}$. That is, the probability of a unique pair of players (i, j) to be selected is independent of the current urnings and is equal to the inverse of the total number of unique pairs. $Y_{I,J,\mathbf{R}}$ proposes a new state with the values for the selected players specified according to Equations 3 and 4, and $\check{r}_k = r_k, \forall k \notin \{i, j\}$. $Z_{\check{\mathbf{R}},I,J,\mathbf{R}}$ accepts the proposed value with probability given by Equation 7. The properties of the Markov chain for such a tournament of N players are summarised in Theorem 3 (see Section A.3 in the appendix for the proof).

Theorem 3 *In a tournament with N players, when players for each game are selected randomly, the proposed value for the selected players is based on the difference between the observed and simulated outcome of the game between them, and the proposal is accepted with probability given in Equation 7, then P is a reversible ergodic Markov chain with the limiting stationary distribution that satisfies detailed balance and is given by*

$$p_{\boldsymbol{\pi}}(\mathbf{r}) = \frac{\prod_k \binom{n}{r_k}\pi_k^{r_k}(1-\pi_k)^{n-r_k}}{F(\boldsymbol{\pi}, r_{+N})}\mathcal{I}_{\sum_k r_k = r_+}, \tag{9}$$

where $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_N\}$, and $F(\boldsymbol{\pi}, r_+)$ is the normalizing constant.

In the actual applications of rating systems, the current ratings are often used to match players with each other, select products for users and exercises for learners. This is not an innocent activity, as it can potentially change the invariant distribution of the ratings. This is a general phenomenon which seems to have received little attention in other works on rating systems (see Hofman et al., 2020, p. 12). In the context of the urnings, we have that if the choice of players is dependent on the current urnings, then the distribution of the updated urnings is no longer equal to the desired invariant distribution:

$$p(\mathbf{r}^*) = \sum_{\mathbf{r}}\sum_{\check{\mathbf{r}}}\sum_{i}\sum_{j>i} p_Z(\mathbf{r}^* \mid \check{\mathbf{r}}, i, j, \mathbf{r})p_Y(\check{\mathbf{r}} \mid i, j, \mathbf{r})p_X(i, j \mid \mathbf{r})p_{\boldsymbol{\pi}}(\mathbf{r}) \neq p_{\boldsymbol{\pi}}(\mathbf{r}^*), \tag{10}$$

where the probability of selecting players i and j depends on the current urnings. The ratings keep the memory of which games have been played, which distorts their invariant distributions. Correcting for the dependence on the

current state can be achieved by adding a Metropolis-Hastings step that maintains detailed balance. The ratio of selection probabilities after and before the update are added to the to the acceptance probability:

$$p_Z(\mathbf{r}^* \mid \tilde{\mathbf{r}}, i, j, \mathbf{r}) = \begin{cases} \min\left(1, \frac{r_i(n-r_j)+(n-r_i)r_j}{\tilde{r}_i(n-\tilde{r}_j)+\tilde{r}_j(n-\tilde{r}_i)} \frac{p_X(i,j \mid \tilde{\mathbf{r}})}{p_X(i,j \mid \mathbf{r})}\right) & \text{if } \mathbf{r}^* = \tilde{\mathbf{r}} \\ \max\left(0, 1 - \frac{r_i(n-r_j)+(n-r_i)r_j}{\tilde{r}_i(n-\tilde{r}_j)+\tilde{r}_j(n-\tilde{r}_i)} \frac{p_X(i,j \mid \tilde{\mathbf{r}})}{p_X(i,j \mid \mathbf{r})}\right) & \text{if } \mathbf{r}^* = \mathbf{r} \\ 0 \text{ otherwise.} \end{cases} \tag{11}$$

The correction means that if the proposed update makes a future match between players i and j less likely, then it is less likely to be accepted. That is, the extra Metropolis-Hastings step undoes the memory of kernel selection. Put differently, it ensures fair selection, by sometimes ignoring the observation.

It is important to note that if ratings are used for matchmaking (which is in many applications their primary use), we can correct for that if, and only if, we have an explicit selection mechanism $X_{\mathbf{R}}$. It does not matter what the matchmaking mechanism is (except for some trivial conditions), it does not need to be constant, but it should be explicitly known and corrected for when urnings are updated.

We note one more important condition for matchmaking: $p_X(i,j \mid \mathbf{r})$ should be chosen in such as way that the Markov chain is irreducible. That is, there should be a path with positive probability that joins any starting $\mathbf{r} \in S$ with any ending $\dot{\mathbf{r}} \in S$. Since the other two components of the update ($Y_{I,J,\mathbf{R}}$ and $Z_{\tilde{\mathbf{R}},I,J,\mathbf{R}}$) do not threat irreducibly of the chain, choosing $p_X(i,j \mid \mathbf{r})$ in such a way that all players are connected to each other (i.e., if there is a non-zero probability that player i is matched to player j, and that player j is matched to player k, then i and k are connected) guarantees irreducibility. What should not happen is that, for instance, in a chess competition males only play against males and females only play against females, since in that case the Markov chain will be reducible.

The full urnings algorithm is as follows:

Select players i and j according to $p_X(i,j \mid \mathbf{r})$

Modeled reality:

repeat
  $Y_i \sim Bernoulli(\pi_i)$
  $Y_j \sim Bernoulli(\pi_j)$
until $Y_i \neq Y_i$
return (Y_i, Y_j)

Rating system:

repeat
  $Y_i^* \sim Bernoulli(r_i/n)$
  $Y_j^* \sim Bernoulli(r_j/n)$
until $Y_i^* \neq Y_j^*$
return (Y_i^*, Y_j^*)

Proposed update:

$$\tilde{r}_i = r_i + Y_i - Y_i^*$$
$$\tilde{r}_j = r_j + Y_j - Y_j^*$$

Metropolis-Hastings: accept $\tilde{\mathbf{r}}$ with probability:

$$\min\left(1, \frac{r_i(n-r_j)+(n-r_i)r_j}{\tilde{r}_i(n-\tilde{r}_j)+(n-\tilde{r}_i)\tilde{r}_j} \times \frac{p_X(i,j \mid \tilde{\mathbf{r}})}{p_X(i,j \mid \mathbf{r})}\right).$$

Theorem 4 *When the urnings are updated according to the full urnings algorithm, assuming that $p_X(i, j \mid \mathbf{r})$ is chosen such that P is irreducible, then P is a reversible ergodic Markov chain with the limiting stationary distribution given in Equation 9 that satisfies the detailed balance condition.*

The proof of Theorem 4 is given in the appendix, Section A.4. The invariant distribution of the urnings of all players is the product of the binomial distributions conditional on the total sum of the urnings (i.e., the total number of green balls in the system is constant). Compared to the product of independent binomials, there is a small negative dependence between the urnings and their standard errors are slightly smaller (i.e., they are conservative in that the standard errors of the binomial give an upper bound to the actual standard errors). However, when the number of players (products and users, learners and exercises) is large this difference in standard errors is very small and can be ignored for all practical purposes. Furthermore, for each player the expected value of R_i/n is extremely close to π_i. To make the dependence between the urnings of different players explicit, we characterize the invariant distribution in terms of elementary symmetric polynomials in Theorem 5.

Theorem 5 *If the R_i are independent binomial random variables with parameters n and π_i, the distribution of $\mathbf{R}$ conditional on $\sum_i R_i = r_+$ is*

$$p(\mathbf{r}|r_+) = \frac{\prod_{i=1}^{N} \binom{n}{r_i} \rho_i^{r_i}}{\gamma_{r_+}(\boldsymbol{\rho})} \tag{12}$$

where $\rho_i = \frac{\pi_i}{1-\pi_i}$, $\gamma_{r_+}(\cdot)$ is the elementary symmetric polynomial of order r_+, and $\boldsymbol{\rho} = \{\underbrace{\rho_1, \ldots, \rho_1}_{n \text{ times}}, \ldots, \underbrace{\rho_N, \ldots, \rho_N}_{n \text{ times}}\}$.

The proof can be found in the appendix, Section A.5.

