EPJ B
Condensed Matter and Complex Systems
www.epj.org

Eur. Phys. J. B **67**, 473–481 (2009) DOI: 10.1140/epjb/e2008-00405-5

Understanding baseball team standings and streaks

C. Sire and S. Redner

EDP SCIENCES
Società Italiana di Fisica
Springer

Eur. Phys. J. B **67**, 473–481 (2009)
DOI: 10.1140/epjb/e2008-00405-5

THE EUROPEAN
PHYSICAL JOURNAL B

Regular Article

Understanding baseball team standings and streaks

C. Sire[1] and S. Redner[2,a]

[1] Laboratoire de Physique Théorique - IRSAMC, CNRS, Université Paul Sabatier, 31062 Toulouse, France
[2] Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA

Received 29 July 2008 / Received in final form 8 October 2008
Published online 5 November 2008 –

Abstract. Can one understand the statistics of wins and losses of baseball teams? Are their consecutive-game winning and losing streaks self-reinforcing or can they be described statistically? We apply the Bradley-Terry model, which incorporates the heterogeneity of team strengths in a minimalist way, to answer these questions. Excellent agreement is found between the predictions of the Bradley-Terry model and the rank dependence of the average number team wins and losses in major-league baseball over the past century when the distribution of team strengths is taken to be uniformly distributed over a finite range. Using this uniform strength distribution, we also find very good agreement between model predictions and the observed distribution of consecutive-game team winning and losing streaks over the last half-century; however, the agreement is less good for the previous half-century. The behavior of the last half-century supports the hypothesis that long streaks are primarily statistical in origin with little self-reinforcing component. The data further show that the past half-century of baseball has been more competitive than the preceding half-century.

PACS. 89.75.-k Complex systems – 02.50.Cw Probability theory

1 Introduction

The physics of systems involving large numbers of interacting agents is currently a thriving field of research [1]. One of its many appeals lies in the opportunity it offers to apply precise methods and tools of physics to the realm of "soft" science. In this respect, biological, economic, and a large variety of human systems present many examples of competitive dynamics that can be studied qualitatively or even quantitatively by statistical physics. Among them, sports competitions are particularly appealing because of the large amount of data available, their popularity, and the fact that they constitute almost perfectly *isolated* systems. Indeed, most systems considered in econophysics [2] or evolutionary biology [3] are strongly affected by external and often unpredictable factors. For instance, a financial model cannot predict the occurrence of wars or natural disasters which dramatically affect financial markets, nor can it include the effect of many other important external parameters (China's GDP growth, German exports, Google's profit...). On the other hand, sport leagues (soccer [4], baseball [5], football [6]...) or tournaments (basketball [7,8], poker [9]...) are basically isolated systems that are much less sensitive to external influences. Hence, despite their intrinsic human nature, which actually contribute to their appeal, competitive sports are particularly suited to quantitative theoretical modeling. In this spirit, this work is focused on basic statistical features of game outcomes in Major-League baseball.

In Major-League baseball and indeed in any competitive sport, the main observable is the outcome of a single game – who wins and who loses. Then at the end of a season, the win/loss record of each team is fundamental. As statistical physicists, we are not concerned with the fates of individual teams, but rather with the average win/loss record of the 1st, 2nd, 3rd, etc. teams, as well as the statistical properties of winning and losing streaks. We concentrate on major-league baseball to illustrate statistical properties of game outcomes because of the large amount of available data [10] and the near constancy of the game rules during the so-called "modern era" that began in 1901.

For non-US readers or for non-baseball fans, during the modern era of major-league baseball, teams have been divided into the nearly-independent American and National leagues [11]. At the end of each season a champion of the American and National leagues is determined (by the best team in each league prior to 1961 and by league playoffs subsequently) that play in the World Series to determine the champion. As the data will reveal, it is also useful to separate the 1901–1960 early modern era, with a 154-game season and 16 teams, and the 1961–2005 expansion era, with a 162-game season in which the number of teams expanded in stages to its current value of 30, to highlight systematic differences between these two periods. Our data is based on the 163674 regular-season games that have

[a] e-mail: uebing@mpie-duesseldorf.mpg.de

occurred between 1901 and the end of the 2005 season (72741 between 1901–60 and 90933 between 1961–2005).

While the record of each team can change significantly from year to year, we find that the *time average* win/loss record of the rth-ranked team as a function of rank r is strikingly regular. One of our goals is to understand the rank dependence of this win fraction. An important outcome of our study is that the Bradley-Terry (BT) competition model [12,13] provides an excellent account of the team win/loss records. This agreement between the data and theory is predicated on using a specific form for the distribution of team strengths. We will argue that the best match to the data is achieved by using a uniform distribution of teams strengths in each season.

Another goal of this work is to understand the statistical features of consecutive-game team winning and losing streaks. The existence of long streaks of all types of exceptional achievement in baseball, as well as in most competitive sports, have been well documented [14] and continue to be the source of analysis and debate among sports fans. For long consecutive-game team winning and team losing streaks, an often-invoked theme is the notion of reinforcement – a team that is "on a roll" is more likely to continue winning, and vice versa for a slumping team on a losing streak. The question of whether streaks are purely statistical or self reinforcing continues to be vigorously debated [15]. Using the BT model and our inferred uniform distribution of team strengths, we compute the streak length distribution. We find that the theoretical prediction agrees extremely well with the streak data during 1961–2005. However, there is a slight discrepancy between theory and the tail of the streak distribution during 1901–60, suggesting that non-statistical effects may have played a role during this early period.

