Mathematics and Sports

Distribution of performances and scoring in athletics

B. Grammaticos[1,*], J. Meloun[2], J. G. Purdy[3]

[1]Université Paris-Saclay & Université de Paris, CNRS/IN2P3, IJCLab, 91405 Orsay, France, [2]Barry University, 11300 NE 2nd Ave, Miami Shores, FL 33161, USA, [3]Florida Atlantic University, 777 Glades Rd, Boca Raton, FL 33431, USA

[*]Corresponding Author E-mail:
basigram@gmail.com

Abstract

We address the question of scoring the athletic performance and claim that scoring should be linked to the distribution of performances among a selected population. We start by reviewing four different scoring methods (Letessier, Harder, Purdy, and World Athletics) that have been proposed in the past. We proceed next to formulate the mathematical link between scoring and performance distributions. Based on the explicit expressions for the Purdy and World Athletics scoring tables, we derive the corresponding performance distributions, finding that they involve the well-known Gompertz and Weibull distributions. Starting from a distribution assumption, we propose a new version of the Purdy scoring method, which could be the basis for new scoring tables for combined events, as well as for popular endurance events. Finally, we present a bold claim that the distribution of performances of a very large human population can be represented by a skew-logistic distribution.

Keywords: scoring, performances distribution, athletics, combined events

1 Introduction

The spirit of competition is deeply ingrained in human nature. What started as a survival trait became ritualized in sport, where the natural proclivity for play of most predating species became, in the case of humans, codified and institutionalized. The origins of sports [Crowther 2007] can be traced back to the Neolithic and, supposedly, even to the upper Palaeolithic era. However it is in Ancient Greece [Golden 2004] that sports found their fullest expression and became a permanent fixture of the society. Competitions in the Ancient Greek tradition [Miller 2004] were governed by the quest for excellence. The most important thing was victory [Leonard 2016]. The glory went to the winner of the contest; placing second was considered as being the first of the defeated. In that spirit, the Ancient Greek sport did not have an exigency of precise measurements. This is indeed reflected in the fact that very few data on the performances of the ancient Greek athletes have been preserved. However, with the evolution of the society and the progress of technology, sport became progressively more quantitative. Athletics was probably the first sport to profit from accurate time and distance measurements.

With the advent of quantitative assessment of the competition's results came the possibility of performance comparison even in the absence of direct confrontation. Although diachronic comparisons can be (and are being) criticized, maintaining a list of records or of top performances is a regular practice in today's sports [Hymans 2021]. The possibility of objective comparisons led naturally to the quest for the best athlete. While it is natural to assume that the champion of some discipline is better than a beginner, even in some other discipline, the situation is less clear when one tries to compare the champions of two different disciplines. The question could have been considered of pure academic interest were it not for the introduction of combined events.

The history of combined events is a long one. It goes back to the Ancient Greek Olympic Games and the introduction of the pentathlon [Zarnowski 2013]. The latter consisted of five events: discus throw, long jump, javelin throw, stadium race (192 m in Olympia) and wrestling. Again, the important thing in the Ancient Greek tradition was to designate the winner. Unfortunately, no writings explaining the precise classification method of the pentathlon exist and one can only formulate educated guesses [Grammaticos 2013] about it. Modern versions of combined events made their appearance in the second half of the 19th century. Two different methods were used in order to establish a classification. The first consisted in tallying the places of the competitors in each event, the victor being the athlete with the smallest number of points. Although simple in its conception, this method is both participant-dependent and unfair. Indeed the place of a given athlete in each event depends on the other participants. Moreover the same number of points will separate two competitors, based on their relative positions, independently of the magnitude of the difference of their performances. In a sense, this consists in replacing a fine measurement by a yes/no answer. The second method is more complicated, but when properly implemented, quite fair. It consists in attributing a number of points, which according to the current tradition span the range from 0 to 1000 (and beyond), to each performance. This led to the appearance, already at the end of the 19th century, of what are now known under the moniker of "scoring tables". Scoring tables and their theoretical foundation will be the object of the present paper. For the most part, in our presentation we shall focus on Athletics. However at the end of the paper we shall revisit the question of scoring the human performance in general, in the light of the developments presented in the rest of the paper. Before proceeding, we feel that a caveat is in order. Scoring the athletic performance is a delicate business. One can tell by the fact that it took close to a century for the international athletics federation [Formerly IAAF, currently World Athletics (WA)] to get it right.

There are two main difficulties one faces when attempting the development of scoring tables for athletics. The Letessier brothers, authors of the homonymous tables [Letessier 1961] used in France for close to 50 years, had dubbed them "horizontal" and "vertical". The horizontal task is to establish the equivalent performances between the various disciplines. For this, one needs a solid statistical basis of performances, ranging from the ones of the sport's elite to those of the beginners. The second task, the vertical, deals with attributing points to the various performances. Here, one needs a theory based on biomechanics and physiology, while keeping in mind the fact that the vertical task in never completely independent of the horizontal. In this work we shall deal with the second task, mainly developing a "vertical" scoring system i.e. devising a way to attribute points to performances of a single discipline, without addressing the question of the correspondence between different disciplines. Our approach will be one based on mathematical modelling.

The idea of using some physical quantity, other than the duly measured time, in the construction of the tables can be found in the work of J. Ohls who proposed a scoring system for track events based on the velocity. His work led to the adoption of what came to be known as the Finnish tables [Sarnowski 1989] in 1934. K. Ulbrich [Ulbrich 1950], who is at the origin of the 1962 IAAF scoring tables, formulated a strong critique concerning the use by Ohls of length for the scoring of field events. He insisted that even for these events the proper scoring ansatz should be linear in velocity. Unfortunately Ulbrich was wrong and this led to the introduction of inadequate scoring tables which plagued international athletics for more than two decades, despite the efforts [Purdy 1974] of one of the authors, (G.P.), to remedy this. The first, genuinely physics-based approach to scoring was to be found in the works of M. Woolf. In his article [Woolf 1968] Isomarks he stated epigrammatically "the tables should carry the germ of some physical truth". He went on to make a point that scoring should be related to the work required in order to produce the performance. This resonates with the approach [Grammaticos 2007] of one of the authors, (B.G.), who claimed that the points attributed to a performance should be related to free energy, i.e. the amount of energy that can be converted into work. A prerequisite for the construction of scoring tables is thus the accurate knowledge of the energetic cost [DiPrampero 1986] of the various disciplines. Put in this way, the approach just sketched is not limited to athletics, but can be applied to all sports where the performance depends mainly on physiological factors (like swimming or cycling).

In this paper, we shall address the question of the link between distribution of performances and scoring. We shall start by analyzing four different approaches to the construction of tables due to Letessier [Letessier 1961], Harder [Harder 2001], Purdy [Purdy 1974] (we refer to the tables proposed by one of us, G.P., as the "Purdy tables", in order to avoid awkward circumlocutions) and tables used by World Athletics. The first is, according to the author(s), based on an explicit assumption of an underlying performance distribution. In the case of Harder's approach, the link to a performance distribution was given by one of the authors (B.G.) [Grammaticos 2007]. The tables proposed by another of the authors, (G.P.), were constructed using a sample of thousands of performances, but they were not based on an explicit performance distribution. This will be remedied in what follows. The WA federation, currently maintains two sets of tables. The ones due to B. Spiriev [Spiriev 2017] (known as the Hungarian tables), are currently in use by WA for the scoring of performances of all events, not strictly confined to the combined ones. The second set is the one drawn up under the direction of V. Trkal [Trkal 2016] and is devoted to the scoring of combined events. However, from a mathematical point of view, their structure is similar and there is no need to study them separately. The relation of scoring to the performance distribution will be studied in this paper. Based on our findings, we shall propose a modification of the Purdy approach, which in some sense, would blend it with the one proposed by another of the authors (B.G.). We believe that the scoring method, thus obtained, is the most appropriate one for the scoring of athletic performances. Subsequent work will explain why this new approach, while better at solving the scoring table problem, can concurrently be applied to a broad class of athletic performance. This is in the same spirit as one of the authors (G.P) work [Purdy 1972] on the running curve) with the objective of helping people train while minimizing injury risks.

2 Four scoring approaches

In this section we shall present four different scoring methods. The Letessier approach is based on an explicit hypothesis of the distribution of performances and is thus a precursor of the ideas presented in this paper. The Purdy method is the one proposed by one of the authors, (G.P.), and will be the basis of further developments. The Harder method will also be presented together with the results of one of the authors, (B.G.), that make it possible to link the scoring proposed to a distribution of performances. Finally the tables currently used by World Athletics for the scoring of performances will also be included in our analysis.

The scoring tables are traditionally constructed so as to cover the range from 0 to 1000 points while providing scoring for performances beyond the ones obtaining 1000 points (and which can go beyond the current world record). The Letessier tables constitute an exception since their authors decided to limit the maximum of points to 500. The Letessier brothers argue that the precision necessary to cover a wider range is not guaranteed and thus a doubling of the range is to their eyes unnecessary. We are here in presence of an example of what was touted at the time as the ‘exception française’, something that has since largely disappeared. Before proceeding, it is interesting to discuss the question of the null score, i.e. the performance that obtains zero points in scoring.

2.1 The question of null score

The Letessier brothers do not offer any explanation as to how the zero point of their tables was fixed. They explain that adjusting the distribution of performances is a statistical task and it is clear that, being professors of physical education they had a sufficient sample of performances (albeit covering only a certain age group) at their disposal. However a close inspection of the values presented shows that while for short distances the null score velocity is of the order of 4-4.5 m/s, its value diminishes fast with the distance of the race and would correspond to leisurely walking velocities already for races of 1000 m. The situation is equally puzzling for the jumps. While the null score for the high jump is just 58 cm, where a step over the bar would suffice, the corresponding performance for the long jump is 1.78 m, which requires some (admittedly, very modest) effort.

Harder decided, somewhat arbitrarily, to attribute zero points to a performance that can be realized by 95 % of the population. Of course knowing the precise value of this for each discipline is a tall order but, if we knew the distribution function of performances for the discipline in question (and this is a big ‘if’) fixing the null score would be straightforward.