From the following property of elementary symmetric functions: $\gamma_{r_+}(c\boldsymbol{\rho}) = c^{r_+}\gamma_{r_+}(\boldsymbol{\rho})$, we directly deduce that we can multiply all the ρ_i by some number c without changing the distribution in Equation 12. In other words, from the urnings $\mathbf{r}$ the odds $\boldsymbol{\rho}$ are not identifiable. This does not pose a problem as the response probability in Equation 2 only depends on the odds ratios, in which c cancels out:

$$\Pr(X_{ij} = 1 | \pi_i, \pi_j) = \frac{\frac{\pi_i}{1-\pi_i}\frac{1-\pi_j}{\pi_j}}{1 + \frac{\pi_i}{1-\pi_i}\frac{1-\pi_j}{\pi_j}}. \tag{13}$$

A direct consequence of Theorem 5 is that as long as we change the urn size (n) and/or the number of players (N) in such a way that the proportion $r_+/(nN)$ converges in probability, the urnings converge in law to independent non-identical binomial random variables. So for small N, the exact distribution is available and for the intended use case of the rating system (large N) the difference between the exact invariant distribution and independent binomial distributions can be ignored. Clearly, for different choices of this limiting proportion of green balls, we would get different distributions of the urnings. For every $r_+/(nN)$ the invariant distribution would be close to a product of independent binomials, but with a specific set of probability parameters. The corresponding odds, however, are proportional to one another. That is, the choice of the proportion $r_+/(nN)$ implies a choice for a constraint to identify the probability parameters.

2.2 | Convergence of the urnings

The urnings' update generates a Markov chain, as the new urnings depend only on the current urnings and the observed outcome. This Markov chain has a known invariant distribution, at every moment in time. That is, should the parameters of interest no longer change, the urnings would tend to a binomial distribution. The question arises of how, and how fast the urnings converge to their invariant distribution when the parameters do change.

As an illustration of how the conception of our urning update as a Markov chain helps in answering this question, we present a simple result, Theorem 6, that shows that every observation brings us closer in Kullback-Leibler distance to the invariant distribution (Brinkhuis and Maris, 2019). We formulate Theorem 6 in a general form, as it pertains to all Markov chains, and not only to our urning update.

Theorem 6 *If $p_t(X = x)$ is the (unknown) distribution of the random variable X at time t, and $p_\infty(X = s|Y = y)$ a transition kernel with $p_\infty(X = x)$ as its invariant distribution, then*

$$\sum_{x=0}^{n} \log\left(\frac{p_\infty(X=x)}{p_{t+1}(X=x)}\right) p_\infty(X=x) \leq \sum_{x=0}^{n} \log\left(\frac{p_\infty(X=x)}{p_t(X=x)}\right) p_\infty(X=x) \tag{14}$$

The proof can be found in the appendix, Section A.6. Theorem 6 formalizes the self-correcting property of any good rating system: If one's rating is off, having more observations will make the ratings change in the right direction.

3 | RESULTS

To illustrate the various uses of our rating system we consider simulated and real-life examples. We apply our method to three data sets: 1) historic data of competitive chess; 2) movie reviews from an online platform; 3) responses from an online learning system. Using historic data has the disadvantage that the matchmaking mechanism is not known, and hence we can not correct for it. This potentially brings some bias in the ratings. However, the advantage of using real-life data is that it shows all the complexity of real-life rating applications. For instance, the distribution of observation frequency typically has power law tails. Many players are observed only a few times, whereas a few account for the majority of games. Furthermore, parameters of interest change over time, and we deliberately selected data sets that extends over long periods of time.

3.1 | Simulated example

We simulated $N = 1000$ players playing 100 000 000 games. The true values of the player's abilities were generated such that their logits matched 1 000 equally spaced quantiles of the standard normal distribution and were kept constant throughout the simulation. The matchmaking probabilities depended on the current ratings in the system, proportional to $\exp(-2(\ln((R_i+1)/(n-R_i+1)) - \ln((R_j+1)/(n-R_j+1)))^2)$, such that players with similar ratings are more likely to be matched to each other. Each player started with $R_i = 50$ (i.e., $R_+ = 50 \times N$), and urn size n was set to 100. This starting value for R_+ was chosen to make it easier to illustrate the properties of the algorithm, since this choice guarantees that both the average logit of π used to generate data and the average logit of $\hat{R}/n$ are equal to 0 (i.e., the same identifying constraints are used for the parameters). Additionally, we present the results of simulations in which each player started with $R_i = 20$ (i.e., $R_+ = 20 \times N$) and with $R_i = 80$ (i.e., $R_+ = 80 \times N$) to show how the invariant distributions of the urnings depend on the choice of R_+. Data were simulated using R (R Core Team, 2018).

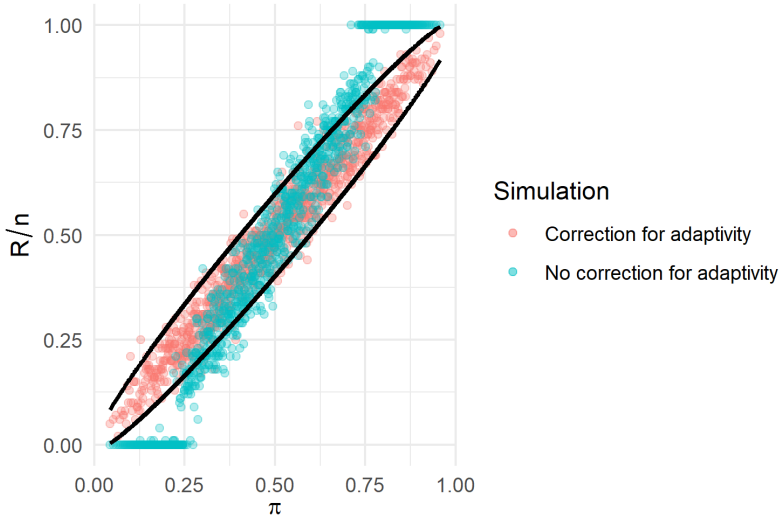


FIGURE 1 Simulated example results: True skills of the players vs. their scaled urnings for the simulations with (in red) and without (in blue) the correction for adaptive matchmaking.

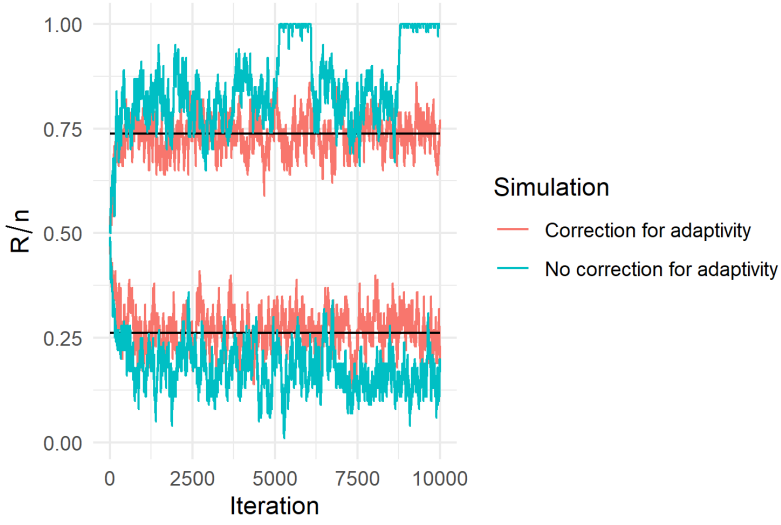


FIGURE 2 Simulated example results: Time series of the scaled urnings of two players with (in red) and without (in blue) the correction for adaptive matchmaking.

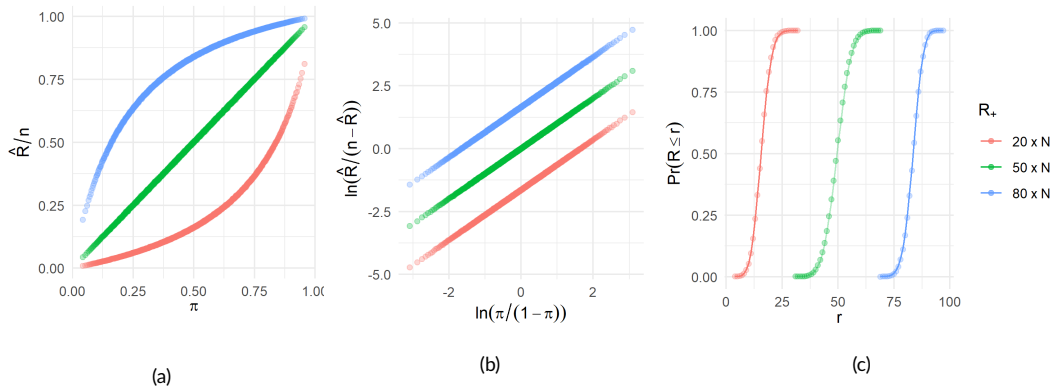


FIGURE 3 Dependence of the invariant distributions of the urnings on the choice of the total number of green balls in the system (R_+; indicated by different colors): (a) Relationship between the true values and average scaled urnings on the probability scale; (b) Relationship between the true values and the average scaled urnings on the logit scale; (c) Marginal invariant distributions of the urnings of one player (indicated with dots) compared to theoretical binomial cumulative distribution functions (indicated with lines).