As a byproduct of our study, we find clear evidence that baseball has been more competitive during 1961–2005 than during 1901–60 and feature that has been found previously [16]. The manifestation of this increased competitiveness is that the range of team records and the length of streaks was narrower during the latter period. This observation fits with the general principle [17] that outliers become progressively rarer in a highly competitive environment. Consequently, extremes of achievement become less and less likely to occur.

2 Statistics of the win fraction

2.1 Bradley-Terry model

Our starting point to account for the win/loss records of all baseball teams is the BT model [12,13] that incorporates the heterogeneity in team strengths in a natural and simple manner. We assume that each team has an intrinsic strength x_i that is fixed for each season. The probability that a team of strength x_i wins when it plays a team of strength x_j is simply

$$p_{ij} = \frac{x_i}{x_i + x_j}. \tag{1}$$

Thus the winning probability depends continuously on the strengths of the two competing teams [18]. When two equal-strength teams play, each team has a 50% probability to win, while if one team is much stronger, then its winning probability approaches 1.

The form of the winning probability of equation (1) is quite general. Indeed, we can replace the team strength x_i by any monotonic function $f(x_i)$. The only indispensable attribute is the ordering of the team strengths. Thus the notion of strength is coupled to the assumed form of the winning probability. If we make a hypothesis about one of these quantities, then the other is no longer a variable that we are free to choose, but an outcome of the model. In our analysis, we adopt the form of the winning probability in equation (1) because of its simplicity. Then the only relevant unknown quantity is the probability distribution of the x_i's. As we shall see in the next section, this distribution of team strengths can then be inferred from the season-end win/loss records of the teams, and a good fit to the data is obtained when assuming a uniform distribution of team strengths. Because only the ratio of team strengths is relevant in equation (1), we therefore take team strengths to be uniformly distributed in the range $[x_{min}, 1]$, with $0 \leq x_{min} \leq 1$. Thus the only model parameter is the value of x_{min}.

For uniformly distributed team strengths $\{x_j\}$ that lie in $[x_{min}, 1]$, the average winning fraction for a team of strength x that plays a large number of games N, with equal frequencies against each opponent is

$$\begin{aligned} W(x) &= \frac{1}{N} \sum_{j=1}^{N} \frac{x}{x + x_j} \\ &\to \frac{x}{1 - x_{min}} \int_{x_{min}}^{1} \frac{dy}{x + y} \\ &= \frac{x}{1 - x_{min}} \ln\left(\frac{x + 1}{x + x_{min}}\right), \end{aligned} \tag{2}$$

where we assume $N \to \infty$ in the second line. We then transform from strength x to scaled rank r by $x = x_{min} + (1 - x_{min})r$, with $r = 0, 1$ corresponding to the weakest and strongest team, respectively (Fig. 1). This result for the win fraction is one of our primary results.

To check the prediction of equation (2), we start with a value of $x_{\rm min}$ and simulate 10^4 periods of a model baseball league that consists of: (i) 16 teams that play 60 seasons of 154 games (corresponding to 1901–60) and (ii) 30 teams that play 45 seasons of 162 games (1961–2005), with uniformly distributed strengths in $[x_{min}, 1]$ for both cases, but with different values of x_{min}. Using the winning probability p_{ij} of equation (1), we then compute the average win fraction $W(r)$ of each team as function of its scaled rank r. We then incrementally update the value of x_{min} to minimize the difference between the simulated values of $W(r)$ with those from game win/loss data. Nearly the same results are found if each team plays every opponent with equal probability or equally often, as long as the number of teams and number of games is not unrealistically small. The BT model, with each team playing each opponent with the same probability, gives very good fits to the

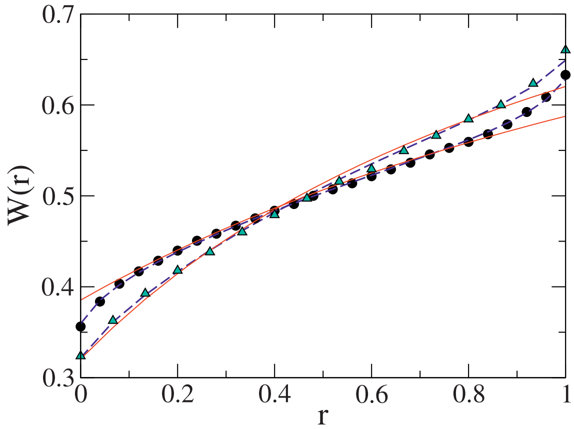

Fig. 1. Average win fraction $W(r)$ versus scaled rank r for 1901–60 (△) and 1961–2005 (○). For these periods, the dashed lines are simulation results for the BT model with $x_{min} = 0.278$ and 0.435 respectively. The solid curves represent equation (2), corresponding to simulations for an infinitely long season and an infinite number of teams.

