FREE Article

Scaling laws and forecasting in athletic world records.

Magazines

Journal of Sports Sciences; 7/1/2001; SAVAGLIO, SANDRA

In this study, we analysed running world **records** and found that the mean speed of the race, [Mu], as a function of the record time, [Tau], can be described asymptotically by two well-defined scaling laws of the form u ~ [[Tau].sup.-[Beta]]. There is a break in the scaling laws (~1000 m) between the shorter and the longer races at a characteristic time of around 150-170 s, after which a new scaling regime emerges. This is the first occasion that this characteristic time has been clearly found in physical terms; we interpreted it as the transition time between the anaerobic and the aerobic energy expenditure of athletes. This phenomenon is independent of the athletes' sex and is also found in swimming races with similar values of the characteristic time. We also investigated the forecasting of world **records** using historical data. Using an approach based on the identification of non-Poissonian events for a sequence of temporal point processes, we found that the sequence of improvements in all athletic **records** from 1900 to the present day cannot be considered as a sequence of completely random events.

Keywords: athletic world **records,** forecasting, scaling laws.

Introduction

The study of the potentiality of the human body offered by the monitoring of sport performance over 100 years is an indirect and unexplored way of investigating mathematically how society has evolved and will evolve in the third millennium. In particular, world **records** in athletics are very useful, both from a physical (Katz and Katz, 1999) and physiological (di Prampero, 1981; Mognoni et al., 1982) point of view. World **records** are measured in standard external conditions and represent the most reliable and up-to-date measure of human performance.

The interest of physicists and physiologists in developing mathematical models for running dates back to 1900 (Kennelly, 1906; Henry, 1955; Katz and Katz, 1999). Scientists have tried to show that a scaling relation exists between distance, d, and time of running, [Tau]. This relation, represented by a power law of the kind

(1) [Tau] = [cd.sup.n]

(where c is a constant and n is a characteristic scaling exponent), has been found by looking at data on **records** (see, for example, Katz and Katz, 1999). The range of distances covered by relation (1) appears to be very wide, from sprints of 100 m to the 10,000 m, with a scaling exponent that depends on the epoch. For example, n = 1.141 for the **records** of 1925, while n = 1.123 for the **records** of 1995 (Katz and Katz, 1999). Figure 1 plots the logarithm of times, log [Tau], as a function of the logarithm of distances, log d, for both running and swimming and for men and women (here and throughout the paper we use decimal logarithms). In this way, by applying the logarithm operation to both sides of equation (1), we get log [Tau] = log ([cd.sup.n]) = log c + n log d and relation (1) is transformed into a linear relationship with a slope given by the angular coefficient n. The scaling index is thus obtained by measuring the angular coefficient.

[GRAPH OMITTED]

An immediate consequence of relation (1) is the existence of a scaling invariance for human performances, for both running and swimming. This means that, if we change the race distance by a factor [Lambda] (d [right arrow] [Lambda]d), the time, which is a function of d, changes by a factor [[Lambda].sup.n]:

(2) [Tau]([Lambda]d) = c[[Lambda].sup.n][d.sup.n] = [[Lambda].sup.n][Tau](d)

The ratio between times at the two different distances is [Tau]([Lambda]d)/[Tau](d) = [[Lambda].sup.n] and depends only on the scaling index n. The presence of a scaling invariance indicates an invariance of the statistical properties of the system over all scales, implying that it is impossible to define a characteristic scale for the phenomenon. In physics, a statistical system at a critical point -- for example, at a phase transition -- displays this kind of behaviour with a power law distribution. In analogy with this, we call the behaviour of the human body during running and swimming a critical phenomenon.

Measured world **records,** however, show a deviation from equation (1). By taking a closer look at Fig. 1, we note that the points are not well distributed along a straight line and that a `knee' around d = 1000 m is present. To make this clearer, we use the mean speed at distance d, u = d/[Tau]. The distance d of the race is given a priori, thus the variable u better represents the phenomenon, because it is directly related to the average kinetic energy of the athlete ([varies] [u.sup.2]) or, in other words, to the energy expenditure during the race. Since the slope of the scaling law (1) is close to n [approximately equals] 1 (see Fig. 1), if we consider the average speed, we expect that u = d/ [Tau] ~ [[Tau].sup.-[Beta]] with a slope [Beta] = 1 - 1/n. However, that n is only slightly greater than 1 leads to the conclusion that even small deviations from the scaling law (1) are amplified if the relation between u and [Tau] is considered.

Scaling laws in running and swimming

To investigate scaling laws in world **records,** we used running and swimming data for both men's and women's events. The results are summarized in Fig. 2. Unlike (1), there are two distinct asymptotic regimes, abruptly separated by a characteristic scaling time [Tau]* in the range 2.2-2.8 min. For each regime, we have assumed a scaling law of the form

(3) u = c[[Tau].sup.-[Beta]]

with two distinct values of the critical exponent [Beta]. To determine the relationship (3) in the two regimes, we performed a maximum likelihood analysis, assuming that the point distribution can be described by a broken power law and arbitrarily giving to the single points an error close to their fluctuations around a theoretical expectation.