Figure 1 shows in red the recovery of the true level of skill, with the black ellipse indicating the expected theoretical confidence intervals for the scaled urnings ($\pi_i \pm 1.96\sqrt{\pi_i(1-\pi_i)}/\sqrt{n}$). The correlation between the true values and the urnings (i.e., reliability) depends on the urn size and was in this simulation equal to .978. Figure 2 shows the time series of scaled urnings of two players (in red). One can see that they indeed fluctuate around the corresponding true values within the area of expected variability.

To illustrate the importance of correcting for the matchmaking probability, we repeated the simulation with the update without the ratio of matchmaking probabilities in the Metropolis-Hastings acceptance ratio. The blue dots in Figure 1 show the recovery of the true values for this scenario. One can see that unlike in the simulation with the proper update, here the scaled urnings do not lie within the expected theoretical bounds. In Figure 2 we can see in blue that the urning of the player with low ability was underestimated, while the urning of the player with high ability was overestimated. That is, the variance of the urnings was overestimated. The Metropolis-Hastings step of the algorithm corrects for this effect and prevents the inflation of the variance.

Figures 3a and 3b show how the average scaled urnings (computed without the first 10% of the sampled values) differ when different R_+ are used. In Figure 3a the values are shown on the probability scale, and in Figure 3b they are shown on the logit scale. For the simulation with $R_+ = 50$ the relationship between the values used to simulate the data and the average urnings is close to identity, while for the other two choices of R_+ the logits of the average scaled urnings are subject to a linear shift compared to the values used in a simulation. That is, the log-odds ratios for every pair of players computed using the average urnings (i.e., the quantity determining the probability of one player winning over the other) would be the same regardless the value of R_+. For one particular player Figure 3c shows the marginal invariant distributions of the urnings: All three distributions are close to binomial distributions with probability parameters depending on the choice of R_+ (.16 for $R_+ = 20$, .50 for $R_+ = .50$, and .84 for $R_+ = 80$).

3.2 | Historical chess data

A data set with professional chess games from 1990 till 2018 has been obtained from `kingbase-chess.net`. The data set contains games with at least 6 moves in which at least one of the players has an Elo rating of at least 2 000. The games for which the outcome of the game or the month and day of the game were missing (about 12%) were removed from the data set. The remaining 1 971 197 games of 63 734 players were used in the analysis. The games were sorted in chronological order and after each game the urnings of the players who participated in the game were updated.

One of the characteristics of chess that we focus on in this analysis is that there is an asymmetry between the players in the game: The player who plays white (i.e., who makes the first move) has an advantage over the player who plays black. Therefore, the probability that player i wins over player j depends on who plays white. One can also ask a question whether the advantage of playing white is the same for every player, in other words whether the individual differences in the chess skill are the same for playing white and playing black, or these are two different (but strongly related) skills. Having a rating system in which the invariant distribution of the ratings is known is useful, because it allows one to test a statistical hypothesis about whether the skills involved are the same or not. To be able to test such a hypothesis, we set up two urnings for each player - one for playing white and one for playing black.

Another characteristic of chess that is important for how the rating system is set up, is that there are three different outcomes possible (white wins, black wins, or a draw) instead of two outcomes as in the algorithm described in the Methods section. Therefore, we extended the algorithm to allow for three outcomes (see Section B for details). Since preliminary analysis has demonstrated that for players of equal skills the probability of a tie is larger than .5, a modified model for the probabilities of the game outcomes was used with the probability of a draw proportional to $2.5\pi_i(1-\pi_i)\pi_j(1-\pi_j)$.

Both for playing black and for playing white each player started with $R_i = 100$ and $n = 200$. Since we are using the algorithm on historic data, we do not know how exactly the players were matched to each other and cannot correct for the probability of matchmaking $p_X(i, j \mid \mathbf{r})$.

First, we evaluate the predictive power of our rating system. Before each game by two players that already have played 200 or more games, the difference between the logits of their scaled urnings was recorded. These differences were binned in bins of width .04 and the observed proportions of the different game outcomes within each bin were plotted against the expected model-based probabilities (see Figure 4). The sizes of the dots in the plot are proportional to the number of games within the bins. Figure 4 demonstrates that even though we could not correct for the matchmaking mechanism, the predictions on actual game outcomes based on urnings are quite accurate. There is some discrepancy for the probability of winning with black and draws for games in which the white player has the larger urning. Whether this discrepancy reflects genuine model misfit or is the result of not correcting for the matchmaking mechanism is unclear.

Second, we evaluate the hypothesis about the difference between playing black and playing white. Figure 5 shows the white and black scaled urnings of the players at the end of the analyzed period of time. The graph is based on the urnings of all players who played at least 100 games with black and 100 games with white. The black ellipse in the graph shows the expected relationship between the scaled urnings if they were based on the same skill. One can see that for a large number of players, their scaled urnings are outside of the ellipse, indicating that it is unlikely that the white and black urnings are based on the same skill. The observed correlation between the white and the black urnings was equal to .897, and the average difference between the white and black scaled urnings was equal to .047. These observed statistics were compared to the corresponding reference distributions to test whether the differences of these sizes are to be expected from sampling variability when the two sets of urnings are based on the same skill.

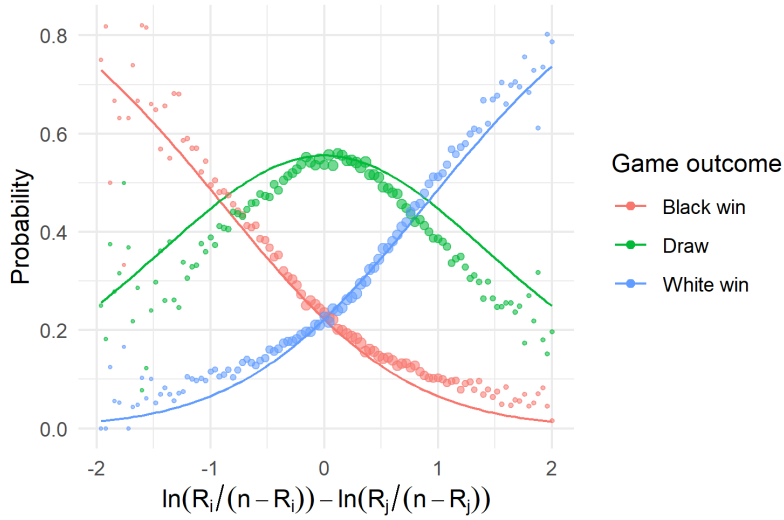


FIGURE 4 Probability of a loss, a draw and a win based on (binned) logit differences: Observed (dots with the size of the dot proportional to the number of observation in the bin) and expected (lines).

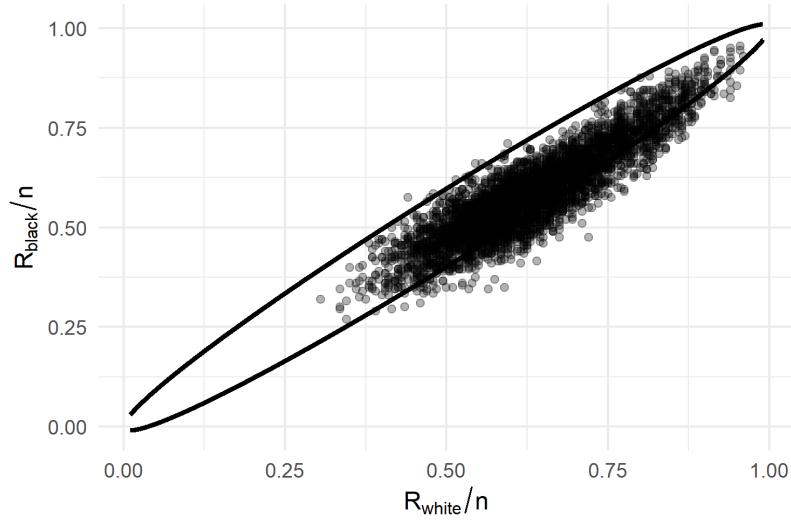


FIGURE 5 Scaled urnings for playing white versus playing black. The ellipse is the expected relation under the assumption that there is no difference.