Purdy went one step further and, at least for track events and jumps, presented some physical arguments in support of his choice of null score. The argument goes as follows. For running events he assumed that a walking velocity is not running and can be continued almost indefinitely. The expression of the null score velocity is $v_0 = 2 - d/99950$ m/sec where d is the distance of the event. For vertical jumps the null height was set at that of the average center of gravity fixed at $L_0 = 80$ cm. For horizontal jumps one assumes that a large walking step of $L = 1$ m is not jumping (and thus for the triple jump the null score would be 3 m). There was no physical guide for the throwing events and so Purdy decided to fix the null score distance at 5 % of the one that corresponds to 1000 points. And it turned out that the null scores in the World Athletics tables are not very different form those of Purdy, at least as far as field events are concerned.

While Purdy’s approach is quite realistic, it can definitely be improved. Concerning the vertical jumps what should determine the null score is the height of the crotch and not that of the center of mass. Still, 80 cm is not off the mark, but, if we take into account the average height of men, 75 cm would be more realistic (and perhaps 70 cm for women). The estimate of 3 m for triple jump is not quite right since the jumping style is fixed and imposes a hop on the first jump. However a null-score mark between 2 and 3 m would be acceptable. For the throws, a suggestion in the spirit of the previous ones would be to decide that the null-score mark corresponds to an arm’s length for the heavy throws, around 75 cm (perhaps, slightly larger in the case of the hammer throw). For the javelin one should add the length of the javelin from tip to center of mass, roughly a meter, so the null-score mark would be around 1.80 m. However a problem does persist, and it has to do with hurdle races. The WA rules [World Athletics 2021] are clear on this point. The athletes must jump each hurdle, and they can be disqualified if they deliberately knock down any hurdle. So, in principle, in order to score some points the hurdler has to clear 1.06 m hurdles, for men’s 110 m race, and 0.84 cm for the women’s 100 m one. It is only in masters races, where for over M70+ the hurdles have a height of 76.2 cm and 68.6 cm for M80+ men and M60+women, that stepping over them without jumping becomes feasible. (Steeple hurdles do not pose a problem since one can step on them). Unfortunately given the current state of affairs there is no satisfactory answer for hurdle races.

In [Grammaticos 2007] one of the authors, (B.G.), discussing the problem of the null score, pointed out that a scoring table that aims at wide applicability (and not just reserved to competitors of a certain level) should attribute zero points to a zero-level performance i.e. zero velocity or zero length in the case of athletics. Moreover there exists a (small) percentage of the total population unable to realize any performance (usually for severe health reasons). Thus a null score, like the one proposed in the Purdy tables, based on physiological and/or biomechanical arguments is most appropriate when one develops scoring table targeted at sport practitioners. However, when the aim is to score the human performance as a whole, then the null score can be set at zero performance. Interestingly, as we shall see in Section 5, the new proposal for a scoring approach makes this point immaterial.

2.2 The Letessier tables

The Letessier tables were not proposed in relation to combined events but were introduced for educational purposes. When physical education became an essential component of the French educational system and students were to be attributed marks according to their performance in sports activities, various scoring scales were proposed. The first attempts were disastrous. In some events more than 50 % of the students were obtaining the maximal mark, while in some other less than 1 % managed to do so. It became thus clear that a serious overhaul of the scoring process was in order. Jean Letessier, a physical educator had already pointed out in 1951 the necessity of a systematic approach [Fortune 2008] to the question.

What came to be known as the ‘Letessier tables’ is the product of a close collaboration of Jean Letessier with his brother Pierre. According to the Letessier brothers the construction of the tables must be based on a statistical approach. However, as they correctly remark, a statistical comparison has meaning only if all the distributions of the performances follow the same law. They define two tasks, which they dub horizontal and vertical. The horizontal is to establish the equivalent performances between the various disciplines. The vertical one is to attribute points to the various performances. The authors of the tables are aware of the difficulties presented by these two tasks. While it is not so difficult to find equivalences between high level performances, when one gets to the bottom of the tables, things can get iffy.

The Letessier tables make the assumption that the distribution of the performances follows a specific distribution which can described in mathematical terms. In this sense they are the first to introduce a rigorous approach to the construction of scoring tables. However their choice of distribution was rather unfortunate, since they opted for a 'log-normal' distribution. Figure 1, below, shows how the performances of a large population ensemble would be distributed if they were following the log-normal distribution.

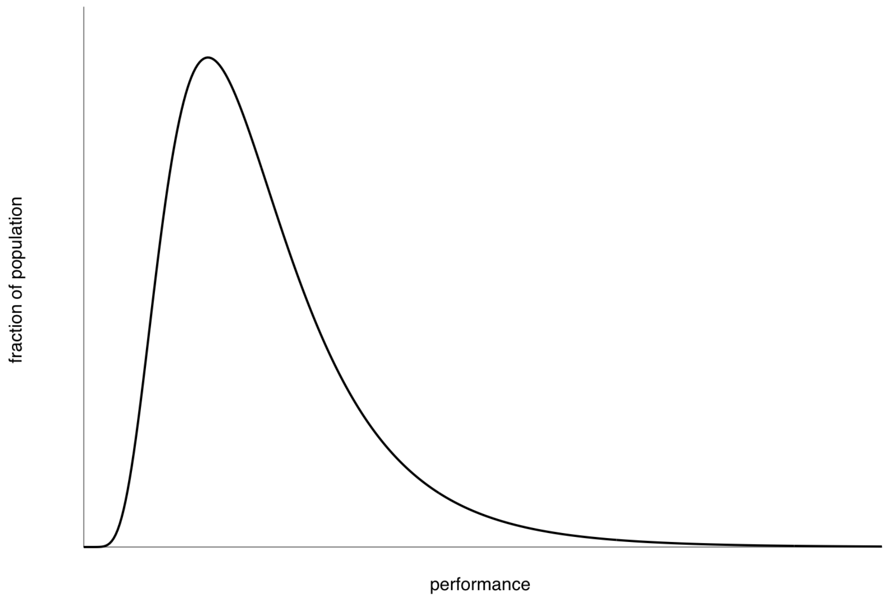

Figure 1. *The log-normal distribution (fraction of population vs. performance) used in the Letessier tables.*

The motivation of the Letessier brothers for the use of this particular distribution is that it appears in the description of natural phenomena, and in particular, in the case of animals, it is related to the distribution of the size, be it height or weight. In the case of athletic performance, their was probably also influenced by the fact that the distribution curve goes to zero for zero performance and in fact stays close to zero for a whole range of (small) performances. However when applied to athletic performances it leads to a regressive scoring. Figure 2 below shows an example of scoring for the long jump.

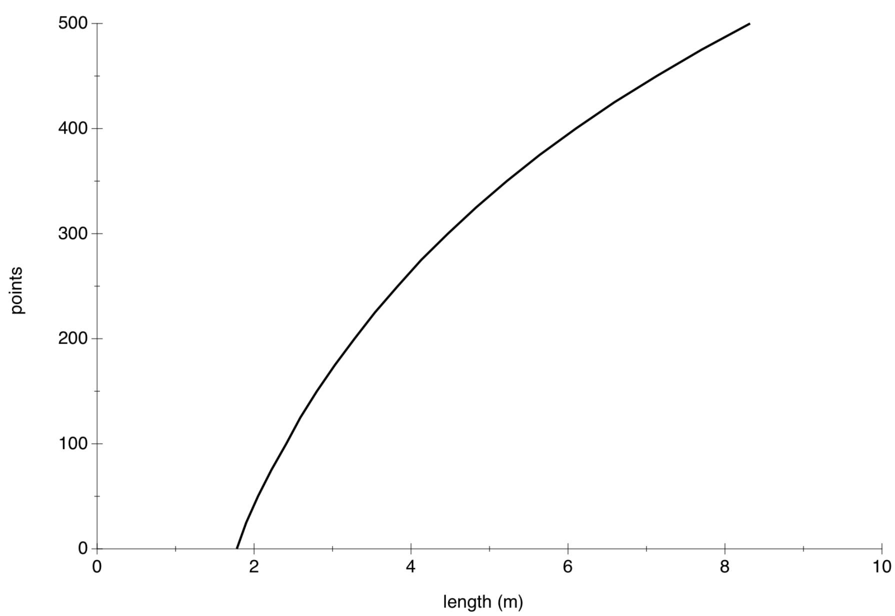

Figure 2. *Scoring of long jump (points vs. length of jump) according to the Letessier tables.*

In fact, the Letessier tables are regressive for all events be they track or field. The regressive character is an intrinsic feature of scoring based on the log-normal distribution. This is something that was known to the Letessier brothers and was considered as a desirable feature. They say that a progressive scoring would incite the athletes to try to improve their strong events and neglect the weak ones (and they give the example of the 1500 m for the latter). Unfortunately, the experience of several decades of regressive decathlon tables has shown that the exact opposite is true: the athletes would not make extra efforts for an ever diminishing reward.

2.3 The Purdy tables

At the beginning of the 70s one of us, (G.P.), proposed a new scoring system for athletics [Purdy 1972]. The main motivation to develop scoring tables was the fact that the scoring tables in use at that time were (due to a basic reasoning flaw [Ulbrich 1950]) regressive for the field events with a resulting stagnation of performances. The Purdy tables were based on a set of 10 principles. We shall not go into all the details here, but point out that one of those principles was that the tables should be progressive throughout. Moreover, for track events, instead of being constructed around time, they were using velocity, which, as explained in the introduction, is the proper choice if one wishes to build tables based on physics.

A substantial advantage of the Purdy tables was that they were based on a simple, explicit, mathematical expression

$$p = g(x-z) + a\left(e^{b(x-z)} - 1\right), \tag{1}$$

where x is the performance, z is the zero-scoring performance, and a, b, g are three adjustable parameters. Thus, the Purdy system comprises a linear part and an exponential one. For a whole range of (low) performances the scoring is linear but when the performances reach a high level the exponential term dominates and leads to a strongly progressive scoring.