[GRAPH OMITTED]

The first scaling law is seen for distances of 200-1000 m and 50-200 m for running and swimming, respectively. The second scaling law is seen for distances of 1500-42,195 m (marathon) and 400-1500 m for running and swimming, respectively. Figure 2 also shows the `flying' speed in the men's 100 m track race (average speed in the last 50 m; data taken from Ward-Smith and Radford, 2000), which confirms expectations. Short races are separated in a distinct way from long races by a characteristic time [Tau]*, which means that running and swimming are characterized by two critical phenomena. In other words, there are two distinct types of running and swimming that are sharply separated. This is well known to individuals directly involved in sport, such as athletes and coaches, but this is the first occasion that this difference has been shown convincingly by the data **(Savaglio** and Carbone, 2000). What is noteworthy is that there are two distinct scaling laws for both men's and women's races and for running and swimming, and the scaling time is very similar.

The appearance in all cases of two universal scaling laws must be interpreted within the physiological models of muscular exercise. The transition between both regimes occurs when the athletes' anaerobic energy expenditure becomes too great; it is then substituted by aerobic energy expenditure. By looking at world **records** from different epochs, we note two distinct scaling laws that are independent of years and appear consistently throughout the history of athletics. Table 1 shows the scaling exponents for the anaerobic and aerobic regimes ([[Beta].sub.an] and [[Beta].sub.ae]) of running in men's and women's events in different epochs at intervals of 10 years (before 1951, the year of the first official record for the 200 m, **records** for this distance have been substituted by the best European or US times; the effect of this on the values of [[Beta].sub.an] obtained using data before 1951 was negligible).

To interpret our results, we invented the `super-athlete', an ideal athlete who is able to run or swim all races at world-record times. It is clear that the smaller the scaling exponent, the better the performance of the super-athlete, in the sense that the dissipated power will decrease more slowly during longer races. At the limit of human capabilities, the super-athlete can perform all races at the same very high speed. This would correspond to stretching the power laws of Fig. 2 and reducing the slopes to zero. The slopes of the two scaling laws can be used to determine how the performance evolved and how close to the limit it is when comparing athletes of different age and sex, from different countries, and so on. In running, men and women show the same efficiency with the same slopes for both the anaerobic and aerobic regimes. The common belief that women are physically better than men when running long-distance races is not confirmed by these data. The slope for men is flatter than, although consistent with, that of women in the anaerobic regime; this could be due to the different budgets of anaerobic energy, which is a fixed quantity and favours men. The opposite happens in swimming, where small differences are most probably due to differences in the buoyancy which favours female swimmers in short races. The slopes are significantly steeper for running than for swimming, perhaps because swimming demands a small aerobic contribution for short races, which helps for example with buoyancy, as the human body is lighter in the water when the lungs are filled with air.

Even if top human performances follow a simple double power law (3), is it possible to construct a physiological model that describes this behaviour? This kind of modelling was introduced, for example, by di Prampero (1981, 1984). The model is based on forecasting the maximum speed in running races (from 400 to 10,000 m), which has been reported to be proportional to the maximum metabolic power of the athlete for the duration of the race and the energetic expenditure for the race at a given constant speed. The model (di Prampero, 1984) provides a set of values for maximum speed, which are shown in Fig. 3 together with the present world **records.** As can be seen (di Prampero, 1984), there is good agreement in terms of the absolute value of the record, with an error close to 4%. However, there is no agreement between the functional shape of the model and that shown by the data. The model does not predict a double power law, rather an exponential decrease in average speed.

[GRAPH OMITTED]

Periodicity and forecasting in the time sequence of world **records**

Periodicity signal through wavelet analysis

It is interesting to look at all world **records** together from 1900 to the present day. Let $X_n(t_i)$ be the n-th athlete's record (a time for running, a length for jumping, etc., for all athletics disciplines) obtained at a discrete epoch t_i, the time at which the (i - 1)-th record has been improved. The relative improvement (Δ_n) of a given record is:

(4) $\Delta_n(t_i) = |X_n(t_i) - X_n(t_{i-1})| / X_n(t_{i-1})$

A plot of this quantity as a function of epoch shows that improvements in world **records** in men's athletics (that for females is much less well documented) tend to cluster; that is, they are intermittently localized during short periods (see Fig. 4). The fundamental period seems to be about 1 year, evidently related to the periodicity of training and official events of athletics.