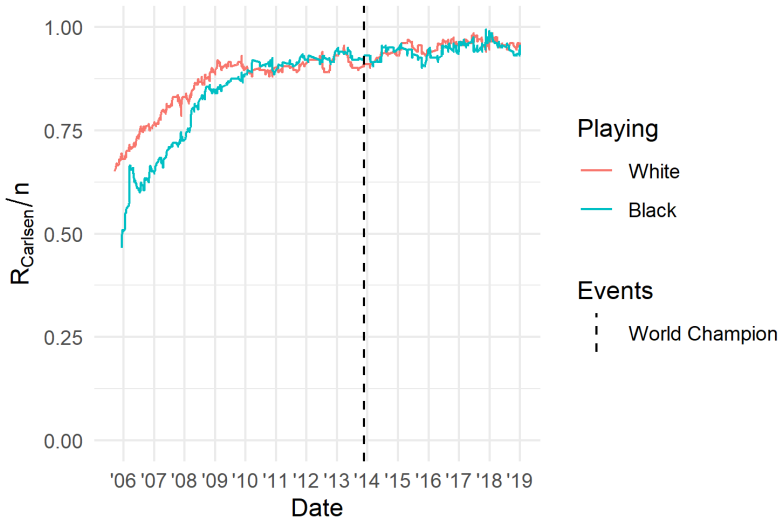


FIGURE 6 Scaled urnings (white and black) for Magnus Carlsen over time.

For that purpose the white and black urnings were summed and re-distributed. All of the replicated correlations were larger than the observed, and all the replicated differences were smaller than the observed. Therefore, we reject the hypothesis that playing white and playing black represent the same skill.

Given our conclusion about there being a difference between the skills for playing black and playing black, it is interesting to study the separate development of these skills for particular players. Here, we consider the historic record of the current world champion Magnus Carlsen. The urnings of Carlsen are displayed in Figure 6 starting from the moment when he had at least 200 games played with white and black. One can see that the white and black urnings of Carlsen were generally increasing over time, and that the difference between the white and the black urnings was decreasing. Interestingly, while on average for players the black urning is lower than the white urning, for Carlsen the two urnings became almost the same not long before he became the World Champion, suggesting that having a very high skill of playing black is very important for world class chess players.

3.3 | Movielens 20M

The MovieLens 20M data set consists of 20 000 263 reviews of 27 278 movies by 138 493 over a period from January 9, 1995 to March 31, 2015 (Harper and Konstan, 2015). In this application, the urnings of the users reflect how easy or hard it is to satisfy the user. If urnings are high, a user is likely to dislike even the very best movies. Similarly, if urnings are low, a user will like even movies of mediocre quality. The urning of a movie represents the general likeability or attractiveness of a movie. If high, it is liked by many persons. Combined, the difference between the urning of a movie and a user (on the logit scale) can determine the probability that a specific individual likes (e.g., provides a positive review of) a specific movie.

For the present analysis, movie reviews (0.5 - 5 stars with half-star increments) were dichotomized with four stars and higher comprising a *like* (1), all others a *dislike* (0). This gives a (close to) equal split between likes and dislikes. The

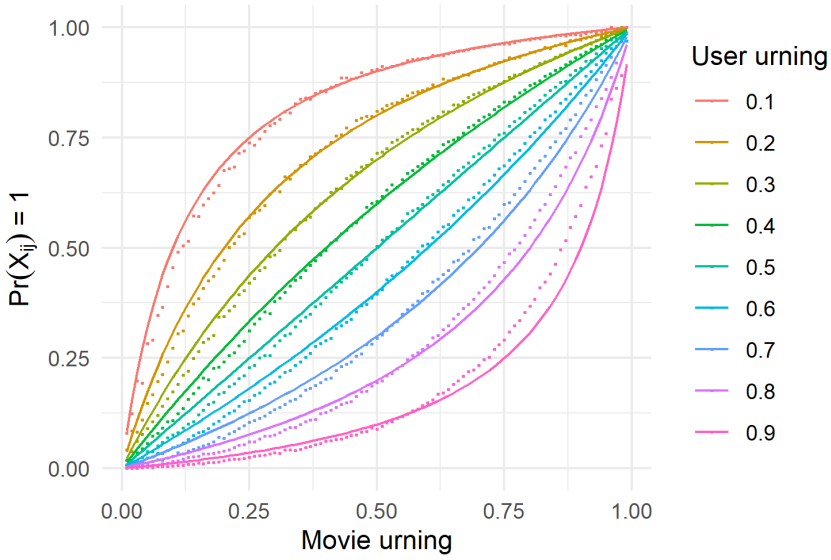


FIGURE 7 Expected and observed proportions of positive reviews (on the y-axis) for different combinations of the urnings of the movie (on the x-axis) and the urnings of the user (corresponding to a particular color).

observations were processed in chronological order. Since there is asymmetry between movies and raters in terms of the number of observations (i.e., there are a lot more observations per movie than there are per rater), we used different urn sizes: $n_{movie} = 100$, $n_{user} = 10$. For all raters and movies the urnings were initialized at random values sampled from uniform distributions.

Figure 7 shows the fit of the urnings to the observations. On the x-axis are the scaled urnings of the movies taken after the update, on the y-axis are the probabilities of a positive review; and each color corresponds to a particular value of the scaled urning of the user after the update. The lines represent the expected proportions of a positive review, and the points are the observed proportions. Only observations for which a movie has already been reviewed at least $3n_{movie}$ times and a user has already provided at least $3n_{user}$ reviews were included to remove the effect of the starting values. The observed proportions follow the expected ones rather closely. Larger deviations are present only for the combinations of high urnings of the movies and high urnings of the users. These deviations may be there because in the update we did not take into account how the movies were selected for the users, which could have been adaptive. Alternatively, these deviations could be a sign of model misfit. However, we can conclude that with just one parameter for the movie and one parameter for the user we already obtain a good fit.

Using the output of our analysis we can follow the development of the urnings of particular movies. Figure 8 shows the urnings of six Star War movies. One can see that the urnings of the three older episodes are higher than those of the three newer episodes. Note that for Episodes I, II, and III likeability starts relatively high (close to Episodes IV, V, and VI), but over time they decrease to values much lower than that of the earlier episodes.

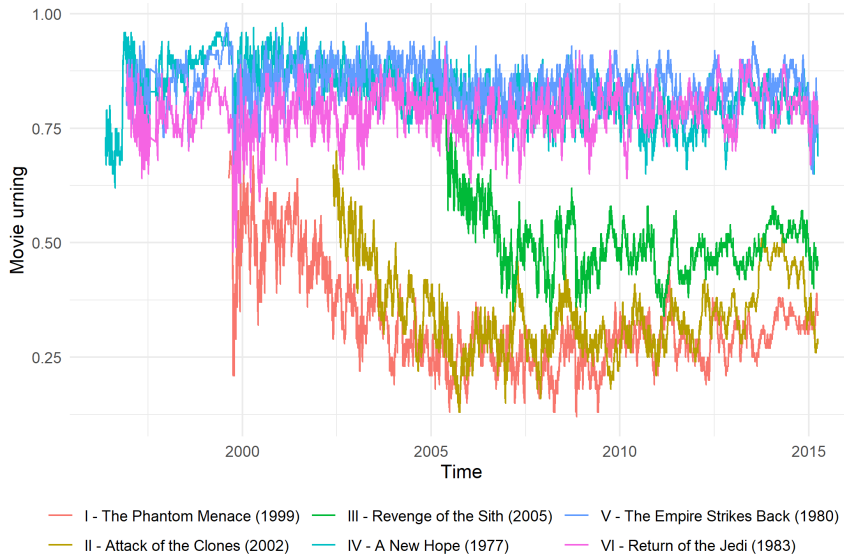


FIGURE 8 Development of the scaled urnings (likeability) of the six episodes of 'Star Wars' over time.