The strong progressiveness can be assessed by the example on long jump scoring as per the original Purdy proposal:

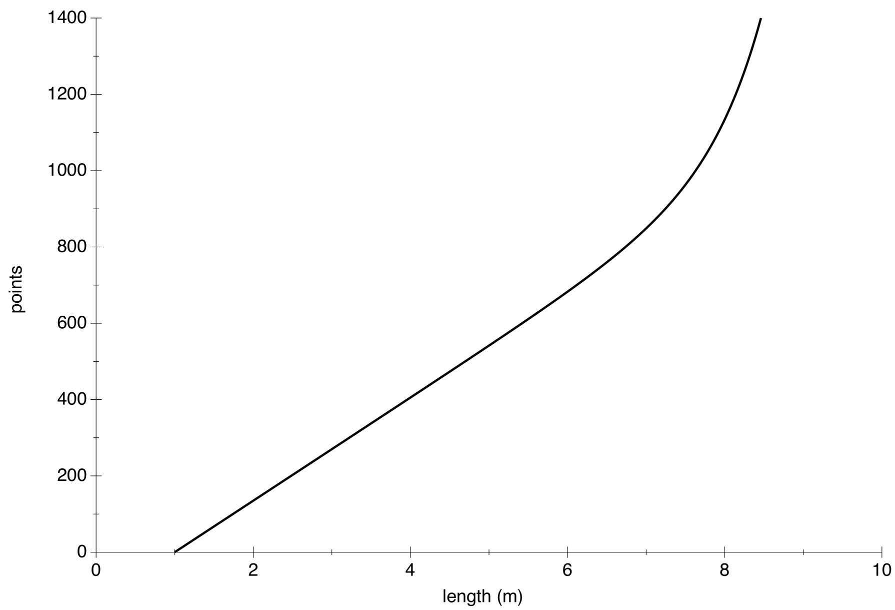

Figure 3. *Scoring of long jump (points vs. length of jump) according to the original Purdy proposal.*

The same progressiveness is also present in the track events. In Figure 4, below, we give the scoring for the 100 m, as a function of the mean velocity of the race.

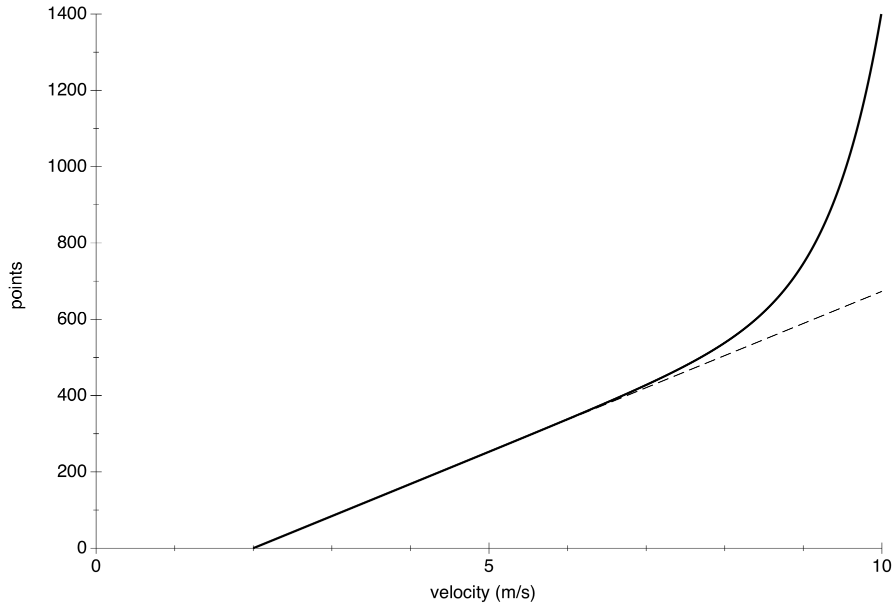

Figure 4. *Scoring of the 100 m (points vs. velocity) according to the original Purdy proposal. The dashed line corresponds to the scoring obtained from the linear term in (1).*

The thin, dashed line in the Figure is obtained from just the linear term of the scoring, using (1). We see that the contribution of the exponential term is negligible below roughly 400 points, which, in the case of the 100 m, correspond to a velocity of circa 6.5 m/s (and a time of 15.4 s).

While the strong progressiveness may look somewhat excessive, in the light of the evolution of the scoring tables which followed Purdy's work, it is definitely something that can be adjusted without necessitating a complete reshuffling of the model. We shall revisit the Purdy model in the next sections.

2.4 Harder's approach

D. Harder developed a theory of scoring of very wide applicability. The main idea behind his approach is that "two performances are equivalent when the number of athletes who obtain them is comparable", where by "number of athletes" Harder refers to the fraction of the practitioners. The application of this principle allows meaningful comparisons even between totally different sports like, say, between car racing and mountain climbing. What one needs are detailed statistics of each sport, covering the largest possible domain of performances. Stated thus, the approach is mainly qualitative. However Harder goes one step further and proposes a quantitative approach that can be used for the construction of scoring tables. The method is the following:

A mark of 100 points is attributed to some performance if a fraction of 0.5 of the population can realize a score equal or better than this. For 200 points, only a fraction of 0.05 of the population can do better than this performance. The next 100 points, i.e. 300, correspond to a performance realized by just 0.005 of the population and so on up to 1000 points where only a fraction 5.10^{-10} of the population can realize the corresponding performance. A distinctive feature of Harder's approach is that the scoring of performances of the lower 50 % of competitors follows a different curve. In fact he makes an attempt to accommodate an as-large-as-possible part of the population (in practice 95 %, which means that 0 corresponds to the performance obtained by at least 95 % of the total population).

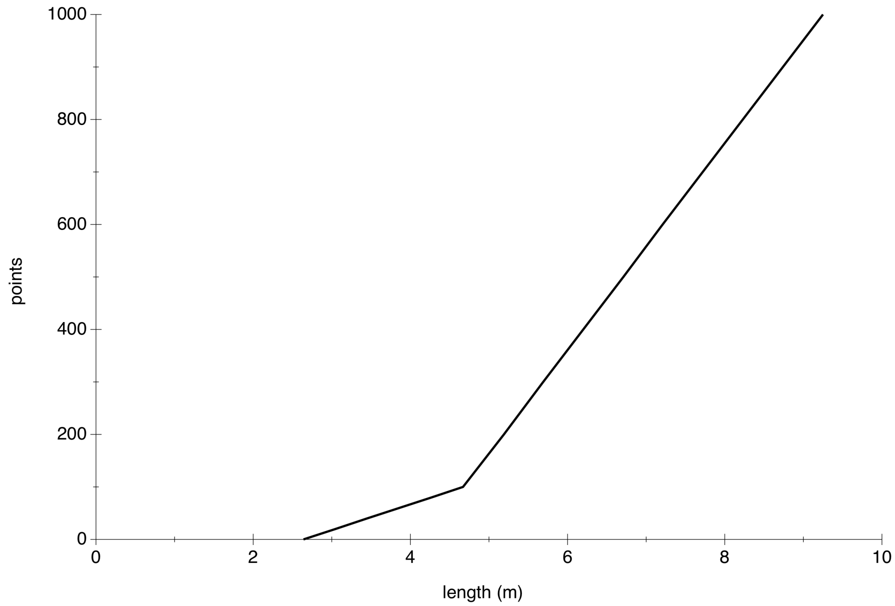

Figure 5. *Long jump scoring (points vs. length of jump) according to Harder.*

In the example presented in Figure 5 Harder's scoring follows a linear relation of points to performance, all the way down to 100 points, whereupon the slope changes and we have again a linear dependence up to zero points. Thus Harder's scoring is neither regressive nor progressive, attributing the same increase in points to the same increase of performance (past 100 points, of course) independently of the value of the performance.

2.5 The World Athletics tables

The World Athletics international federation maintains two scoring tables. The tables due to B. Spiriev (also known as Hungarian tables) are used for the scoring of all events. On the other hand the Technical Committee working group proposed scoring tables, aimed specifically at combined events, adopted in 1984 and have been in use since that date. While the details of the two set of tables do differ, they are both based on a simple mathematical expression (which we give in the case of jumps and throws)

$$p = a(x - z)^c \tag{2}$$

where p are the points, x the performance, z is the zero-scoring performance, and a, c two parameters. The graphic below shows the scoring for long jump obtained from the decathlon tables.

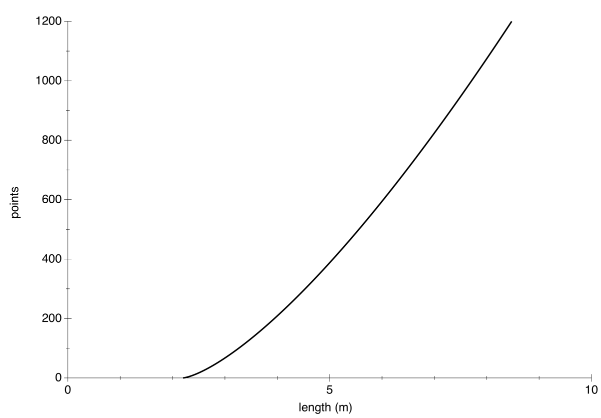

Figure 6. *Long jump scoring (points vs. length of jump) according to the decathlon tables.*

We remark that the scoring is moderately progressive. The same is true for all jumps and, to a lesser degree, for throws.

The scoring of track events asks for special attention. Here, the points are related to time, t, by the formula

$$p = a(z - t)^c. \tag{3}$$

It is somewhat curious that the authors of the World Athletics scoring tables opted for this nonphysical construction. Admittedly, time is the quantity in which the results of all races are expressed. It is the quantity that speaks directly to sportsmen and the public. However from a physical point of view the quantity which should have been used in order to link points to performance is the velocity v (where this term refers to the mean velocity of the race $v = s/t$, s being the length of the race). Using the velocity we can write (3) as

$$p = a\left(z - \frac{s}{v}\right)^c. \tag{4}$$

And the question which naturally arises is how does the relation of points to velocity look. The result for 100 m is presented in the Figure below..

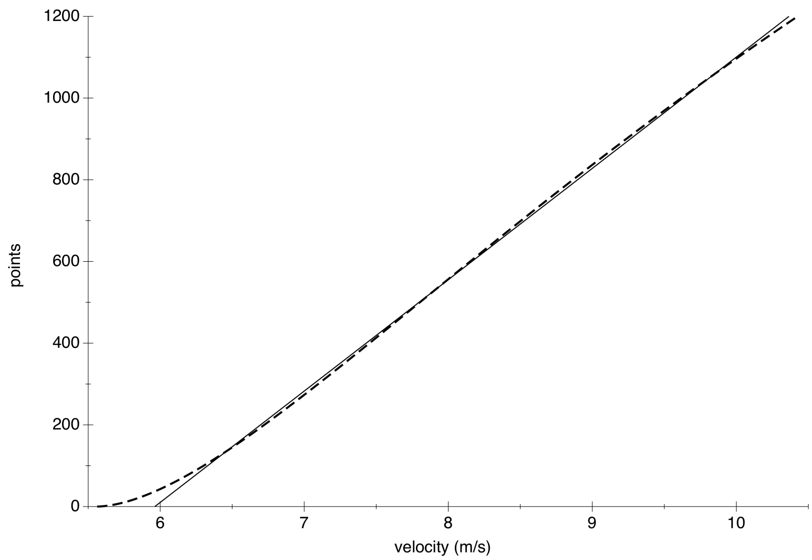

Figure 7. *Points versus velocity from the WA scoring tables for the 100 m. (The straight line fit for performances in the 300-1200 range of points is there in order to guide the eye).*

We remark that, given the values of the parameters and the small range of velocities involved, the relation is almost rectilinear for the major part. However it is clear from the expression (4) that, were one to consider non-realiztically high velocities, the relation would become regressive (and this can already be seen in Figure 7 thanks to the eye-guide straight line). One can wonder why an expression such as (4) was chosen rather than one where the velocity would appear directly in a numerator. To this end we fitted the points in Figure 7 by an expression of the form $p = A(v-B)^C$. Having obtained the best fit, we proceeded to compare it to the curve relating points to time. The results are given in Figure 8 that follows.