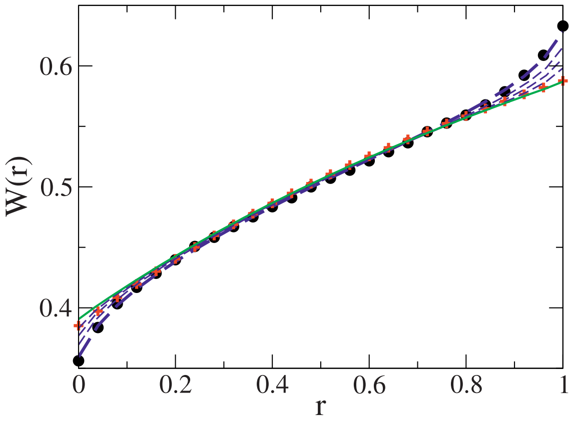

Fig. 2. Convergence of $W(r)$ versus scaled rank r as a function of season length for 1961–2005, using $x_{min} = 0.435$ and 30 teams. The circles and the thick dashed curve are the baseball data and the corresponding BT model data for a $n = 162$ game season. The thin dashed lines are model data for a season of $n = 300, 500$, and 1000 games averaged over 100000 seasons. The full line corresponds to the model for an infinitely long season with 30 teams. Finally, the + symbols give the result of equation (2), which corresponds to an infinite-length season and an infinite number of teams.

data by choosing $x_{min} = 0.278$ for the period 1901–60, and $x_{min} = 0.435$ for 1961–2005 (Fig. 1). If the actual game frequencies in each season are used to determine opponents, x_{min} changes slightly – to 0.289 for 1901–60 – but remains unchanged for 1961–2005.

Despite the fact that the number of teams has increased from 16 to 30 since in 1961, the range of win fractions is larger in the early era (0.32–0.67) than in the expansion era (0.36–0.63), a feature that indicates that baseball has become more competitive. This observation accords with the notion that the pressure of continuous competition, as in baseball, gradually diminishes the likelihood of outliers [17]. Given the crudeness of the model and real features that we have ignored, such as home-field advantage (approximately 53% for the past century and slowly decreasing with time), imbalanced playing schedules, and in-season personnel changes due to trades and player injuries, the agreement between the data and simulations of the BT model is satisfying.

It is worth noting in Figure 1 is that the win fraction data and the corresponding numerical results from simulations of the BT model deviate from the theoretical prediction given in equation (2) when $r \to 0$ and $r \to 1$. This discrepancy is simply a finite-season effect. As shown in Figure 2, when we simulate the BT model for progressively longer seasons, the win/loss data gradually converges to the prediction of equation (2).

The present model not only reproduces the average win record $W(r)$ over a given period, but it also correctly explains the season-to-season fluctuation $\sigma^2(r)$ of the win fraction defined as

$$\sigma^2(r) \equiv \frac{1}{Y}\sum_{j=1}^{Y}(W(r) - W_j(r))^2, \tag{3}$$

where $W_j(r)$ is the winning fraction of the rth-ranked team during the jth season and

$$W(r) = \frac{1}{Y}\sum_{j=1}^{Y} W_j(r),$$

is the average win fraction of the rth-ranked team and Y is the number of years in the period. These fluctuations are the largest for extremal teams (and minimal for average teams). There is also an asymmetry of $\sigma(r)$ with respect to $r = 1/2$. Our simulations of the BT model with the optimal x_{min} values that were determined previously by fitting to the win fraction quantitatively reproduce these two features of $\sigma(r)$.

In addition to the finite-season effects described above, another basic consequence of the finiteness of the season is that the intrinsically strongest team does not necessarily have the best win/loss record. That is, the average win fraction W does not necessarily increase with team strength. By luck, a strong team can have a poor record or vice versa. It is instructive to estimate the number of games G that need to be played to ensure that the win/loss record properly reflects team strength. The difference in the number of wins of two adjacent teams in the standings is proportional to $G(1 - x_{min})/T$, namely, the number of games times their strength difference; the latter is proportional to $(1 - x_{min})/T$ for a league that consists of T teams. This systematic contribution to the difference should significantly exceed random fluctuations, which are of the order of $\sqrt{G}$. Thus we require

$$G \gg \left(\frac{T}{1 - x_{min}}\right)^2 \tag{4}$$

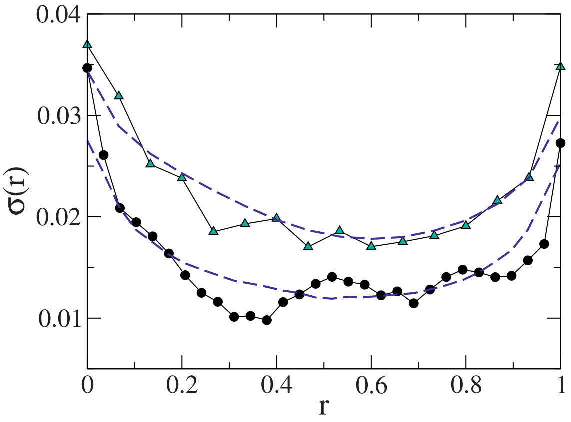

Fig. 3. Season-to-season fluctuation $\sigma(r)$ for 1901–60 (△) and for 1961–2005 (○). The dashed lines are numerical simulations of the BT model for 10^4 periods with the same x_{min} as in Figure 1.

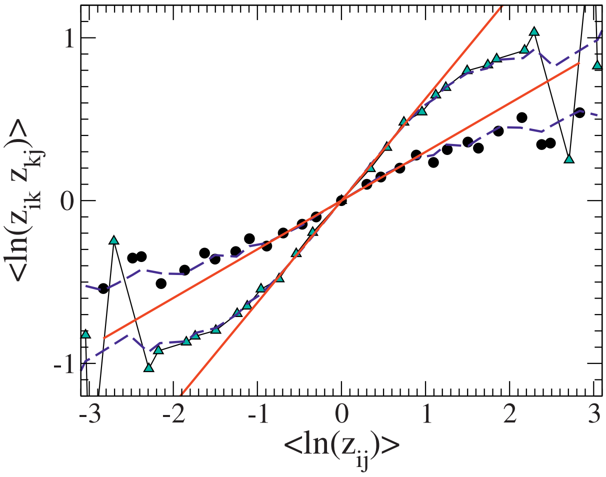

Fig. 4. Comparison of the detailed balanced relation equation (8) for baseball data to the results of the BT model over 10^4 periods (dashed lines), where each period corresponds to the results of all baseball games during either 1901–60 (triangles) or 1960–2005 (circles). The x_{min} values are the same as in Figure 1. The straight lines are guides for the eye, with slope 0.63 for the data for 1901–60 and 0.30 for 1961–2005.