3.4 | Math Garden

Our third application is to the data of an online learning system Math Garden (Klinkenberg et al., 2011; Hofman et al., 2020), which is a platform for practicing primary school mathematics used widely in the Netherlands. Learners can practice various mathematics skills such as addition, fractions, and multiplication using series of exercises that are tailored to the level of their skill. For this analysis, we focused on the tables of multiplication exercises (100 in total) solved by a single birth cohort of learners (14 175 in total) over the course of three years (1 696 112 responses in total)[3]. Similar to the previous application, the number of observations per exercise versus per learner differ, therefore different urn sizes were used: $n_{exercise} = 200, n_{learner} = 20$. For all learners and all exercises the urnings were initialized at random starting values from 0 to n.

Figure 9 shows the predictive power of the urnings. All the responses given after the learner has already responded to $3n_{learner}$ exercises and after the exercise has been practiced by $3n_{exercises}$ learners were analysed. The dots represent the observed proportions of correct responses for each combination of a value of the urning of the exercise and the urning of the learner, while the line shows the expected proportion based on the urning values. There is generally a good match between expected and observed data.

For different exercises we can track how their difficulty was changing over time. In the tables of multiplication it is interesting to see how the difficulties of the exercises that differ only in the order of the numbers relate to each other. As an illustration we monitored the exercises '5x9' and '9x5' (see Figure 10, the urnings of each exercise are shown from the moment there were $3n_{exercise}$ responses to it). In the beginning when the learners are only starting with studying the tables of multiplication, the urning of '9x5' is slightly higher than that of '5x9' (i.e., it is more difficult), which is logical since children usually start with learning the tables of lower numbers first. After that the two urnings become very similar. Both items start as relatively difficult but their difficulty decreases over time.

[3] The same data set as in Brinkhuis and Maris (2020) was used

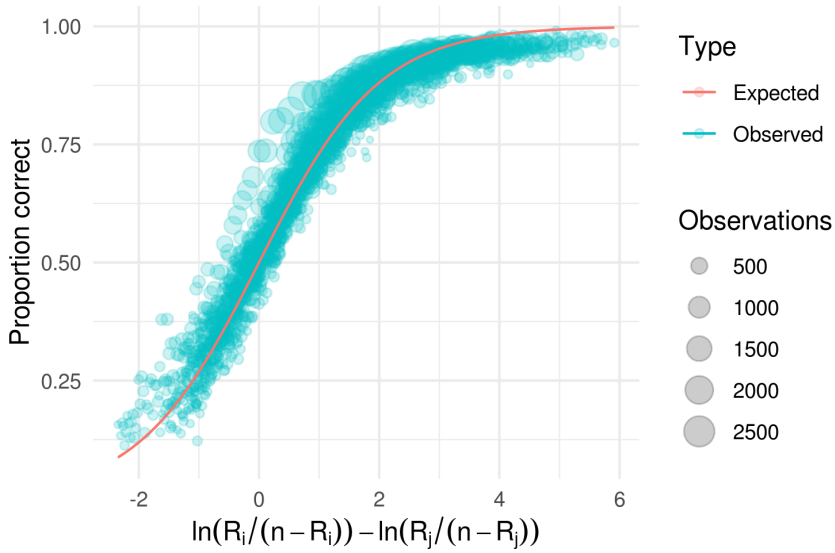


FIGURE 9 Expected and observed proportions of correct responses (on the *y*-axis). Each dot corresponds to a different combination of the urnings of the learner and the exercise in the system before the response (only combinations with at least 100 observations are plotted).

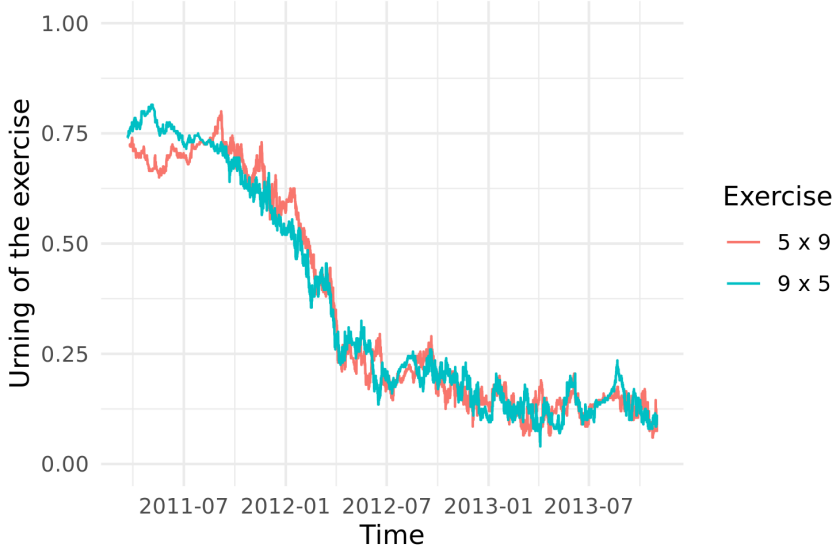


FIGURE 10 Development of the scaled urnings (difficulty) of the exercises '5x9' and '9x5' over time.

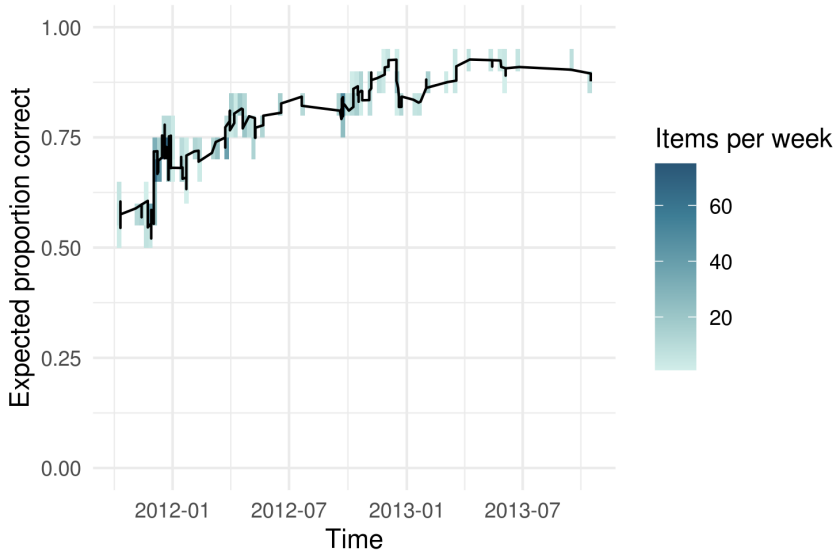


FIGURE 11 Development of a selected learner over time: Expected proportion of correct responses on the hundred tables of multiplication exercises

Analogously, we can monitor the development of each learner. In addition to simply looking at the development of the urnings, we can also monitor how the expected number of correct responses (based on the urning values) to the table of multiplication items changes over time. Figure 11 shows an example for one particular learner (shown from the moment that the learner responded to $3n_{learner}$ exercises).

4 | DISCUSSION

Our new rating system can be compared to other ratings system such as the popular Elo rating system (Elo, 1978; Batchelder and Bershad, 1979). Though parts are similar, like the conservation of the total number of rating points in the system, and the update based on the difference between the observed and the expected/simulated outcome, other aspects of the systems are different. Specifically, we can regard our rating system as a *tracker*, with the following two ideal properties (Brinkhuis and Maris, 2019): being able to estimate dynamically changing parameters and thus adapting to possible changes in their values, and provide unbiased estimates with a known error variance if no change occurs for some time. The Elo rating system holds this first property, but provides no standard errors that can be used for inference on the ratings. Other rating systems with standard errors exist (Glickman, 2001; Herbrich et al., 2006), which approximate the (possibly changing) variance of the ratings. The main difference with our rating system is that here variances are not approximated, but are derived from the invariant distribution of the ratings if no development in the parameters takes place. A unique feature that we have not yet seen in other rating systems is that in our rating update we explicitly take the adaptive matchmaking of players into account (or in the context of adaptive learning systems the adaptive selection of items for learners). As we have shown in the simulation study, this is important because if adaptive matchmaking is *not* taken into account, the variance of the ratings would artificially increase.

Clearly, such a rating variance inflation is problematic for making substantive inferences about the growth of different parameters in relation to each other. The good statistical properties of our rating system come at the price of ignoring some observations. This is undesirable in applications where the stakes of an individual game outcome are high, such as in some sports. However, it is not a problem in many other applications, especially if a lot of data is available, such as in learning systems, product ratings, etc.