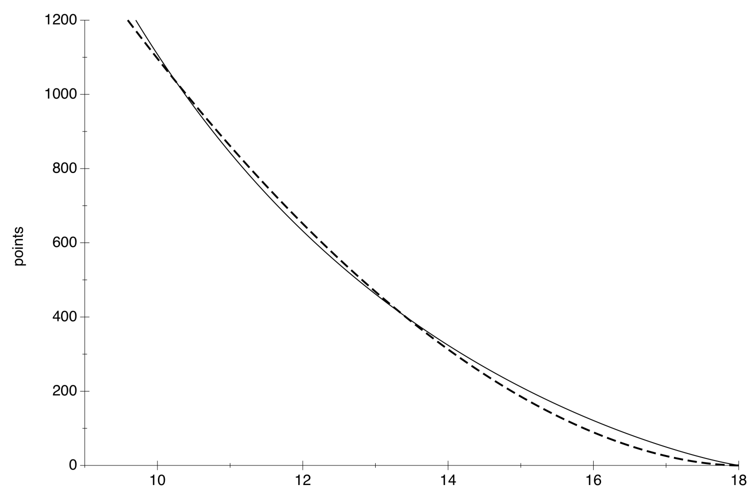

Figure 8. *Points versus time from the WA scoring tables for the 100 m. The dashed line is obtained from (3) while the continuous line corresponds to what is obtained from the best fit with $p = A(v-B)^C$, given here as points vs time.*

The agreement is acceptable over the whole range of times, except perhaps for the ones corresponding to very few points. So, World Athletics could very well have based their scoring tables on the more physical expression $p = a(v-z)^c$.

3 A mathematical approach to scoring

The main idea, already encountered in the Letessier approach, is that the scoring must somehow be related to the number of athletes realizing some performance. This means that one must have knowledge of the performances' distribution. It is a well-established statistical fact that the distribution of human performance in various domains follows a, perhaps asymmetrical, bell-shaped curve. The details may vary but the fact remains that most individuals perform close to the median with only a very small percentage registering exceptionally good or bad performances. Given this, the fraction of the population realizing a performance better than some threshold should be given by a sigmoid curve, which is just the integral of the bell-like one.

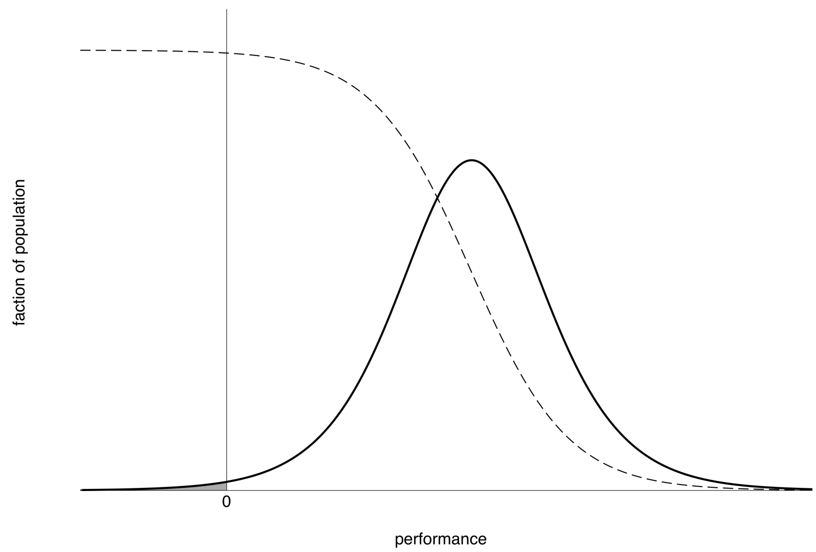

Figure 9. *An example of distribution density of performances (continuous line) and of the corresponding (complement of the) cumulative function (dashed curve).*

A remark is in order concerning Figure 9 above. It is interesting to draw the attention to the small shaded area lying beyond the zero of the performance axis. Of course this is an unphysical region since negative performances do not have a meaning (although the term is often encountered in a figurative sense). What is interesting is that the distribution curve crosses the vertical axis at a positive value. This is what we were referring to in the discussion of the zero-point when we pointed out that "there exists a (small) percentage of the total population unable to realize any performance".

Going back to Harder's relation of points to performance, we remark that a linear progression in points corresponds to a geometric progression in the fraction of the population. Thus it is clear that some logarithmic dependence is in play here. This is the relation postulated by one of us (B.G.) in [Grammaticos 2007] and which was used in order to propose a new scoring system. It will be the basis for the developments presented in this article. But, before proceeding further, we shall need a little of mathematical formalism.

We start by assuming that the fraction of population who realize a given performance follows a distribution like the one in Figure 9 (or perhaps 1). We represent it by $f(x)$, where x is a measure of the performance, for instance a velocity or a length, when we are interested in athletics. The meaning of $f(x)$, which is known as the probability density function, is that if we consider a small interval Δx of performances, the fraction of the population realizing performances in this interval is $f(x)\Delta x$. The cumulative distribution function is the area under the curve of $f(x)$ or more precisely $G(x) = \int_{-\infty}^{x} f(s)ds$ (and thus $G(\infty) = 1$). For our approach we need the complement of the latter i.e. $F(x) = 1 - G(x)$. (In what follows we shall refer to it by the acronym CCD). This is the curve represented by the dashed line in Figure 9. The meaning of the CCD function $F(x)$ is that it gives the percentage of population who can realize a performance equal or higher than x.

The mathematical formulation proposed in [Grammaticos 2007] is that the points p of a consistent scoring system should be related logarithmically to F. Thus the basic expression is

$$p(x) = -\log F(x)/\lambda, \tag{5}$$

where log is the logarithm and λ a positive constant. Using the properties of the logarithm it is more convenient to represent the relation (5) as $p(x) = \log(1/F(x))/\lambda$. If we start from an known expression for $p(x)$ we can simply obtain the CCD function $F(x)$ by

$$F(x) = e^{-\lambda p(x)}, \tag{6}$$

where e is the basis of the natural logarithms. Having $F(x)$ we can obtain the distribution density function as $f(x) = -F'(x) = \lambda F(x)p'(x)$ where the prime $'$ indicates a derivation with respect to x.

The scoring system proposed in [Grammaticos 2007] by one of the authors (B.G.) was based on the assumption that the distribution of performances is a logistic one. The explicit expression is

$$F(x) = \frac{1}{1+ce^{dx}}, \tag{7}$$

and its graphical representation follows the dashed line in Figure (9). When $x = 0$ the value of F is equal to $1/(1+c)$, which, for small c, is roughly equal to $1-c$. Thus c is the fraction of the population unable to obtain even zero performance. (Here 'zero' means infinite time or zero meters). If one assumes, like Harder, that 5 % of the population cannot obtain the zero-point performance then c would have to be equal to 0.05. (But we must stress here that, in the case of Harder, the zero point is *not* a real zero, being, for instance, 2.5 m for the long jump). In what follows we shall work with smaller values for c in the range of 0.01-0.0001.

Having an explicit expression for $F(x)$, we can, using (5), obtain the expression for the scoring points

$$p = g\log\left(1+c(e^{dx}-1)\right), \tag{8}$$

(where $c(e^{dx}-1)$ is introduced instead of ce^{dx} in order to ensure that $p = 0$ for $x = 0$). In the figure below we give an example of scoring obtained with (8) which has been adjusted so as to follow the one of Harder for long jump, while ensuring that p is non-zero for any positive x.

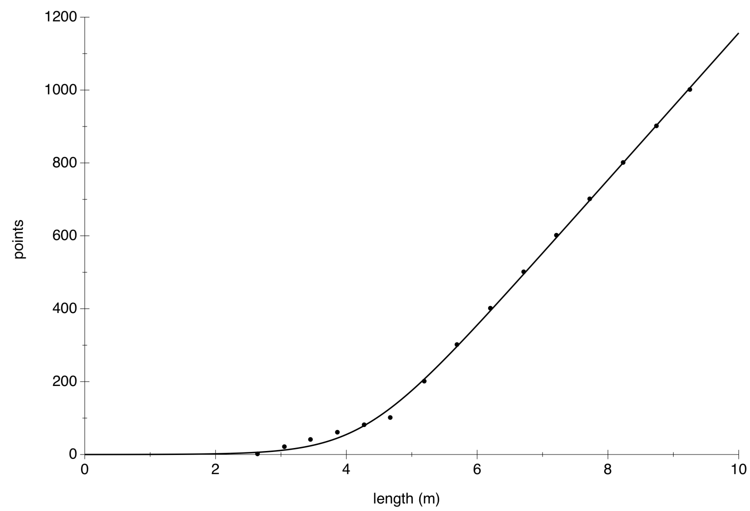

Figure 10. *Scoring (points vs. length of jump) with expression (7) in the case of men's long jump, following Harder. The points are those given by Harder while the continuous curve is the best fit using (8).*

The agreement is remarkable, and this is true for all field events. On a pure anecdotal level, the performance in men's long jump that would score 1 point, according to the fit of figure 10 is 1.65 m.

As a conclusion, if one knows the distribution of performances, calculating the corresponding scoring function is straightforward. The reverse is also true and will be the object of the next section.

4 From scoring to distributions

The Letessier tables were based on an explicit assumption of an underlying performance distribution while the Harder approach has served as a basis for the proposal of the link between performance distribution and scoring presented in [Grammaticos 2007]. In this section we shall use the theory outlined above in order to obtain the distributions which would correspond to the Purdy and World Athletics scoring systems.