for the end-of-season standings to be ordered by team strength. Figures 2 and 3 illustrate the fact that this effect is more important for the top-ranked and bottom-ranked teams. During the 1901–60 period, when major-league baseball consisted of independent American and National leagues, $T = 8$, $G = 154$, and $x_{min} \approx 0.3$, so that the season was just long enough to resolve adjacent teams. Currently, however, the season length is insufficient to resolve adjacent teams. The natural way to deal with this ambiguity is to expand the number of teams that qualify for the post-season playoffs, which is what is currently done.

2.2 Applicability of the Bradley-Terry model

Does the BT model with uniform teams strength provide the most appropriate description of the win/loss data? We perform several tests to validate this model. First, as mentioned in the previous section, the assumption (1) for the winning probability can be recast more generally as

$$p_{ij} = \frac{f(x_i)}{f(x_i) + f(x_j)}, \tag{5}$$

so that an arbitrary $X_i = f(x_i)$ reduces to the original winning probability in equation (1). Hence the crucial model assumption is the separability of the winning probability. In particular, the BT model assumes that $p_{ij}/p_{ji} = p_{ij}/(1-p_{ij})$ is *only* a function of characteristics of team i, divided by characteristics of team j. One consequence of this separability is the "detailed-balance" relation

$$\frac{p_{ik}}{1-p_{ik}}\frac{p_{kj}}{1-p_{kj}} = \frac{p_{ij}}{1-p_{ij}}, \tag{6}$$

for any triplet of teams. This relation quantifies the obvious fact that if team A likely beats B, and B likely beats C, then A is likely to beat C. Since we do not know the actual p_{ij} in a given baseball season, we instead consider

$$z_{ij} = \frac{W_{ij}}{G_{ij} - W_{ij}}, \tag{7}$$

where W_{ij} is the number of wins of team i against j, and G_{ij} is the number of game they played against each other in a given season. If seasons were infinitely long, then $z_{ij} \to p_{ij}/(1-p_{ij})$, and hence

$$z_{ik}z_{kj} = z_{ij}. \tag{8}$$

To test the detailed balance relation equation (8), we plot $\langle \ln(z_{ik}z_{kj})\rangle$ as a function of $\langle \ln(z_{ij})\rangle$ from game data, averaged over all team triplets (i,j,k) and all seasons in a given period (Fig. 4). We discard events for which $W_{ij} = G_{ij}$ or $W_{ij} = 0$ (team i won or lost all games against team j). Our simulations of the BT model over 10^4 realizations of the 1901–60 and 1961–2005 periods with the same G_{ij} as in actual baseball seasons and with the optimal values of x_{min} for each period are in excellent agreement with the game data. Although $z_{ik}z_{kj}$ in the figure has a sublinear dependence of z_{ij} (slope much less than 1 in Fig. 4), the slope progressively increases and ultimately approaches the expected linear relation between $z_{ik}z_{kj}$ and z_{ij} as the season length is increased (Fig. 5). We implement an increased season length by multiplying all the G_{ij} by the same factor M. Notice also that $\langle \ln(z_{ik}z_{kj})\rangle$ versus $\langle \ln(z_{ij})\rangle$ for the 1901–60 period has a larger slope than for 1961–2005 because the G_{ij}'s are larger in the former period ($G_{ij} = 22$) than in the latter (G_{ij} in the range 5–19).

This study of game outcomes among triplets of teams provides a detailed and non-trivial validation for the BT form equation (2) for the winning probability. As a byproduct, we learn that cyclic game outcomes, in which team A beats B, B beats C, and C beats A, are unlikely to occur.

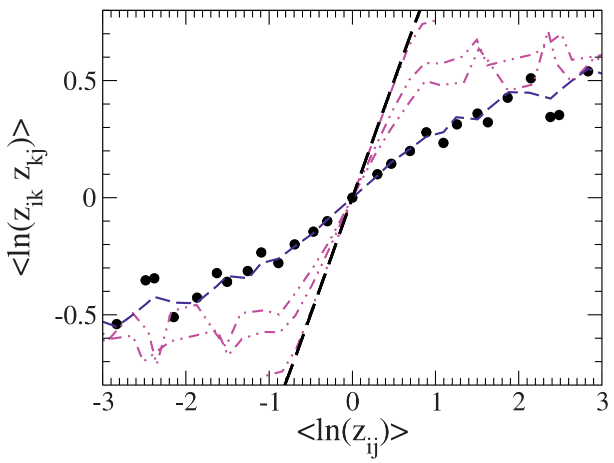

Fig. 5. Dependence of $\langle \ln(z_{ik} z_{kj}) \rangle$ vs. $\langle \ln(z_{ij}) \rangle$ on season length for the 1961–2005 period. All G_{ij}'s are multiplied by $M = 5, 10, 100$ (steepening dot-dashed lines). The thick dashed line corresponds to $M = 10^4$ and is indistinguishable from a linear dependence with unit slope.