Rating systems such as Elo's have parameters to tune the bias/variance trade-off. In practice these are important to account for fast changing ratings, for which large update steps are needed, or to be able to measure more precisely when no change occurs (Klinkenberg et al., 2011). In our rating system, the urn sizes (n) determines how fast ratings are allowed to change. With small n, a single point is quite a big step and hence the step size is rather large: the bias will therefore be relatively small, and the variance relatively large. Larger values of n will allow for more precise measurements if the true parameters are rather stable. Though implementing dynamically changing urn size is possible, how to optimally tune this parameter in different situations is a topic for future research. Related to the rate of change is the convergence of the invariant distribution. As we have shown in a simulation, the distribution of the urnings converges to the expected distribution, but a test for convergence in a dynamically changing system is also a topic for further research.

In this paper we introduced the Urnings algorithm for simple dichotomous observations (e.g., correct/incorrect responses in a learning system) and showed how it can be extended to polytomous outcomes (e.g., win/draw/loss in chess). Generalization for more than three possible outcomes, or more than two players are possible. Basically whenever a game can be conceptualized as resulting from drawing balls from infinite urns, we can set up a rating system using urns of finite size. For example, a rating system can be developed for learning systems where the probability of answering an item correctly depends on multiple abilities which we want to track over time. Furthermore, the updates in the rating system can be based not just on a single variable (e.g., accuracy of the response), but also other variables related to the same observation (e.g., response time, see Deonovic et al., 2020) or are based on a continuous outcome (Maris, 2020).

A | THEOREM PROOFS

A.1 | Proof of Theorem 1

The Markov chain resulting from the two players repeatedly playing each other is irreducible since every state in S can be reached from any other state in S in a finite number of steps and aperiodic since for every state there is a non-zero probability of staying in this state (i.e., when $X^*_{1,2} = X_{1,2}$), therefore it is ergodic.

Let us derive the distribution of (R^*_1, R^*_2) when (R_1, R_2) is distributed according to Equation 5. There are four possible ways in which the updated (R^*_1, R^*_2) can be in state (r_1, r_2):

$$\begin{aligned} p(R^*_1 = r_1, R^*_2 = r_2) = {} & p_{\pi_1,\pi_2}(r_1, r_2)\, p(X^*_{1,2} = 1 \mid R_1 = r_1, R_2 = r_2)\, p(X_{1,2} = 1) + \\ & p_{\pi_1,\pi_2}(r_1, r_2)\, p(X^*_{1,2} = 0 \mid R_1 = r_1, R_2 = r_2)\, p(X_{1,2} = 0) + \\ & p_{\pi_1,\pi_2}(r_1 + 1, r_2 - 1)\, p(X^*_{1,2} = 1 \mid R_1 = r_1, R_2 = r_2)\, p(X_{1,2} = 0) + \\ & p_{\pi_1,\pi_2}(r_1 - 1, r_2 + 1)\, p(X^*_{1,2} = 0 \mid R_1 = r_1, R_2 = r_2)\, p(X_{1,2} = 1). \end{aligned} \quad (15)$$

Using the following binomial identities:

$$\binom{n}{r-1} = \frac{r}{n-r+1}\binom{n}{r} \tag{16}$$

and

$$\binom{n}{r+1} = \frac{n-r}{r+1}\binom{n}{r} \tag{17}$$

we can show that

$$p_{\pi_1,\pi_2}(r_1-1, r_2+1)p(X^*_{1,2}=0 \mid R_1=r_1, R_2=r_2)p(X_{1,2}=1) = p_{\pi_1,\pi_2}(r_1,r_2)p(X^*_{1,2}=1 \mid R_1=r_1, R_2=r_2)p(X_{1,2}=0), \tag{18}$$

from which follows that the sum of the first and the fourth element in Equation 15 is equal to

$$p_{\pi_1,\pi_2}(r_1,r_2)p(X^*_{1,2}=1 \mid R_1=r_1, R_2=r_2). \tag{19}$$

Analogously, it can be shown that the sum of the second and the third elements in Equation 15 is equal to

$$p_{\pi_1,\pi_2}(r_1,r_2)p(X^*_{1,2}=0 \mid R_1=r_1, R_2=r_2). \tag{20}$$

Therefore,

$$p(R^*_1=r_1, R^*_2=r_2) = p_{\pi_1,\pi_2}(r_1,r_2)(p(X^*_{1,2}=1 \mid R_1=r_1, R_2=r_2)+p(X^*_{1,2}=0 \mid R_1=r_1, R_2=r_2)) = p_{\pi_1,\pi_2}(r_1,r_2), \tag{21}$$

which completes the proof.

A.2 | Proof of Theorem 2

Ergodicity of the chain is still satisfied because adding the acceptance probability does not prevent the chain from going from every state in S to every other state in S or from staying in the same state with non-zero probability.

The detailed balance condition which we need to prove is:

$$p_Z(r^*_1, r^*_2 \mid r_1, r_2, \tilde{r}_1, \tilde{r}_2)p_Y(\tilde{r}_1, \tilde{r}_2 \mid r_1, r_2)p_{\pi_1,\pi_2}(r_1,r_2) = p_Z(r_1, r_2 \mid r^*_1, r^*_2, \tilde{r}_1, \tilde{r}_2)p_Y(\tilde{r}_1, \tilde{r}_2 \mid r^*_1, r^*_2)p_{\pi_1,\pi_2}(r^*_1, r^*_2), \tag{22}$$

where $p_X(1,2 \mid r_1, r_2)$ and conditioning on $I=1$ and $J=2$ are not included into the equation because $(1,2)$ is the only pair that can be selected. The distribution of the updated urnings given the current urnings and the proposed values is the following

$$p_Z(r^*_1, r^*_2 \mid r_1, r_2, \tilde{r}_1, \tilde{r}_2) = \begin{cases} \min\left(1, \frac{r_1(n-r_2)+(n-r_1)r_2}{\tilde{r}_1(n-\tilde{r}_2)+(n-\tilde{r}_1)\tilde{r}_2}\right) \text{ if } (r^*_1, r^*_2) = (\tilde{r}_1, \tilde{r}_2) \\ \max\left(0, 1-\frac{r_1(n-r_2)+(n-r_1)r_2}{\tilde{r}_1(n-\tilde{r}_2)+(n-\tilde{r}_1)\tilde{r}_2}\right) \text{ if } (r^*_1, r^*_2) = (r_1, r_2) \\ 0 \text{ otherwise.} \end{cases} \tag{23}$$

Since

$$p_Y(\tilde{r}_1, \tilde{r}_2 \mid r_1, r_2) = \begin{cases} \frac{r_1(n-r_2)\pi_1(1-\pi_2)+r_2(n-r_1)\pi_2(1-\pi_1)}{(r_1(n-r_2)+r_2(n-r_1))(\pi_1(1-\pi_2)+\pi_2(1-\pi_1))} \text{ for } (\tilde{r}_1, \tilde{r}_2) = (r_1, r_2), \\ \frac{r_2(n-r_1)\pi_1(1-\pi_2)}{(r_1(n-r_2)+r_2(n-r_1))(\pi_1(1-\pi_2)+\pi_2(1-\pi_1))} \text{ for } (\tilde{r}_1, \tilde{r}_2) = (r_1+1, r_2-1), \\ \frac{r_1(n-r_2)\pi_2(1-\pi_1)}{(r_1(n-r_2)+r_2(n-r_1))(\pi_1(1-\pi_2)+\pi_2(1-\pi_1))} \text{ for } (\tilde{r}_1, \tilde{r}_2) = (r_1-1, r_2+1), \\ 0 \text{ for all other values of } (\tilde{r}_1, \tilde{r}_2), \end{cases} \tag{24}$$

it is sufficient to show that detailed balance is satisfied for the three updates that are possible. For the first one it is trivial. For the second one, one can derive using the binomial identities in Equations 16 and 17 that both the right and the left side of Equation 22 are equal to

$$\frac{\binom{n}{r_1}\pi_1^{r_1+1}(1-\pi_1)^{n-r_1}\binom{n}{r_2}\pi_2^{r_2}(1-\pi_2)^{n-r_2+1}\mathcal{I}_{r_1+r_2=r_+}}{\max\left((r_1(n-r_2)+r_2(n-r_1)), ((r_1+1)(n-r_2+1)+(r_2-1)(n-r_1-1))\right)(\pi_1(1-\pi_2)+\pi_2(1-\pi_1))Z^*}. \tag{25}$$

Finally, an analogous derivation can be done for the third update, which completes the proof.

A.3 | Proof of Theorem 3

Ergodicity of the chain follows by induction from ergodicity of the chain for just two players since there is a non-zero probability for each pair of players to play against each other.