4.1 The distributions underlying the Purdy approach

As we have seen in the section 2, the Purdy tables are based on an explicit mathematical expression linking the points to the performance. It is thus straightforward to apply the theory of section 3 and obtain the CCD function F. Using expression (6) we find readily

$$F(x) = e^{-g\lambda(x-z)}e^{-a\lambda(e^{b(x-z)}-1)}. \tag{9}$$

Thus the performance distribution corresponding to the Purdy system is the product of an exponential distribution (in fact, a Poisson distribution since the performances are measured with a finite time or length step) and a Gompertz one. The presence of a Gompertz distribution, which plays a major role in biology (but also in computer science) is remarkable. In the figure below we give a schematic representation of the CCD and density distribution functions corresponding to expression (8) for some choice of the parameters a, b, g, λ.

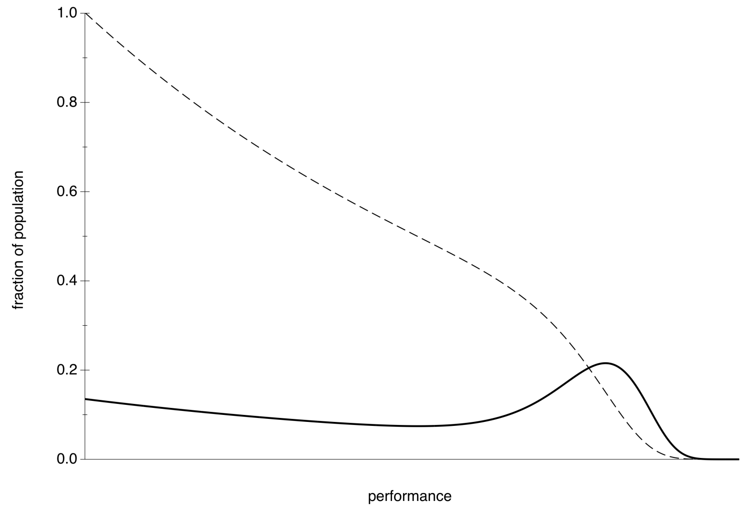

Figure 11. *The density distribution functions (continuous line) and CCD function (dashed line) obtained from the original Purdy proposal.*

Two remarks are in order at this point. First, because of the late dominance of the exponential part in the scoring, the density function is concentrated around rather elevated performances. There is nothing astonishing here. On the contrary, since this scoring method is tailored to the scoring of combined events and thus targeting experienced athletes, it is expected that the population concerned, as resulting from the distribution functions, comprise athletes realizing statistically elevated performances. Figure 12 highlights the importance of the Gompertz part for the high level performances.

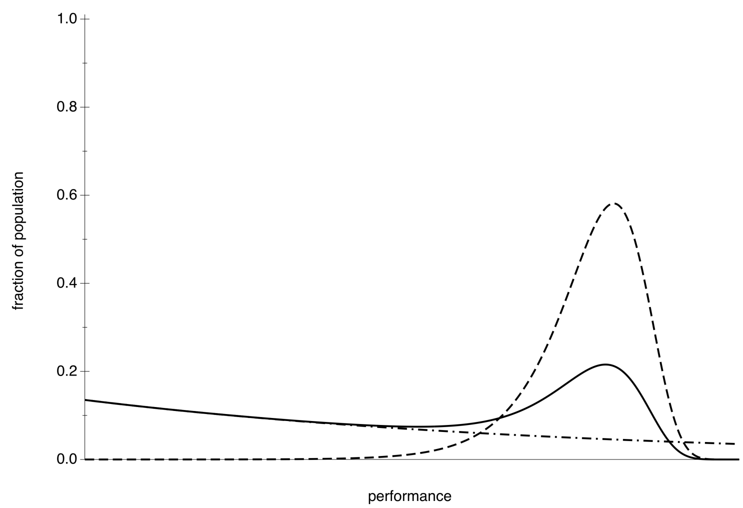

Figure 12. *The distribution density function (continuous line) from the original Purdy proposal together with the contribution from the exponential (dot-dash curve) and the Gompertz (dashed curve) parts.*

Second, the influence of the linear part in the scoring dominates the low-performance part of the distribution functions. Had the Gompertz part of the distribution been absent, we would have obtained a pure exponential/Poisson distribution as can be seen from Figure 12. This is tantamount to saying that a large part of the population can only realize modest athletic performances, as the maximum frequency is situated at the zero point. We shall return to this question in the next sections.

4.2 The distributions for the World Athletics tables

The scoring for the WA tables is also given by an explicit mathematical formula and this allows us to obtain the underlying distribution of performances. Starting from an expression like (2) for jumps and throws we obtain readily for the CCD function

$$F(x) = e^{-a\lambda(x-z)^c}, \tag{10}$$

which is precisely the (complement of) the distribution function for the Weibull distribution. A graphical representation of the distribution functions $F(x)$ and $f(x) = -F'(x)$ is shown in Figure 13 below, in the case of the long jump.

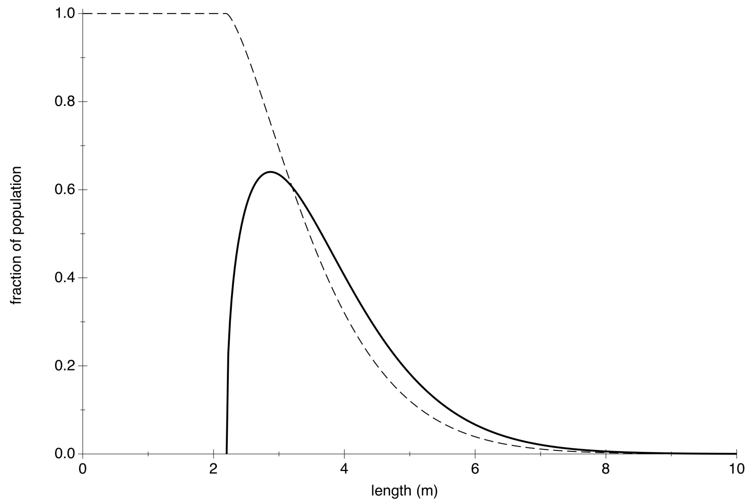

Figure 13. *The density distribution function (continuous line) and CCD function (dashed line) for the WA scoring tables.*

In the case of track events the situation is slightly more complicated since the dependence of the points p on the velocity v is $p = a(z - s/v)^c$. Still, the shape of the Weibull distribution is preserved in this case also. Figure 14 below shows the dependence of the distribution density function on the velocity

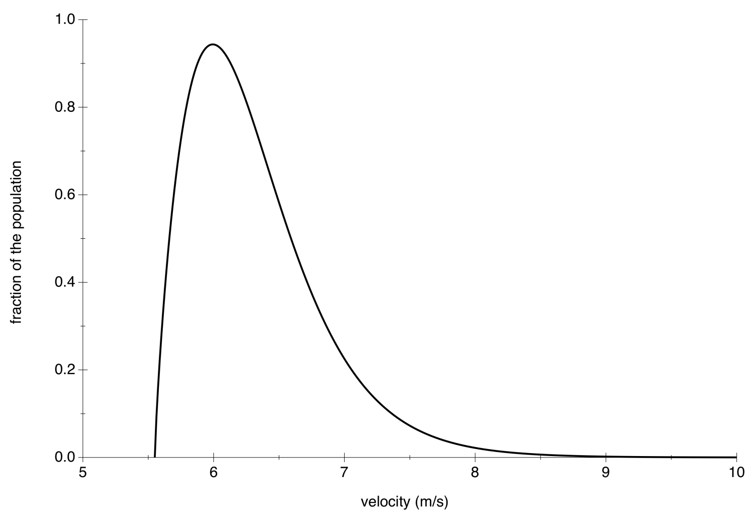

Figure 14. *The performance distribution density function for track events.*

A similar distribution function would have been obtained had we used the expression $p = a(v - z)^c$ that we proposed in section 2 for the scoring of track events.

The appearance of the Weibull distribution related to the scoring of athletic performances is quite remarkable. The Weibull distribution has already been proposed as an adequate descriptor of the human performance [Berry 1981]. Still the difference between the context of [Berry 1981] and the present one is essential. While the cited work concerns a non-specialized population, where the peak of the distribution density is close to zero, in the athletic scoring case, we are addressing a population of specialists, resulting in the distribution starting some distance from a zero performance.

It is interesting that in the two cases where scoring tables were proposed without an explicit assumption of a distribution of performances, our theory led to two well-known distributions, Gompertz and Weibull. On the other hand the presence of the exponential/Poisson distribution in the case of the Purdy approach is perhaps suboptimal. Improving upon this assumption will be the object of the next section.

5 Improving the Purdy approach

AAs we have remarked in the previous section, a drawback of Purdy's original scoring approach was the linear part which is associated to an exponential/Poisson distribution, leading to an unrealistic distribution of performances like the one of Figure 11. Still the advantages of the Purdy approach are, first, a (very) low null-score performance and, second, the progressiveness associated to the exponential part. Those are desirable features which should be preserved in any modification. This led us naturally to consider blending (a part of) the Purdy scoring to the one proposed by another of the authors (B.G.) based on the hypothesis of the logistic distribution. The new scoring we propose is thus

$$p = a\left(e^{bx} - 1\right) + g \log\left(1 + c(e^{dx} - 1)\right), \tag{11}$$

where, with respect with the original Purdy approach, we have now one extra parameter (but, as we shall explain, some of the parameters can be easily fixed). In order to illustrate the proposed scoring we show in Figure 15 a scoring for long jump where we have just arranged the parameters of (11) so as to give the same scoring as the WA tables for $p = 700$.