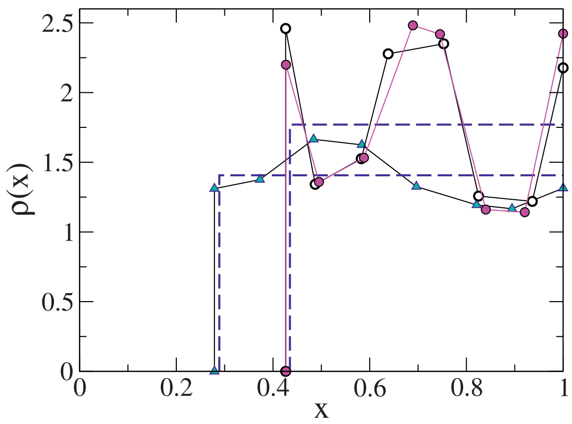

Fig. 6. Optimized strength distributions $\rho(x)$ for 1901–60 (triangles) and 1961–2005 (circles), together with the optimal uniform distributions (dashed). For 1961–2005, we also show the final distributions starting from y_i's equally spaced between $y_1 = 0.1$ and $y_8 = 1$ with the distribution ρ: (a) uniform on $[0.1, 1]$ (open circles), and (b) a symmetric V-shape on $[0.1, 1]$ (full circles).

2.3 Distribution of team strengths

Thus far, we have used a uniform distribution of team strengths to derive the average win fraction for the BT model. We now determine the most likely strength distribution by searching for the distribution that gives the best fit to the game data for $W(r)$ by minimizing the deviation Δ between the data and the simulated form of $W(r)$. Here the deviation Δ is defined as

$$\Delta^2 = \frac{\sum_r [W(r) - W(r;\rho)]^2}{\sum_r W(r)^2}, \tag{9}$$

where $W(r;\rho)$ is the winning fraction in simulations of the BT model for a trial distribution $\rho(x)$ in which the actual game frequencies G_{ij} were used in the simulation, and $W(r)$ is the game data for the winning fraction.

We assume that the two periods 1901–60 and 1961–2005 are long enough for $W(r)$ to converge to its average value. We parameterize the trial strength distribution as a piecewise linear function of n points, $\{\rho(y_i)\}$, with $y_i \in [0,1]$ and $y_n \equiv 1$. We then perform Monte Carlo (MC) simulations, in which we update the y_i and $\rho_i = \rho(y_i)$ by small amounts in each step to reduce Δ. Specifically, at each MC step, we select one value of $i = 1, ..., n$, and

- with probability 1/2 adjust y_i (except $y_n = 1$) by $\pm u\, \delta y/10$, where δy is the spacing between y_i and its nearest neighbor, and u is a uniform random number between 0 and 1;
- with probability 1/2, update $\rho(y_i)$ by $\pm u\, \rho(y_i)/10$.

If Δ decreases as a result of this update, then y_i or $\rho(y_i)$ is set to its new value; otherwise the change in the parameter value is rejected. We choose $n = 8$, which is large enough to obtain a distribution with significant features and for which typically 1000–2000 MC steps are sufficient for convergence. A larger n greatly increases the number of MC steps necessary to converge and also increases the risk of being trapped in a metastable state because the size of the phase space grows exponentially with n. To check that this algorithm does not get trapped in a metastable state, we started from several different initial states and found virtually identical final distributions (Fig. 6). The MC-optimized distribution for each period is remarkably close to uniform, as shown in this figure.

Although the optimal distributions are visually not uniform, the small difference in the relative errors, the closeness of y_1 and x_{min}, and the imperceptible difference in the r dependence of $W(r)$ for the uniform and optimized strength distributions suggests that a uniform team strength distribution on $[x_{min}, 1]$ describes the game data quite well.

For completeness, we also considered the conventionally-used log-normal distribution of team strengths [5,19]:

$$\rho(x) = \frac{1}{\sqrt{2\pi}\kappa x} \exp\left[-\frac{1}{2\kappa^2}\left(\ln\left(\frac{x}{\bar{x}}\right) + \frac{\kappa^2}{2}\right)^2\right]. \tag{10}$$

With the normalization convention of equation (10), the average team strength is simply $\bar{x}$, which can be set to any value due to the invariance of p_{ij} with respect to the transformation $x \to \lambda x$. Hence, the only relevant parameter is the width κ. Using the same MC optimization procedure described above, we find that a log-normal *ansatz* for the strength distribution with optimal parameter κ gives a visually inferior fit of the winning fraction in both periods compared to the uniform strength distribution, especially for r close to 1 (see Fig. 7). The relative error for the log-normal distribution is also a factor of 6 and 3 larger, respectively, than for the optimal distribution in the 1901–60 and 1961–2005 periods. However, we do reproduce the feature that the optimal log-normal distribution for 1961–2005 is narrower ($\kappa = 0.238$) than that

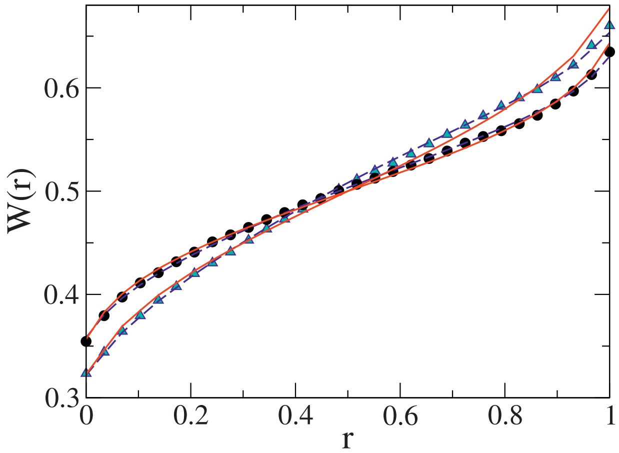

Fig. 7. Comparison of the winning fraction $W(r)$ extracted from the actual baseball data (symbols) to the model with a constant $\rho(x)$ (dashed lines), and with the optimal log-normal distribution $\rho(x)$ (full lines). The log-normal fit to the data is inferior to that provided by a uniform distribution, as illustrated in Fig. 1.