For $\mathbf{r}^* = \mathbf{r}$ the detailed balance condition is trivial. Among $\mathbf{r}^* \neq \mathbf{r}$ there are only two possible updates for each unique pair (i,j): $r_i^* = r_i + 1, r_j^* = r_j - 1, r_k^* = r_k, \forall k \neq i,j$ and $r_i^* = r_i - 1, r_j^* = r_j + 1, r_k^* = r_k, \forall k \neq i,j$. For the first one, both $p_P(\mathbf{r}^* \mid \mathbf{r})p_{\boldsymbol{\pi}}(\mathbf{r})$ and $p_P(\mathbf{r} \mid \mathbf{r}^*)p_{\boldsymbol{\pi}}(\mathbf{r}^*)$ are equal to:

$$\frac{2\binom{n}{r_i}\pi_i^{r_i+1}(1-\pi_i)^{n-r_i}\binom{n}{r_j}\pi_j^{r_j}(1-\pi_j)^{n-r_j+1}\mathcal{I}_{\sum_k r_k=r_+}\prod_{k\neq i,j}\binom{n}{r_k}\pi_k^{r_k}(1-\pi_k)^{n-r_k}}{\max\left((r_i(n-r_j)+r_j(n-r_i)), ((r_i+1)(n-r_j+1) + (r_j-1)(n-r_i-1))\right)(\pi_i(1-\pi_j)+\pi_j(1-\pi_i))N(N-1)F(\boldsymbol{\pi}, r_+)}, \tag{26}$$

with an analogous expression for the second one. Hence, detailed balance is satisfied, which completes the proof.

A.4 | Proof of Theorem 4

The chain is ergodic since it is irreducible and aperiodic (there is a non-zero probability to stay in the same state after the update). For $\mathbf{r}^* = \mathbf{r}$ the detailed balance condition is trivial. Among $\mathbf{r}^* \neq \mathbf{r}$ there are only two possible updates for each unique pair (i,j): $r_i^* = r_i + 1, r_j^* = r_j - 1, r_k^* = r_k, \forall k \neq i,j$ and $r_i^* = r_i - 1, r_j^* = r_j + 1, r_k^* = r_k, \forall k \neq i,j$. For the first one, both $p_P(\mathbf{r}^* \mid \mathbf{r})p_{\boldsymbol{\pi}}(\mathbf{r})$ and $p_P(\mathbf{r} \mid \mathbf{r}^*)p_{\boldsymbol{\pi}}(\mathbf{r}^*)$ are equal to:

$$M\frac{\binom{n}{r_i}\pi_i^{r_i+1}(1-\pi_i)^{n-r_i}\binom{n}{r_j}\pi_j^{r_j}(1-\pi_j)^{n-r_j+1}\mathcal{I}_{\sum_k r_k=r_+}\prod_{k\neq i,j}\binom{n}{r_k}\pi_k^{r_k}(1-\pi_k)^{n-r_k}}{F(\boldsymbol{\pi}, r_+)}, \tag{27}$$

where

$$M = \min\left(\frac{p_X(i,j \mid \mathbf{r})}{r_i(n-r_j)+(n-r_i)r_j}\frac{p_X(i,j \mid R_i = r_i+1, R_j = r_j-1, R_k = r_k, \forall k \neq i,j)}{(r_i+1)(n-r_j+1)+(r_j-1)(n-r_i-1)}\right) \tag{28}$$

with an analogous expression for the second one. Hence, detailed balance is satisfied, which completes the proof.

A.5 | Proof of Theorem 5

The product of independent binomials can be re-written as a function of ρ_i:

$$p(\mathbf{r}) = \prod_{i=1}^{N} \binom{n}{r_i} \pi_i^{r_i} (1-\pi_i)^{n-r_i} = \frac{\prod_{i=1}^{N} \binom{n}{r_i} \rho_i^{r_i}}{\prod_{i=1}^{N} (1+\rho_i)^n}. \tag{29}$$

And its denominator can be re-written as follows:

$$\prod_{i=1}^{N} (1+\rho_i)^n = \prod_{i=1}^{N} \sum_{s=0}^{n} \binom{n}{s} \rho_i^s = \sum_{t=0}^{nN} \gamma_t(\boldsymbol{\rho}). \tag{30}$$

If the R_i are binomally distributed then their sum has the following distribution:

$$p(r_+) = \frac{\gamma_{r_+}(\boldsymbol{\rho})}{\sum_{t=0}^{nN} \gamma_t(\boldsymbol{\rho})}; \tag{31}$$

hence

$$p(\mathbf{r}|r_+) = \frac{p(\mathbf{r})}{p(r_+)} = \frac{\prod_i \binom{n}{r_i} \rho_i^{r_i}}{\gamma_{r_+}(\boldsymbol{\rho})}, \tag{32}$$

which completes the proof.

A.6 | Proof of Theorem 6

From the Markov property, we obtain that

$$p_{t+1}(X = x) = \sum_{y=0}^{n} p_\infty(X = x | Y = y) p_t(Y = y) \tag{33}$$

which using Bayes theorem, we may rewrite as

$$p_{t+1}(X = x) = \sum_{y=0}^{n} \frac{p_\infty(Y = y | X = x) p_\infty(X = x)}{p_\infty(Y = y)} p_t(Y = y) \tag{34}$$

Hence, we find that

$$\frac{p_{t+1}(X = x)}{p_\infty(X = x)} = \sum_{y=0}^{n} \frac{p_t(Y = y)}{p_\infty(Y = y)} p_\infty(Y = y | X = x) \tag{35}$$

such that taking logarithms on both sides and integrating with respect to $p_\infty(X = x)$ we have that

$$\sum_{x=0}^{n} \log\left(\frac{p_{t+1}(X=x)}{p_\infty(X=x)}\right) p_\infty(X=x) = \sum_{x=0}^{n} \log\left(\sum_{y=0}^{n} \frac{p_t(Y=y)}{p_\infty(Y=y)} p_\infty(Y=y|X=x)\right) p_\infty(X=x) \tag{36}$$

Using Jensen's inequality, we find that

$$\begin{aligned}
&\sum_{x=0}^{n} \log\left(\sum_{y=0}^{n} \frac{p_t(Y=y)}{p_\infty(Y=y)} p_\infty(Y=y|X=x)\right) p_\infty(X=x) \\
&\quad \geq \sum_{x=0}^{n}\sum_{y=0}^{n} \log\left(\frac{p_t(Y=y)}{p_\infty(Y=y)}\right) p_\infty(Y=y|X=x) p_\infty(X=x) = \sum_{y=0}^{n} \log\left(\frac{p_t(Y=y)}{p_\infty(Y=y)}\right) p_\infty(Y=y)
\end{aligned} \tag{37}$$

which completes the proof.

B | GENERALISATION TO POLYTOMOUS OUTCOMES

Many games not only allow for wins and losses, but also draws, and we adapt our representation of the observation to reflect these as well. Consistent with its interpretation in chess we consider a draw to represent one win and one loss (in actual fact it would be half a win and half a loss, but we prefer whole scores), so we let our players compete against each other (conceptually) twice:

for $k = 0 \ldots 1$ **do**
  repeat
    $Y_{ik} \sim Bernoulli(\pi_i)$
    $Y_{jk} \sim Bernoulli(\pi_j)$
  until $Y_{ik} \neq Y_{jk}$
end for
return $X_{ij} = \sum_k Y_{ik}$

which gives us the following probabilities for the three outcomes (coded as i wins giving $X_{ij} = 2$, a draw giving $X_{ij} = 1$, and j wins giving $X_{ij} = 0$):

$$\Pr(X_{ij} = 2|\pi_i, \pi_j) = \frac{(\pi_i(1-\pi_j))^2}{(\pi_i(1-\pi_j))^2 + 2\pi_i(1-\pi_j)\pi_j(1-\pi_i) + (\pi_j(1-\pi_i))^2}; \tag{38}$$

$$\Pr(X_{ij} = 1|\pi_i, \pi_j) = \frac{2\pi_i(1-\pi_j)\pi_j(1-\pi_i)}{(\pi_i(1-\pi_j))^2 + 2\pi_i(1-\pi_j)\pi_j(1-\pi_i) + (\pi_j(1-\pi_i))^2}; \tag{39}$$

$$\Pr(X_{ij} = 0|\pi_i, \pi_j) = \frac{(\pi_j(1-\pi_i))^2}{(\pi_i(1-\pi_j))^2 + 2\pi_i(1-\pi_j)\pi_j(1-\pi_i) + (\pi_j(1-\pi_i))^2}, \tag{40}$$

where the 2 for a draw indicates the two ways in which our game of chance can end in a draw. From dividing every numerator and denominator by $(\pi_j(1-\pi_i))^2$, we see that all probabilities only depend on the difference between the logits of π_i and π_j, which makes it a proper generalization of the Bradley-Terry-Luce model.