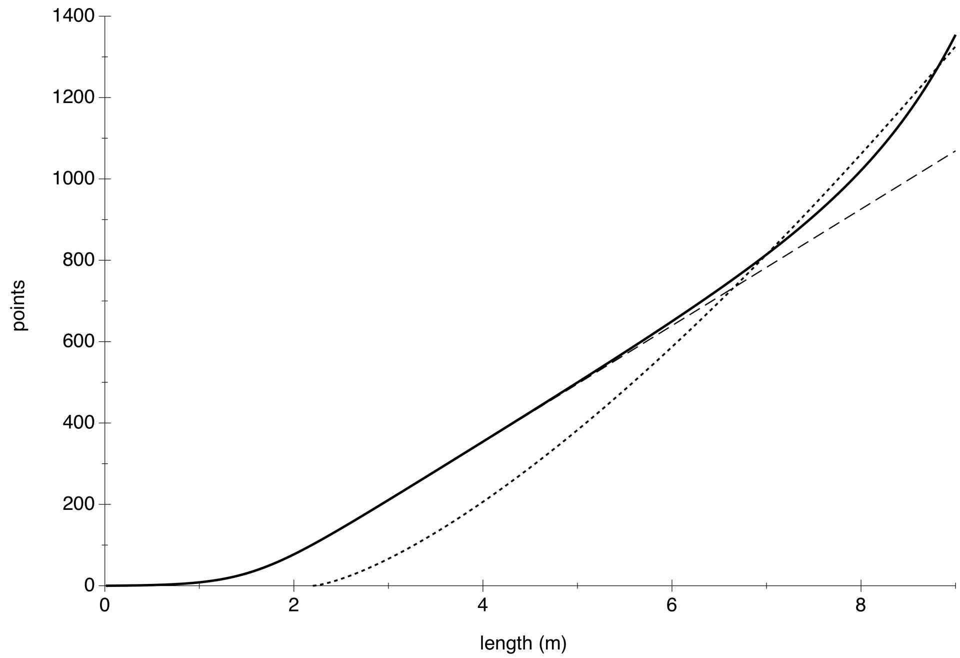

Figure 15. *Scoring of long jump (points vs. length of jump) based on (11).The continuous thick curve corresponds to the scoring from (11) while the dashed line is obtained from just the second part of (11) coming from the logistic hypothesis. The dotted line is the scoring obtained from the formula of World Athletics (3) with the corresponding parameters.*

The two curves from expressions (11) and (3) may look very different, but this is essentially due to the zero point, fixed at 2.2 m, in the case of (3). It is in fact more informative to relinquish the constraint of the null-score performance in the WA scoring formula and obtain a best fit between (11) and (3), with fixed exponent c (equal to 1.4), where z is now a free parameter. Figure 16 shows the result of the best fit.

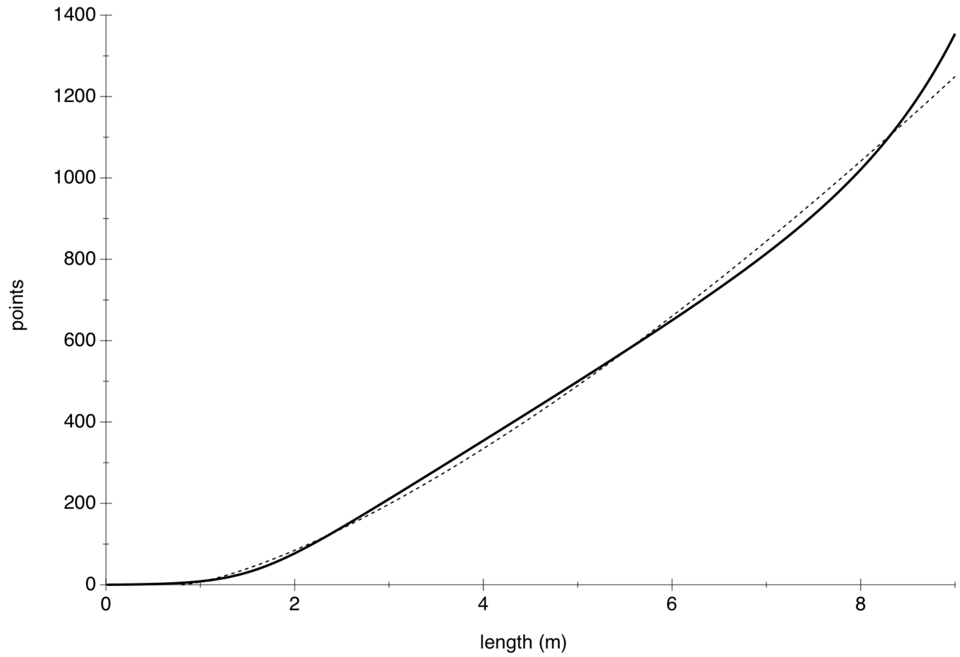

Figure 16. *Fit of the scoring obtained from (11) (thick line) with an expression like (3) (dotted line) with z free.*

The overall fit is now quite satisfactory (and of course it could be improved if we went back and modified the parameters of (11) to values different from the ones used for Figure 15. We find that the value of null-score z corresponding to the best fit is $z = 0.8$ m. And since, due to the logistic hypothesis the zero point for the new tables is 0 m, it is interesting to point out that 1 m (which was the null-score one in the original Purdy scoring) corresponds to just 8 points, which for all practical purposes is as good as zero. This is another advantage of the new method: one does not have to worry particularly about fixing the null-score performance. We will not pursue the question of the parameter fit here. This technical point is discussed in full detail in the Appendix.

Once the scoring formula is given, we can easily calculate the CCD function. We find readily

$$F(x) = \frac{e^{-a\lambda(e^{bx}-1)}}{(1+ce^{dx})^{\lambda}}, \tag{12}$$

i.e. a Gompertz term (coming from the exponential part of the Purdy scoring) and a logistic term. (Note that we have used, for the logistic term, a slightly different expression from the one in (11). In fact the expression $1+ce^{dx}$ is the one expected from the logistic assumption and the modification $1+c(e^{dx}-1)$ was introduced so as to make sure that the scoring curve starts at 0). Due to the relation (6), a parameter λ does appear in the distribution function and in particular as an exponent in the denominator. This is far from insignificant since it transforms the logistic distribution to what is known as skew-logistic. We shall return to this point in the final discussion. In Figure 17 we present the CCD (dashed line) and the density (continuous line) distribution functions (the latter being equal to minus the derivative of the first, as we have seen in section 3), again in the case of the long jump.

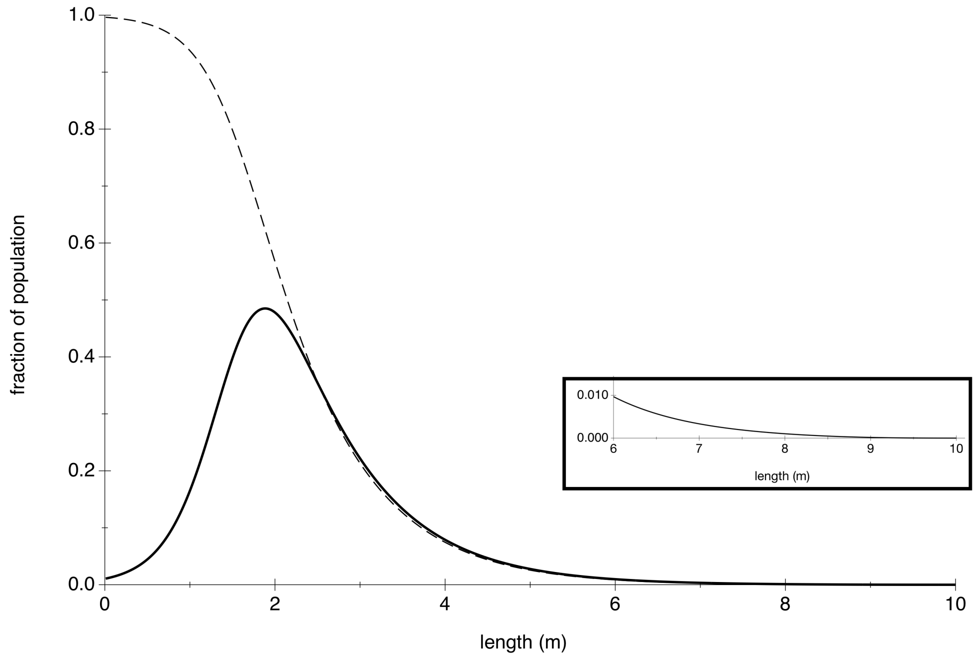

Figure 17. *The density distribution function (continuous line) and CCD function (dashed line) obtained from (12), with an inset in order to visualize the tail of the distributions at high performances.*

As expected from the logistic hypothesis, the value of the CCD function is not strictly equal to 1 at 0 differing from it by 0.003 (something quite acceptable as explained in section 2). It is remarkable that while the distribution has a long tail, the drop-off is exponentially rapid and thus only a very small fraction of the population is able to realize high performances (as one would intuitively expect).

At this point, it is interesting to assess the relative importance of the Gompertz and skew-logistic factors. Figure 18 shows just that.

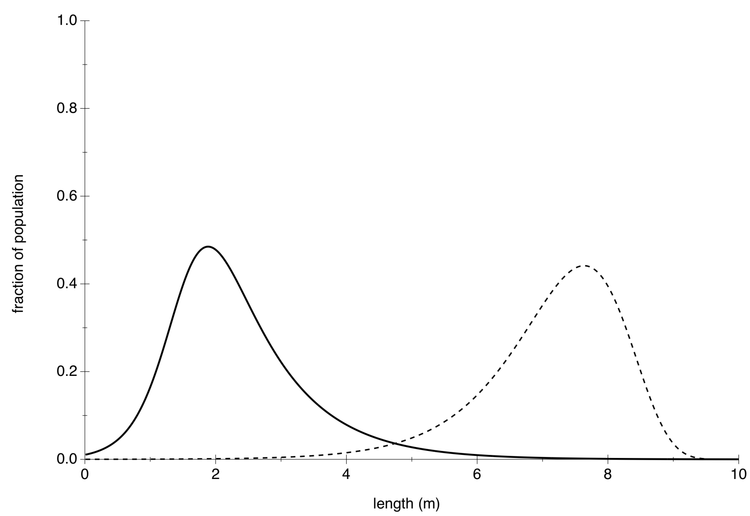

Figure 18. *Skew-logistic (continuous line) and Gompertz (dotted line) distribution density functions in (12).*

What is somewhat unexpected is that the skew-logistic distribution dominates completely. There is practically no visible difference between the total distribution density of Figure 17 and the one obtained from the skew-logistic term in Figure 18. Still, the Gompertz part is quite appreciable. We remark that it is centered around high performances, the ones we would expect from elite athletes. However it is washed out in the total distribution when multiplied by the very-low-value tail of the logistic term. Despite this, its role is essential: without this term present we would not have had the highly progressive part in the scoring. Thus, while the performances distribution naturally represents the athletic elite as a minuscule minority, a correctly built scoring system takes this particular population into account and provides a scoring tailored to its performances.