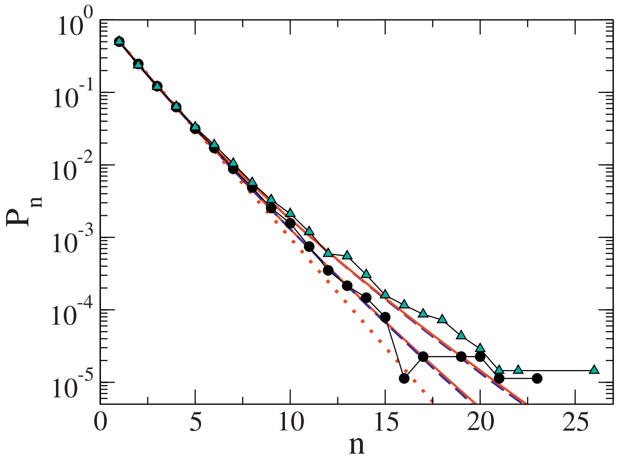

Fig. 8. Distribution of winning/losing streaks P_n versus n since 1901 on a semi-logarithmic scale for 1901–60 ($\triangle$) and 1961–2005 ($\bullet$). The dashed curves are the result of simulations with $x_{\rm min} = 0.278$ and $x_{min} = 0.435$ for the two respective periods. The smooth curves are streak data from randomized win/loss records, and the dotted curve is 2^{-n}.

for 1901–60 ($\kappa = 0.353$), indicating again that baseball is more competitive in the second period than in the first.

3 Winning and losing streak statistics

We now turn to the distribution of consecutive-game winning and losing streaks. Namely, what are the probabilities W_n and L_n to observe a string of n consecutive wins or n consecutive losses, respectively? Because of its emotional appeal, streakiness in a wide variety of sports continues to be vigorously researched and debated [15,20,21]. In this section, we argue that independent game outcomes that depend only on relative team strengths describes the streak data for the period 1961–2005 quite well. The agreement is not as good for the period 1901–60 and suggests that non-statistical effects may have played a role in the longest streaks.

Historically, the longest team winning streak (with ties allowed) in major-league baseball is 26 games, achieved by the 1916 New York Giants in the National League over a 152-game season [22]. The record for a pure winning streak since 1901 (no ties) is 21 games, set by the Chicago Cubs in 1935 in a 154-game season, while the American League record is a 20-game winning streak by the 2002 Oakland Athletics over the now-current 162-game season. Conversely, the longest losing streak since 1901 is 23, achieved by the 1961 Philadelphia Phillies in the National League [23], and the American League losing-streak record is 21 games, set by the Baltimore Orioles at the start the 1988 season. For completeness, the list of all winning and all losing streaks of ≥ 15 games is given in the Appendix.

Figure 8 shows the distribution of team winning and losing streaks in major-league baseball since 1901. Because these winning and losing streak distributions are virtually identical for $n \leq 15$, we consider $P_n = (W_n + L_n)/2$, the probability of a winning *or* a losing streak of length n (Fig. 8). It is revealing to separate the streak distributions for 1901–60 and 1961–2005. Their distinctness is again consistent with the hypothesis that baseball is becoming more competitive. In fact, exceptional streaks were much more likely between 1901–60 than after 1961. Of the 55 streaks of ≥ 15 games, 27 occurred between 1901–30, 13 between 1931–60, and 15 after 1960 [24].

The first point about the streak distributions is that they decay exponentially with n, for large n. This behavior is a simple consequence of the following bound: consider a baseball league that consists of teams with either strengths $x = 1$ or $x = x_{min} > 0$, and with games only between strong and weak teams. Then the distribution of winning streaks of the strong teams decays as $(1 + x_{min})^{-n}$; this represents an obvious upper bound for the streak distribution in a league where team strengths are uniformly distributed in $[x_{min}, 1]$.

We now apply the BT model to determine the form of the consecutive-game winning and losing streak distributions. Using equation (2) for the single-game outcome probability, the probability that a team of strength x has a streak of n consecutive wins is

$$P_n(x) = \prod_{j=1}^{n} \frac{x}{x + x_j} \; \frac{x_0}{x + x_{n+1}} \; \frac{x_{n+1}}{x + x_{n+1}}. \tag{11}$$

The product gives the probability for n consecutive wins against teams of strengths x_j, $j = 1, 2, \ldots, n$ (some factors possibly repeated), while the last two factors give the probability that the 0th and the (n+1)st games are losses to terminate the winning streak at n games. Assuming a uniform team strength distribution $\rho(x)$, and for the case where each team plays the same number of games with every opponent, we average equation (11) over all opponents and then over all teams.