It is clear that for this polytomous extension, the probability of a draw equals 0.5 for two players of equal strength. This is unnecessarily restrictive, and we consider how to relax this assumption. We reject the outcomes of the game with a certain probability, depending on the outcome, and repeat until an outcome is accepted. That is, we make use

of rejection sampling (Ripley, 1987). The new process translates into the following pseudo-code:

repeat
  for $k = 1 \ldots 2$ **do**
    repeat
      $Y_{ik} \sim Bernoulli(\pi_i)$
      $Y_{jk} \sim Bernoulli(\pi_j)$
    until $Y_{ik} \neq Y_{jk}$
  end for
  $u \sim U(0,1)$
until $u \leq \text{accept}(\mathbf{Y}_i, \mathbf{Y}_j)$
return $X_{ij} = \sum_k Y_{ik}$

in which the acceptance probabilities are the following:

$$\text{accept}(\text{win,win}) = a \tag{41}$$

$$\text{accept}(\text{win,loose}) = b \tag{42}$$

$$\text{accept}(\text{loose,loose}) = c \tag{43}$$

which gives us the following probabilities for the three outcomes:

$$p(X_{ij} = 2|\pi_i, \pi_j) = \frac{a(\pi_i(1-\pi_j))^2}{a(\pi_i(1-\pi_j))^2 + 2b\pi_i(1-\pi_j)\pi_j(1-\pi_i) + c(\pi_j(1-\pi_i))^2}; \tag{44}$$

$$p(X_{ij} = 1|\pi_i, \pi_j) = \frac{2b\pi_i(1-\pi_j)\pi_j(1-\pi_i)}{a(\pi_i(1-\pi_j))^2 + 2b\pi_i(1-\pi_j)\pi_j(1-\pi_i) + c(\pi_j(1-\pi_i))^2}; \tag{45}$$

$$p(X_{ij} = 0|\pi_i, \pi_j) = \frac{c(\pi_j(1-\pi_i))^2}{a(\pi_i(1-\pi_j))^2 + 2b\pi_i(1-\pi_j)\pi_j(1-\pi_i) + c(\pi_j(1-\pi_i))^2}. \tag{46}$$

Note that the probabilities only depend on the ratios a/c and b/c, hence giving us two additional degrees of freedom, and that the probabilities still only depend on the difference between the logits of π_i and π_j.

references

Albers, P. C. and Vries, H. d. (2001) Elo-rating as a tool in the sequential estimation of dominance strengths. *Animal Behaviour*, 489–495.

Arulampalam, M. S., Maskell, S., Gordon, N. and Clapp, T. (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on signal processing*, **50**, 174–188.

Batchelder, W. H. and Bershad, N. J. (1979) The statistical analysis of a Thurstonian model for rating chess players. *Journal of Mathematical Psychology*, **19**, 39–60.

Batchelder, W. H., Bershad, N. J. and Simpson, R. S. (1992) Dynamic paired-comparison scaling. *Journal of Mathematical Psychology*, **36**, 185–212.

Bradley, R. A. and Terry, M. E. (1952) Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, **39**, 324–345.

Brinkhuis, M. J. S. and Maris, G. (2009) Dynamic parameter estimation in student monitoring systems. *Measurement and Research Department Reports 09-01*, Cito, Arnhem. URL: `https://www.researchgate.net/publication/242357963`.

— (2019) Tracking ability: Defining trackers for measuring educational progress. In *Theoretical and Practical Advances in Computer-based Educational Measurement* (eds. B. P. Veldkamp and C. Sluijter), Methodology of Educational Measurement and Assessment, chap. 8, 161–173. Cham: Springer International Publishing.

— (2020) Dynamic estimation in the extended marginal Rasch model with an application to mathematical computer-adaptive practice. *British Journal of Mathematical and Statistical Psychology*, **73**, 72–87.

Brinkhuis, M. J. S., Savi, A. O., Coomans, F., Hofman, A. D., van der Maas, H. L. J. and Maris, G. (2018) Learning as it happens: A decade of analyzing and shaping a large-scale online learning system. *Journal of Learning Analytics*, **5**, 29–46.

Deonovic, B., Bolsinova, M., Bechger, T. and Maris, G. (2020) A Rasch model and rating system for continuous responses collected in large-scale learning systems. *Frontiers in psychology*, **11**.

Elo, A. E. (1978) *The rating of chess players, past and present*. London: B. T. Batsford, Ltd.

Glickman, M. E. (2001) Dynamic paired comparison models with stochastic variances. *Journal of Applied Statistics*, **28**, 673–689.

Harper, F. M. and Konstan, J. A. (2015) The MovieLens datasets. *ACM Transactions on Interactive Intelligent Systems*, **5**, 1–19.

Herbrich, R., Minka, T. and Graepel, T. (2006) TrueSkill: A Bayesian skill rating system. In *Advances in Neural Information Processing Systems* (eds. B. Schölkopf, J. Platt and T. Hoffman), vol. 19, 569–576. Cambridge, MA: MIT Press. URL: `http://books.nips.cc/nips19.html`.

Hofman, A. D., Brinkhuis, M. J., Bolsinova, M., Klaiber, J., Maris, G. and van der Maas, H. L. (2020) Tracking with (un) certainty. *Journal of Intelligence*, **8**, 10.

Hvattum, L. M. and Arntzen, H. (2010) Using Elo ratings for match result prediction in association football. *International Journal of forecasting*, **26**, 460–470.

Jansen, B. R., Louwerse, J., Straatemeier, M., Van der Ven, S. H., Klinkenberg, S. and Van der Maas, H. L. (2013) The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences*, **24**, 190–197.

Johnson, N. L. and Kotz, S. (1977) *Urn models and their application*. Wiley series in probability and mathematical statistics. New York: John Wiley & Sons.

Kalman, R. E. (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, **82**, 35–45.

Klinkenberg, S., Straatemeier, M. and van der Maas, H. L. J. (2011) Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, **57**, 1813–1824.

Lehmann, R. and Wohlrabe, K. (2017) Who is the 'Journal Grand Master'? A new ranking based on the Elo rating system. *Journal of Informetrics*, **11**, 800–809.

Luce, R. D. (1959) *Individual Choice Behavior: A Theoretical Analysis*. New York: John Wiley & Sons, Inc.

Mangan, S. and Collins, K. (2016) A rating system for Gaelic football teams: Factors that influence success. *International Journal of Computer Science in Sport*, **15**, 78–90.

Maris, G. (2020) The Duolingo English Test: Psychometric considerations. *Tech. rep.*, Technical Report DRR-20-02, Duolingo.

Minka, T., Cleven, R. and Zaykov, Y. (2018) Trueskill 2: An improved Bayesian skill rating system. *Tech. rep.*, Microsoft Research. URL: `https://www.microsoft.com/en-us/research/publication/trueskill-2-improved-bayesian-skill-rating-system/`.

Neumann, C., Duboscq, J., Dubuc, C., Ginting, A., Irwan, A. M., Agil, M., Widdig, A. and Engelhardt, A. (2011) Assessing dominance hierarchies: validation and advantages of progressive evaluation with Elo-rating. *Animal Behaviour*, **82**, 911–921.

Pelánek, R. (2016) Applications of the Elo rating system in adaptive educational systems. *Computers & Education*, **98**, 169–179.

R Core Team (2018) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: `https://www.R-project.org/`.

Rasch, G. (1960) *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute of Educational Research. Expanded edition, 1980. Chicago: The University of Chicago Press.

Ripley, B. D. (1987) *Stochastic Simulation*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Welch, G. and Bishop, G. (1995) An introduction to the Kalman filter. *Tech. Rep. TR 95-041*, University of North Carolina at Chapel Hill, Department of Computer Science, Chapel Hill, NC, USA. URL: `http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf`. Updated July 24, 2006.

Zohaib, M. (2018) Dynamic difficulty adjustment (DDA) in computer games: A review. *Advances in Human-Computer Interaction*, **2018**, 1–12.