6 Summary and discussion

The first objective of this paper was to establish a link between the distribution of athletic performances and scoring. The idea is not new: it was explicitly presented as the basis of the work of the Letessier brothers for the construction of scoring tables to be used in the French educational system. But, in fact, the same idea has been underlying most scoring approaches based on tables giving a correspondence between performance and points. Four scoring approaches were examined in this paper. As just stated, the Letessier tables are based on the assumption of a specific distribution of performances. However no explanations [Letessier 1992] are offered by the authors as to the construction of the tables themselves. The approach of Harder is based on the idea of a performance distribution involving percentiles of the population but no explicit mathematical formula is offered for the construction of the tables. On the contrary both Purdy's and World Athletics' tables are based on explicit mathematical expressions. They are both progressive tables, contrary to Letessier's (regressive) and Harder's (neutral). While Purdy formulates his tables for track events in terms of the velocity, the WA tables use time in a formula that could lead to a regressive scoring, were it not for the limited range of velocities involved. What was lacking in the case of these two tables is a relation to a distribution of performances. This was remedied in this paper. Our study is based on a previous work of the authors and was expanded here to a full-fledged theory linking scoring and performance distributions. As an application of this theory, we obtained the distributions associated to Purdy's and World Athletics' scoring table. It turned out that in the first case the distribution is a product of an exponential/Poisson distribution with a Gompertz one, while in the second case the distribution is the one known under the name of Weibull.

The second objective of the paper was to propose a novel, improved version of the scoring proposed by Purdy. This was done first at the distribution level by removing the exponential/Poisson part and replacing it by the logistic distribution already proposed in [Grammaticos 2007]. One of the advantages of the inclusion of the latter in Purdy's approach is the fact that one can ignore the question of null-score performances. The new tables attribute zero points to exactly zero performance but, given their structure, the number of points attributed to very low performances is so small that the question of fixing a null-score one becomes irrelevant. We believe that the new scoring expression, were it to be adopted by World Athletics, would turn out be most useful and lead to a fair method of scoring. Calibrating our scoring formula for the various disciplines of athletics is a task that would require additional efforts and was not considered in this paper. Here we have limited ourselves to what is essentially a proof of concept, which can serve as the basis for future developments.

One interesting result of the present study was the discovery that the bell-shaped distribution of performances that was considered in [Grammaticos 2007] is in fact an unwarranted limitation. The proper performance distribution obtained here is a skew-logistic one. There has been of recent quite some excitement [Bernard 2018] over the realization that human performances do not follow a bell curve. Unfortunately oversimplifications led to the conclusion that a power-law, Pareto-like, distribution [Boyle 2012] was the proper parametrization. This would mean that, when it comes to sports, the great majority of the population would realize performances very close to zero. We believe this to be an exaggeration. Although most people would realize rather modest performances, the maximum of the distribution should lie at a (rather small) value away from zero. Thus, the skew-logistic can effectively represent the distribution of human athletic performances for a very large population sample and perhaps even the totality of it. The authors of [Boyle 2012] talk about the "best and the rest" in the title of their paper, but go on to present a single distribution for the performances of both. They rely on the thick, Pareto-like, tail of the distribution in order to distinguish the two classes. We claim that this is an oversimplification. The specialists (the "best" of [Boyle 2012]) have their own distribution.

In order to support this argument, we present below the distribution of velocities of the Boston Marathon finishers obtained from [RS blog 2014]. The choice of this set of data was dictated by fact that the Boston Marathon is not open to everybody; one has to qualify for it and thus the participants are selected among the "best". Figure 19 below shows the velocity distribution along with the best fit by a Gompertz distribution. The agreement is striking.

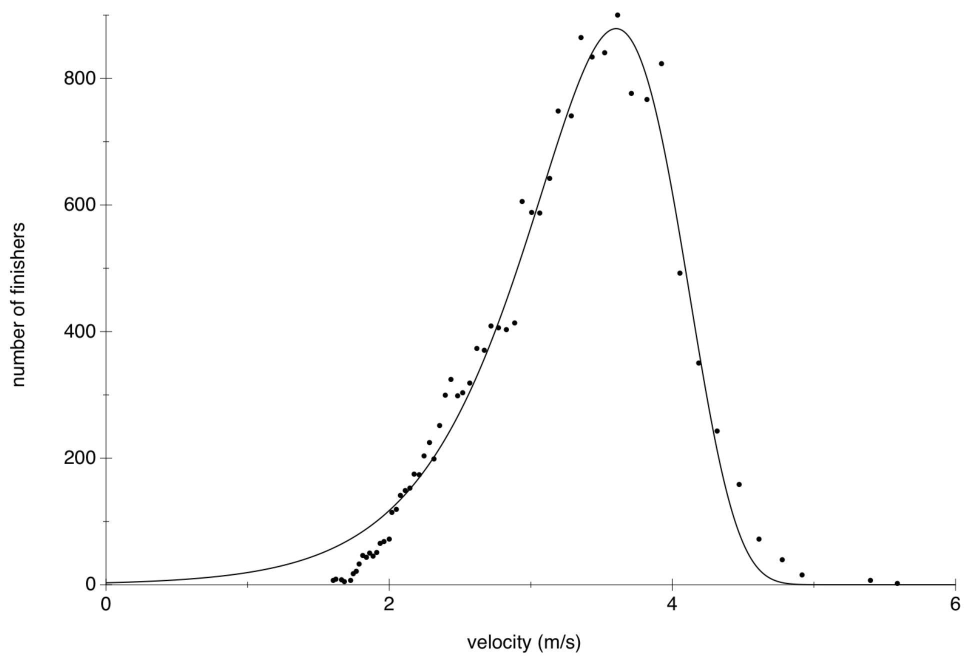

Figure 19. *Velocity distribution (number of finishers vs. velocity) of the Boston marathon finishers.*

It goes without saying, that when one looks at the distribution, that includes everybody, the presence of a certain elite may not make a difference, apart from thickening the distribution tail. However, if one wishes to introduce a scoring for athletic events involving elite athletes, one must consider a distribution of this sub-population and allow for a rewarding scoring of their performances. This does not alter the conclusions concerning the total population and does provide a fair scoring over the whole range of performances.

Acknowledgments

The authors acknowledge interesting discussions with D. Harder, J. McNitt-Gray, C. Robinson, H. Stewart and R. Wilcox.

References

References

[Crowther 2007] N.B. Crowther, *Sport in Ancient Times*, Praeger (Westport CT), 2007.

[Golden 2004] M. Golden, *Sport in the Ancient World*, Routledge (London), 2004.

[Miller 2004] S.G. Miller, *Ancient Greek Athletics*, Yale Univ. Press (New Haven), 2004.

[Leonard 2016] J. Leonard, *The Value of Athletic Glory in Ancient Greece*, at https://www.greece-is.com/author/johnleonard/, 2016.

[Hymans 2021] R. Hymans, *Progression of World Athletics Records*, World Athletics (Monaco), 2021.

[Zarnowski 2013] F. Zarnowski, *The Pentathlon in the Ancient World*, McFarland (Jefferson NC), 2013.

[Grammaticos 2013] B. Grammaticos, *Ancient pentathlon revisited*, online at https://rb.gy/2lurdt.

[Letessier 1961] J. Letessier, *Table de Cotation des Performances Athlétiques*, Education Physique et Sport (Paris), 1961.

[Sarnowski 1989] F. Zarnowski, *The Decathlon*, Leisure Press (New York), 1989.

[Ulbrich 1950] K. Ulbrich, *Leichtathletik Punktwertungen: mathematisch, physikalische, statistische Behandlung*, Fed. Inst. Phys. Ed. Vienna 3 (1950) 73.

[Purdy 1974] J.G. Purdy, *Computer generated track and field scoring tables*, in three parts: I. Historical development, II. Theoretical foundation and development of a model, III. Model evaluation and analysis, Med Sci. Sports 6 (1974) 287, 7 (1975) 111, 9 (1977) 212.

[Woolf 1968] M. Woolf, *Isomarks?*, NUTS Notes 6 (1968) 4.

[Grammaticos 2007] B. Grammaticos, *The physical basis of scoring the athletic performance*, New Stud. Athl. 22:3 (2007) 47.

[DiPrampero 1986] P.-E. DiPrampero, *The energy cost of human locomotion on land and in water*, Int. J. Sports Med. 7 (1986) 77.

[Harder 2001] D. Harder, *Sports Comparisons*, Education Plus (Castro Valley CA) 2001.

[Spiriev 2017] B. Spiriev, *IAAF Scoring Tables of Athletics*, World Athletics (Monaco), 2017.

[Trkal 2016] IAAF Scoring Tables for Combined Events, World Athletics (Monaco), 2016.

[Purdy 1972] J.G. Purdy, *Least square model for the running curve*, Res. Q. Exerc. Sport 45 (1974) 224.

[World Athletics 2021] World Athletics Competition Rules, World Athletics (Monaco), 2021.

[Fortune 2008] Y. Fortune and Y. Saint-Martin, *La table Letessier au service de l'intégration progressive du sport dans l'éducation physique scolaire*, Le Télémaque 34 (2008) 87.

[Purdy 1972] J.G. Purdy, *The application of computers to model physiological effort in scoring tables for track and field*, PhD thesis, Stanford Univ. 1972.

[Berry 1981] G.L. Berry, *The Weibull distribution as a human performnce descriptor*, IEEE transactions SMC-11, (1981) 501.

[Letessier 1992] P. Letessier, *Table Letessier. Rappel de la démarche d'élaboration*, Revue EPS 234 (1992) 55.

[Bernard 2018] V. Bernard, *Human performance: the bell curve is a myth*, online at https://www.linkedin.com/pulse/human-performance-bell-curve-myth-vincent-bernard

[Boyle 2012] E. O'Boyle Jr. and H. Aguinis, *The best and the rest: revisiting the norm of normality of individual performance*, Pers. Psy. 65 (2012) 79.

[RS blog 2014] Runners Stats Blog, *Unusual distribution for Boston marathon finish times*, online at https://rb.gy/cnlzjt.

Appendix: fitting the parameters

In this paper we proposed a new formula for scoring the athletic performances, based on an explicit mathematical expression involving 5 parameters. In view of practical applications it is important to explain how the parameters can be adjusted.