The first average gives:

$$\langle P_n(x)\rangle_{\{x_j\}} = x^n \left\langle \frac{1}{x+y} \right\rangle^n \left\langle \frac{y}{x+y} \right\rangle, \qquad (12)$$

with

$$\left\langle \frac{1}{x+y} \right\rangle = \frac{1}{1-\epsilon} \ln\left(\frac{x+1}{x+\epsilon}\right),$$
$$\left\langle \frac{y}{x+y} \right\rangle = 1 - \frac{x}{1-\epsilon} \ln\left(\frac{x+1}{x+\epsilon}\right)$$

for a uniform distribution of team strengths in $[x_{min}, 1]$. Here we use the fact that each team strength is independent, so that the product in equation (11) factorizes. We now average over the uniform strength distribution, to find, for the team-averaged probability to have a streak of n consecutive wins,

$$\langle P_n\rangle = \frac{1}{1-x_{min}} \int_{x_{min}}^{1} f(x)\, e^{ng(x)}\, dx, \qquad (13)$$

where

$$f(x) = \left[1 - \frac{x}{1-x_{min}} \ln\left(\frac{x+1}{x+x_{min}}\right)\right]^2$$
$$g(x) = \ln x + \ln\left[\frac{1}{1-x_{min}} \ln\left(\frac{x+1}{x+x_{min}}\right)\right].$$

Since $g(x)$ monotonically increases with x within $[x_{min}, 1]$, the integral in equation (13) is dominated by the behavior near the maximum of $g(x)$ at $x = 1$ for large n. Performing the integral by parts [25], the leading behavior is

$$\langle P_n\rangle \sim e^{ng(1)}, \qquad (14)$$

with

$$g(1) = -\ln(1-x_{min}) + \ln\ln\left(\frac{2}{1+x_{min}}\right).$$

As expected, $\langle P_n\rangle$ decays exponentially with n, but with a decay rate that decreases as teams become more heterogeneous (decreasing x_{min}). In the limit of equal-strength teams, the most rapid decay of the streak probability arises, $P_n = 2^{-n}$, while the widest disparity in team strengths, $x_{min} = 0$, leads to the slowest possible decay $P_n \sim (\ln 2)^n \approx (0.693)^n$.

We simulated the streak distribution P_n using the same methodology as that for the win/loss records; related simulations of streak statistics are given in references [19,21]. Taking $x_{min} = 0.435$ for 1961–2005 – the same value as those used in simulations of the win/loss records – we find a good match to the streak data for this period. The apparent systematic discrepancy between data and theory for $n \geq 17$ is illusory because streaks do not exist for every value of n. Moreover, the number of streaks of length $n \geq 17$ is only eight, so that fluctuations are quite important.

For the 1901–60 period, if we use $x_{min} = 0.278$, the data for P_n is in excellent agreement with theory for $n < 17$. However, for n in the range 17–22, the data is a roughly factor of 2 greater than that given by the analytical solution equation (14) or by simulations of the BT model. Thus the tail of the streak distribution for this early period appears to disagree with a purely statistical model of streaks. Again, the number of events for $n \geq 17$ is 5 or less, compared to a total number of ~ 70000 winning and losing streaks during this period. Hence one cannot exclude the possibility that the observed discrepancy for $n \geq 17$ is simply due to lack of statistics.

Finally, we test for the possible role of self-reinforcement on winning and losing streaks. To this end, we take each of the 2166 season-by-season win/loss histories for each team and randomize them 10^5 times. For each such realization of a randomized history, we compute the streak distribution and superpose the results for all randomized histories. The large amount of data gives streak distributions with negligible fluctuations up to $n = 30$ and which extend to $n = 44$ and 41 for the two successive periods. More strikingly, these streak distributions based on randomized win/loss records are virtually identical to the simulated streak data as well as to the numerical integration of equation (13), as shown in Figure 8.

4 Summary

To conclude, the Bradley-Terry (BT) competition model, in which the outcome of any game depends only on the relative strengths of the two competing teams, quantitatively accounts for the average win/loss records of Major-League baseball teams. The distribution of team strengths that gives the best match to these win/loss records was found to be quite close to uniform over a range $[x_{min}, 1]$, with $x_{min} \approx 0.28$ for the early modern era of 1901–1960 and $x_{min} \approx 0.44$ for the expansion era of 1961–2005. This same BT model also reproduces the season-to-season fluctuations of the win/loss records. An important consequence of the BT model is the existence of a non trivial detailed-balance relation which we verified with satisfying accuracy. We consider this verification as a quite stringent test of the theory.

The same BT model was also used to account for the distribution of team consecutive-game winning and losing streaks. We found excellent agreement between the prediction of the BT model and the streak data for $n < 17$ for both the 1901–60 and 1961–2005 periods. However, the tail of the streak distribution for the 1901–60 period with $n \geq 17$ is less accurately described by the BT theory and it is an open question about the mechanisms for the discrepancy, although it could well originate from lack of statistics. We also provided evidence that self-reinforcement plays little role in streaks, as randomizations of the actual win/loss records produces streak distributions that are indistinguishable from the streak data except in for the $n \geq 17$ tail during the 1901–60 period.

We also showed that the optimal team strength distribution is narrower for the period 1961–2005 compared to 1901–60. This narrowing shows that baseball competition is becoming keener so that outliers in team performance over an entire season – as quantified by win/loss records and lengths of winning and losing streaks – are less likely to occur.

We close by emphasizing the parsimonious nature of our modeling. The only assumed features are the Bradley-Terry form equation (2) for the outcome of a single game, and the uniform distribution of the winning probabilities, controlled by the single free parameter x_{min}. All other model features can then be inferred from the data. While we have ignored many aspects of baseball that ought to play some role – the strength of a team changing during a season due to major trades of players and/or injuries, home-field advantage, etc. – the agreement between the win fraction data and the streak data with the predictions of the Bradley-Terry model are extremely good. It will be worthwhile to apply the approaches of this paper to other major sports to learn about possible universalities and idiosyncracies in the statistical features of game outcomes.