[GRAPH OMITTED]

Using wavelets transforms (cf. Meyer, 1992), we can perform a quantitative analysis of the existence of characteristic periods. Such an analysis, unlike Fourier transforms, provides information on the time evolution of characteristic periods (or frequencies) in a given signal. Let us define the function $\Psi_n(t_i)$ as the number of **records** improved each month in athletics. The result of the wavelets transform on this function is a set of wavelet coefficients, $W_n(t_i,f)$, where t_i is the discrete epoch (with time-step equal to 1 month), which characterizes the time evolution of the signal at frequency f. The information contained in the square modulus $|W_n(t_i,f)|^2$ is proportional to the energy contained in the signal at frequency f at time t_i. As a consequence, a map of the function $|W_n(t_i,f)|^2$, obtained through $\Psi_n(t_i)$, provides information on characteristic periods -- proportional to the inverse of f -- relative to the history of athletics. The lower panel of Fig. 5 shows the time evolution of $\Psi_n(t_i)$, while the upper panel shows the values of the square modulus of the wavelet coefficients in the (t_i,f) plane. Darker regions in Fig. 5 correspond to larger values of $|W_n(t_i,f)|^2$. As can be seen, the time evolution at a frequency of 1 year^{-1} shows an evident periodicity. Then, 1 year is the fundamental period of the system. Moreover, large values of $|W_n(t_i,f)|^2$ are visible also for a period of about 4 years, which corresponds to the time between Olympics (for $f \approx 1/4 = 0.25$ year^{-1}). On larger scales (frequencies less than 0.1 year^{-1}), we observe `positive' or `negative' periods for sport. For example, one negative period was the Depression of the 1940s, which was strongly influenced by the Second World War; positive periods for sport were the 1930s and 1960s.

[GRAPH OMITTED]

Forecasting world **records**

Is the improvement of a record in a given athletic contest a completely stochastic event, related to occasional performances of athletes in top races, or is forecasting possible? Remember that there are various stages to forecasting. For example, `strong' forecasting is associated with whether a particular event will happen. In physics, this is related to determinism: given the initial conditions of a deterministic system (whose evolution is described through deterministic equations), we are able to describe its evolution at all times. Of course, this is possible only from a theoretical point of view. On the other hand, the time behaviour of chaotic systems, even if described by deterministic equations, is unpredictable owing to its sensitive dependence on initial conditions. More relaxed types of forecasting are, for example, related to the probability that, at a given time At, an event will occur. This kind of forecasting is related to unpredictable or stochastic events, such as earthquakes and solar flares.

To approach this mathematically, we look at the statistics for the times separating isolated events. In particular, we look for deviations from a local Poisson distribution of these times. Poisson statistics indicate that events are not correlated; that is, there is no relation between an event at time t_i and that at time t_{i+1}. This approach should be useful for all kinds of `events', such as the radioactivity in nuclei, flares in the solar corona and the distribution of large masses in the universe. This approach is useful in our case also, where a new record in a given contest is considered an `event'.

As a first test, we define the time between two consecutive events as the time interval between a record and the next one for a given athletic contest. An exponential statistic of these times is a characteristic of the random Poisson process (Boffetta et al., 1999). However, as we have seen, the local density of races has a strong effect on these times. The distribution of top races, when **records** usually occur, is not absolutely uniform, owing to the natural evolution of society as a whole. Let us use a different approach. If [t.sub.i] represents the date when a record for the n-th contest is improved, let [T.sub.i](n) be the time in days between the dates of two consecutive **records.** Since for each time [t.sub.i] two neighbouring **records** exist, two values of [T.sub.i](n) will exist, [MATHEMATICAL EXPRESSION NOT REPRODUCIBLE IN ASCII]. and [MATHEMATICAL EXPRESSION NOT REPRODUCIBLE IN ASCII]. Let [S.sub.i](n) be the conditioned time interval between the next two neighbouring events: [MATHEMATICAL EXPRESSION NOT REPRODUCIBLE IN ASCII] and [MATHEMATICAL EXPRESSION NOT REPRODUCIBLE IN ASCII], or [MATHEMATICAL EXPRESSION NOT REPRODUCIBLE IN ASCII] and [MATHEMATICAL EXPRESSION NOT REPRODUCIBLE IN ASCII] if [MATHEMATICAL EXPRESSION NOT REPRODUCIBLE IN ASCII]. According to a supposed Poisson statistic, both times [T.sub.i](n) and [S.sub.i](n) are independently distributed with a given local density. After some algebraic calculations, it can be shown that the function [H.sub.i] defined as

(5) [H.sub.i](n) = 2 [T.sub.i](n)/ 2[T.sub.i](n) + [S.sub.i](n)

has a cumulative distribution independent of the local density, and it is uniformly distributed in the interval [0,1] with an average value of 0.5. As a consequence, calculation of the cumulative distribution of this quantity provides information on the randomness of the record improvements. In particular, if local clusters of events exist, the average will be greater than 0.5; if, in contrast, a more regular sequence of events exists, the average will be less than 0.5.