We start with the expression linking the performance x to the number of points p given in Section 5:

$$p = a\big(e^{bx} - 1\big) + g\log\big(1 + c(e^{dx} - 1)\big). \tag{13}$$

It is clear that we can multiply the number of points by an arbitrary factor so as to obtain a number of points p_m for some performance x_m, both decided in advance. For instance we can decide to attribute 1000 points to a (men's) pole vault performance of 5 m. This allows us to divide expression (13) by a factor p_m, asking simply that when $x = x_m$ the number of points is equal to 1. (In order tor recover the proper scoring it suffices to multiply by p_m). Also we can divide x by x_m, which just amounts to modifying the values of b, d by the appropriate factor. The advantage of this scaling of both p and x is that we can work in the interval [0,1] and reason in terms of fractions or percentages instead of absolute values, which vary for each discipline.

Figure 20 below shows a typical scoring curve, where the maximum number of points was scaled to 1 corresponding to an equally scaled maximum performance of 1.

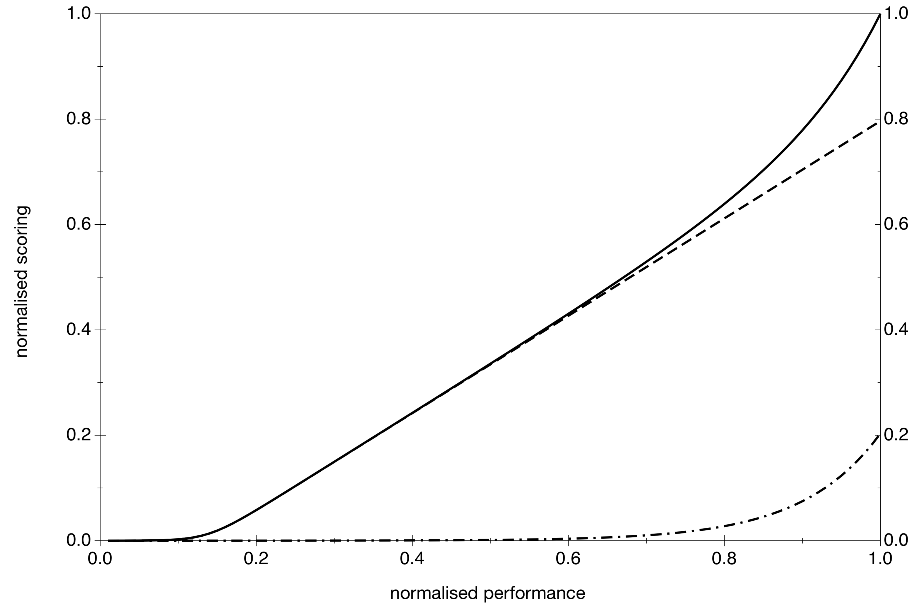

Figure 20. *An example of normalized scoring. The thick continuous line represent the total number of points, the dashed line corresponds to the logistic part and the dot-dashed line comes form the exponential part*

A simple glance at Figure 20 suffices in order to convince one that the contribution of the exponential term in (13) arrives quite late, after the logistic part has stabilized to a linear contribution. This allows us to fit the parameters of the two terms separately. Moreover, as explained in the main body of the paper, the parameter c is directly related to the percentage of the population that cannot realize any performance. We argued there that a good estimate of this quantity is in the 0.01-0.0001 bracket. Thus we can fix the value of c once and for all at 0.001.

In order to fix the parameters g and d we decide that the value of the points corresponding to the null-score performance x_n must be p_n (a small value in the 0.001-0.01 bracket, i.e. 1 to 10 points). The null-score performance was detailed in the original Purdy article and was further elaborated upon in the recent article mentioned at the beginning of this paper. It is determined for each discipline following Purdy's criteria. We have thus a first equation

$$g \log\left(1 + c(e^{dx_n} - 1\right) = p_n. \tag{14}$$

Before pursuing work on the logistic part we turn to the exponential one. First we can decide that the exponential part starts playing a role only after some, not very small, performance. We can thus ask that the contribution of the $a\left(e^{bx} - 1\right)$ term be just p_0 (a small number in the 0.001-0.01 bracket, i.e. 1 to 10 points) for $x = x_m/2$ (we make this choice as it leads to great simplifications is the determination of parameters). Next we make the assumption that the contribution of the exponential term for $x = x_m$ (where $x_m = 1$ in normalized units) to the (normalized) points is just p_1. This allows to fix the values of a and b with elementary calculations. With p_0 and with the number of points for $x = x_m$ being p_1 (a value in the 0.1-0.3 bracket, i.e. 100 to 300 points), we find for the parameters the expressions

$$b = \frac{2}{x_m} \log\left(\frac{p_1 - p_0}{p_0}\right), \tag{15}$$

$$a = \frac{p_0^2}{p_1 - 2p_0}, \tag{16}$$

which, given that p_0 is typically much smaller than p_1 can be further simplified to $b = 2\log(p_1/p_0)/x_m$ and $a = p_0^2/p_1$.

Finally we ask that the total scoring at $x = x_m$ corresponds to p_m points, which gives the equation

$$g \log\left(1 + c(e^{dx_m} - 1\right) + p_1 = p_m. \tag{17}$$

Combining (14) and (17) we can obtain g and d. In particular eliminating g between the two we can solve the transcendental equation for d. Once the latter is obtained we can compute g from either (14) or (17).

To put it in a nutshell, one must decide beforehand what the value of c is and make the assumption that the effect of the exponential term will start being felt at $x = x_m/2$. (This is not a stringent constraint since the value of points for this performance can be freely fixed. However choosing $x_m/2$ greatly simplifies the calculations eschewing having to solve transcendental equations). Next one must compute the null-score performance x_n for the discipline at hand. By choosing the values p_n, p_0 and p_1 of points for $x = x_n, x_m/2$ and x_m one can determine the parameters of the model using (14), (15), (16) and (17). Note that the value of the parameter d can only be obtained numerically. Once the parameters are fixed, (13) gives the correct scoring.

At this point it is interesting to show how the fit procedure works. Let us give an example aiming at providing scoring for the pole vault. We assume that a performance of 5 m corresponds to 1000 points. The null score performance is taken at 0.75 cm and if we normalize the performances is such a way as to have 5 m corresponding to $x_m = 1$, the null score performance is $x_n = 0.15$. Next we decide that the null score obtains just $p_n = 10$ points. Finally we assume that the number of points due to the exponential part of the scoring is $p_1 = 200$ at $x = x_m$ and that at $x = x_m/2$ the exponential part counts only for $p_0 = 10$ points. Using (15) and (16) we obtain readily $b = 5.889$ and $a = 0.5556$. The parameter d is computed using (14) and (17). This can be done very easily graphically as shown in Figure 21 below

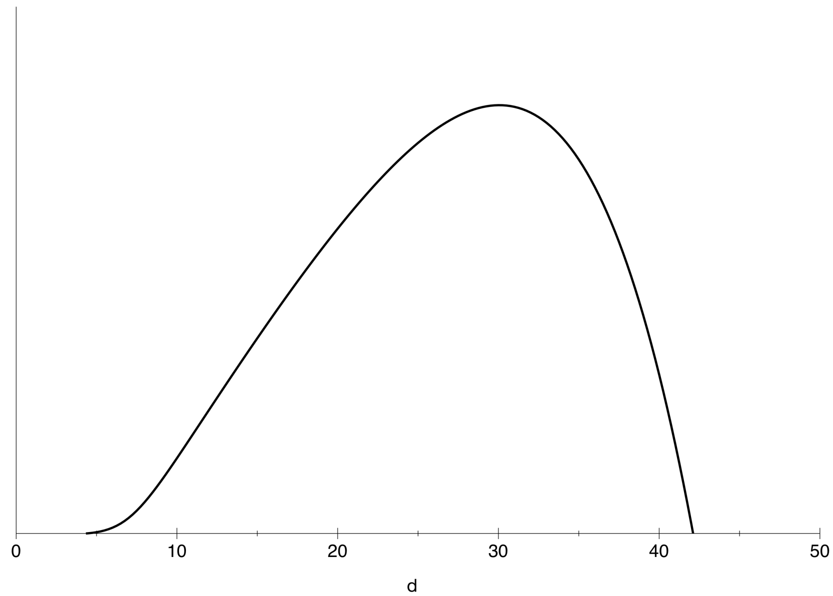

Figure 21. *Graphical representation of the equation, ratio of (17) to (14), having d as a root.*

We find $d = 42.11$ (this is the root of interest to us). Finally we obtain $g = 22.73$. We obtain thus the scoring for the pole vault, shown in Figure 22.

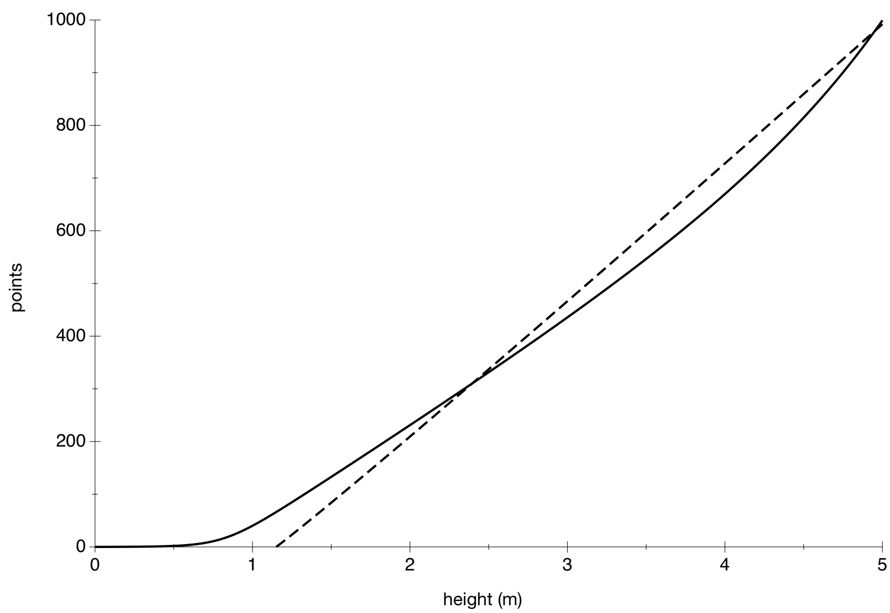

Figure 22. *Scoring for pole vault, (points vs. height of jump) obtained from (13) (continuous line) and from the World Athletics tables (dashed line).*

The comparison of the scoring we propose to that of World Athletics shows that ours is markedly progressive, while the World Athletics scoring follows practically a straight line (a best fit for an expression $p = \alpha(x - z)^\gamma$ yields an exponent $\gamma = 1.03$, uncannily close to 1).