SR thanks Guoan Hu for data collection assistance, Jim Albert for literature advice, and financial support from NSF grant DMR0535503 and Université Paul Sabatier.

Appendix: Team winning and losing streaks

Table 1. Winning streaks of $n \geq 15$ games since 1901.

n	year	team
26	1916	New York Giants (1 tie)
21	1935	Chicago Cubs
20	2002	Oakland Athletics
19	1906	Chicago White Sox (1 tie)
19	1947	New York Yankees
18	1904	New York Giants
18	1953	New York Yankees
17	1907	New York Giants
17	1912	Washington Senators
17	1916	New York Giants
17	1931	Philadelphia Athletics
16	1909	Pittsburgh Pirates
16	1912	New York Giants
16	1926	New York Yankees
16	1951	New York Giants
16	1977	Kansas City Royals
15	1903	Pittsburgh Pirates
15	1906	New York Highlanders
15	1913	Philadelphia Athletics
15	1924	Brooklyn Dodgers
15	1936	Chicago Cubs
15	1936	New York Giants
15	1946	Boston Red Sox
15	1960	New York Yankees
15	1991	Minnesota Twins
15	2000	Atlanta Braves
15	2001	Seattle Mariners

Table 2. Losing streaks of $n \geq 15$ games since 1901.

n	year	team
23	1961	Philadelphia Phillies
21	1988	Baltimore Orioles
20	1906	Boston Americans
20	1906	Philadelphia As
20	1916	Philadelphia As
20	1969	Montreal Expos (first year)
19	1906	Boston Beaneaters
19	1914	Cincinnati Reds
19	1975	Detroit Tigers
19	2005	Kansas City Royals
18	1920	Philadelphia As
18	1948	Washington Senators
18	1959	Washington Senators
17	1926	Boston Red Sox
17	1962	NY Mets (first year)
17	1977	Atlanta Braves
16	1911	Boston Braves
16	1907	Boston Doves
16	1907	Boston Americans (2 ties)
16	1944	Brooklyn Dodgers (1 made-up game)
15	1909	St. Louis Browns
15	1911	Boston Rustlers
15	1927	Boston Braves
15	1927	Boston Red Sox
15	1935	Boston Braves
15	1937	Philadelphia As
15	2002	Tampa Bay
15	1972	Texas Rangers (first year)

References

1. See e.g., W. Weidlich, *Sociodynamics; A Systematic Approach to Mathematical Modelling in Social Sciences* (Harwood Academic Publishers, 2000); *Biological Evolution and Statistical Physics*, edited by M. Lässig, A. Valleriani, (Springer, Berlin, 2002); M. Newman, A.-L. Barabási, D.J. Watts, *The structure and dynamics of networks* (Princeton University Press, 2006)
2. J.-P. Bouchaud, M. Potters, *Theory of financial risk and derivative pricing: from statistical physics to risk management* (Cambridge University Press, 2003)
3. J. Krug, C. Karl, Physica A **318**, 137 (2003); K. Jain J. Krug, J. Stat. Mech., P04008 (2005)
4. E. Bittner, A. Nussbaumer, W. Janke, M. Weigel, Eurpphys. Lett. **78**, 58002 (2007); E. Bittner, A. Nussbaumer, W. Janke, M. Weigel, Nature **441**, 793 (2006)
5. J. Albert, J. Bennett, *Curve Ball: Baseball, Statistics, and the Role of Chance in the Game* (Springer New York, 2001)
6. J. Park, M.E.J. Newman, J. Stat. Mech. P10014 (2005)
7. E. Ben-Naim, S. Redner, F. Vazquez, Europhys. Lett. **77**, 30005 (2007)
8. E. Ben-Naim, N.W. Hengartner, Phys. Rev. E **76**, 026106 (2007)
9. C. Sire, J. Stat. Mech., P08013 (2007)
10. The data presented here were obtained from `www.shrpsports.com`
11. However, since 1997 a small amount of *interleague* play during the regular season has been introduced

12. E. Zermelo, Math. Zeit. **29**, 435 (1929)
13. R.A. Bradley, M.E. Terry, Biometrika **39**, 324 (1952)
14. See, e.g., http://en.wikipedia.org/wiki/List_of_MLB_individual_streaks
15. R.C. Vergin, J. of Sport Behavior **23**, (2000)
16. E. Ben-Naim, F. Vazquez, S. Redner, Journal of Quantitative Analysis in Sports **2**, 1 (2006)
17. S.J. Gould, *Full House: The Spread of Excellence from Plato to Darwin* (Three Rivers Press, New York, 1996)
18. In contrast, in Ref. [16], the winning probability was taken to be independent of the relative strengths of the two teams; the stronger team won with a fixed probability p and the weaker won with probability $1 - p$
19. B. James, J. Albert, H.S. Stern, Chance **6**, 17 (1993)
20. T. Gilovich, R. Vallone, A. Tversky, Cognitive Psychology **17**, 295 (1985)
21. J. Albert, Chance **17**, 37 (2004)
22. http://answers.yahoo.com/question/index?qid=1006053108634. This record is slightly tainted because of a tie during this streak, and ties are no longer allowed to occur; every game that is tied at the end of the regulation 9 innings must continue until one team wins
23. http://en.wikipedia.org/wiki/List_of_worst_MLB_season_records
24. Moreover, three of the post-1960 ≥ 15 game losing streaks occurred during the initial year of necessarily weak expansion teams because they were stocked with the weakest players from established teams (1962 NY Mets, 1969 Montreal Expos, 1972 Texas Rangers)
25. C.M. Bender, S.A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978) Sect. 6.3.