For statistical reasons (relatively few improvements of **records** are seen for each athletics event between 1900 and the present day), we are forced to use a sample of values with **records** relative to all athletic races. We calculate the values of [H.sub.i](n) for each athletic contest, from which we get a set of values of H. The cumulative distribution we calculate is P(H [is less than or equal to] h), which represents the probability (P) of obtaining a value of H less or equal to an assigned h. If the time sequence of **records** is completely random (Poisson hypothesis for the events), then P(H [is less than or equal to] h) = h. This probability can easily be calculated by imposing an increasing value of h in the range 0 [is less than] h [is less than or equal to] 1 and by counting how many values of H fall in the range 0 [is less than] H [is less than] h. If this number is N(h) and [N.sub.Tot] is the total number of values of H, then P(H [is less than or equal to] h) = N(h)/[N.sub.Tot]. Figure 6 shows the line P(H [is less than or equal to] h) = h, together with the curve obtained using the data. There is a substantial difference (at a significance level [is greater than] 99.99% with the Kolmogorov-Smirnov test), so that we can rule out the presence of a Poisson distribution for the events. The average of H is 0.49. This implies that the sequence of **records** in the history of athletics is not distributed in random fashion, rather a regularity exists. From a statistical point of view, this is indicative of a possible correlation -- evidently not known a priori! -- between a given event and the next one. Perhaps this is related to cyclical training or to a regularity in the discovery of elite athletes in consecutive human generations.

[GRAPH OMITTED]

Conclusions

We analysed the time sequence of world **records** from 1900 to the present day as a measure of the evolution of human performance. Our results can be summarized as follows:

* We have shown for the first time, using real data, the existence of two scaling laws in both running and swimming events. This means that there are two critical phenomena, which are different for short and long races, described physiologically as the anaerobic and aerobic processes. This is independent of age and sex and has been constant throughout the whole history of athletics.

* We found periodicity in the time sequence of **records,** obviously related to cyclical training or cyclical elite competitions, as well as to social phenomena on large scales.

* Our results suggest that, unlike commonly believed (di Prampero, 1984), **records** in athletics are not completely random events and, at least from a purely mathematical point of view, forecasting is possible.

Acknowledgements

We thank M. Bruni of the `Scuola dello Sport' (CONI, Italy), O. Florese of CONI (CS, Italy), C. Perrotta, P. Padovani and all our colleagues for their support and interest in our work. Thanks also to S. Katz for sending us a copy of the paper referenced as Katz and Katz (1999). Finally, we are grateful to the captain of the European Southern Observatory swimming team, H. Zodet, for inspiring our work on world **records.**

References

Boffetta, G., Carbone, V., Giuliani, P., Veltri, P. and Vulpiani, A. (1999). Power laws in solar flares: Self-organized criticality or turbulence? Physical Review Letters, 83, 4662-4666.

Di Prampero, P.E. (1981). Energetics of muscular exercise. Review of Physiological and Biochemical Pharmacology, 89, 143-222.

Di Prampero, P.E. (1984). I record del mondo di corsa piana: Costi energetici della corsa e previsione dei record (in Italian). SdS -- Rivista di cultura sportiva, 5, 3-7.

Henry, F.M. (1955). Prediction of world **records** in running sixty yards to twenty-six miles. Research Quarterly, 26, 147-158.

Katz, J.S. and Katz, L. (1999). Power laws and athletic performance. Journal of Sports Sciences, 17, 467-476.

Kennelly, A.E. (1906). An approximate law of fatigue in the speeds of racing animals. Proceedings of the American Academy of Arts and Sciences, 42, 273-331.

Meyer, Y. (1992). Wavelets and Operators. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press.

Mognoni, P., Lafortuna, C., Russo, G. and Minetti, A. (1982). An analysis of world **records** in three types of locomotion. European Journal of Applied Physiology, 49, 287-299.

Savaglio, S. and Carbone, V. (2000). Human performances: Scaling in athletic world **records.** Nature, 404, 244.

Ward-Smith, A.J. and Radford, P. (2000). Investigation of the kinetics of anaerobic metabolism by analysis of the performance of elite sprinters. Journal of Biomechanics, 33, 997-1006.

VINCENZO CARBONE(1)(*) and **SANDRA SAVAGLIO(2)**

(1) Dipartimento di Fisica and Istituto Nazionale di Fisica per la Materia, Unita di Cosenza, Universita della Calabria, 87036 Rende (CS), Italy and (2)Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

(*) Address all correspondence to Vincenzo Carbone, Dipartimento di Fisica, Universita della Calabria, 87036 Rende (CS), Italy. e-mail: carbone@fis.unical.it

Accepted 13 February 2001