# What is the Speed Limit for Men's 100 Meter Dash

***Reza Noubary***

**ABSTRACT**: Sports provide an inexhaustible source of fascinating and challenging problems in many disciplines, including mathematics. In recent years, due to the emergence of some exceptional athletes, prediction of athletic records has received a great deal of attention. For example, mathematicians have tried to model the improvements of records over time in order to forecast future records, including ultimate records. Records set in different sports shed light on human strengths and limitations and provide data for scientific investigations, training, and treatment programs.

This article reviews some common methods used for modeling and analysis of athletic performances and the effect an exceptional individual like Usain Bolt could have on the results. Methods discussed include trend analysis, tail modeling, and methods based on certain results of the theory of records. Data from a few athletic events including the men's 100m dash are used for demonstration.

**1. INTRODUCTION.** At the August, 2009 world track and field competitions in Berlin, Usain Bolt, the Jamaican sprinting sensation put on some amazing performances, shattering his own records in both the 100 and 200 meter lowering both by 0.11 seconds to an amazing 9.58 seconds in the 100 meter and 19.19 in the 200 meter. The man is certainly on another level.

His time is the greatest improvement in the 100 meter record since electronic timing began in 1968. Bolt is not done yet and who knows how fast he can run. In fact, he thinks he can do better. He said this after winning the 100 meter title at the same Olympic Stadium where Jesse Owens won four gold medals at the 1936 Berlin Olympic Games.

Several researchers have done studies to predict how fast a man can run 100 meters. Most models are based on the idea that we are getting close to the limit because the amount of improvements are decreasing. Bolt's results changed this perception.

Bolt, who has set three records at the 100 meter distance with times of 9.72, 9.69 and 9.58, is already looking to go far below some estimates. He thinks the world records will stop at 9.40. Some bookmakers are betting that Bolt will get there. The method for estimating the ultimate record presented in this article gives a 90% confidence interval for the ultimate record that has a lower bound of 9.40.

To measure the effect Bolt has had so far and may have in future we calculated the lower bound excluding his three records. The result is 9.62, greater than the record he set, so Bolt has already beaten the estimated ultimate record based on other runner's times. We have also carried out some other probability calculations by excluding his records. They show how remarkable Bolt's performance has been. He has become the premier runner in the world and has changed our perception of human capabilities.

In the following sections we will present methods we consider relevant for forecasting future records. Other methods can be found in [4], [6], [9], [10] and references therein.

**2. METHODS BASED ON TREND ANALYSIS.** Sports records have improved over the years, often faster than our expectation. To analyze the data many investigators have used models made up of a deterministic component, to account for the trend, and a stochastic component, to account for the random variation. Some attempts have also been made to estimate the limiting times (ultimate records) using models with exponential-decay deterministic component. Although useful it is demonstrated that trend analysis would not produce meaningful performance estimates for future records.

**3. METHODS BASED ON OUTSTANDING VALUES.** Outstanding sports achievements are usually analyzed using one of the following three main methods:

(1) The *Extreme Value Theory* that usually deals with the maxima or minima. This method uses the absolute largest or absolute smallest values in a specific time period.
(2) The *Threshold Theory* that deals with values above or below a specified threshold.
(3) The *Theory of Records* that deals with values larger or smaller than all the previous values.

The frequency of outstanding values is analyzed using *Theory of Exceedances.* This theory deals exclusively with the number of times a chosen threshold is exceeded.

**3.1. Methods Based on Threshold Theory.** In this approach the probabilities of future performances are calculated by developing models for a tail of the distribution for performance measures. Since performance measures above or below a threshold carry more information about the future exceptional performances, methods based on tail modeling are appropriate. Methods based on this approach assume that the tail of the distribution for the performance measure of interest belongs to a parametric family and carry out the inference using excesses, that is, the performance measures greater or smaller than some predetermined value $y_o$. It has been shown that the natural parametric family of distributions to consider for excesses is the generalized Pareto distribution (*GPD)*;

$$P(Y \le y) = 1 - (1 - \frac{ky}{\sigma})^{1/k}$$

where *Y* represents a performance measure and $\sigma > 0$ and $-\infty < k < \infty$ are unknown parameters [7]. The range of *Y* is $0 < y < \infty$ for $k \le 0$, and $0 < y < \sigma/k$ for $k > 0$.

For example, for the men's 100 meter dash a threshold such as 10 seconds may be considered. The GPD includes three specific forms:

1. Long tail Pareto distribution.
2. Medium tail exponential distribution.
3. Short tail distribution with an endpoint.

Most classical distributions have tails that behave like one of these forms. Unfortunately, like most asymptotic results, applying this approach is not free of problems. The obvious problems are the choice of a parametric family, determination of the threshold, and having to deal with intractable equations that need to be solved to obtain estimates of the parameters.

Hill [5] and Davis and Resnick [2] have proposed an approach that is easy to use and is applicable to a wide class of distribution functions possessing medium or long tails. They propose assuming a tail model of the form $F(y) = cy^{-a}$ , $y > y_0$. It is a suitable model for men's 100 meter data considering that the recent records set by Bolt indicate a long tail.

From a random sample $Y_1, Y_2, \ldots, Y_n$ the estimates of the parameters are obtained based on the upper *m = m(n)* order statistics (best times) where *m* is a sequence of integers chosen such that $m \diamond \times$ and $m/n \diamond 0$. Here *c* is estimated using the empirical $1 - m/n$ quantile, $Y_{(m+1)}$ and $1/a$ is estimated by;

$$\hat{a}*(n/\mathrm{m}) = m^{-1} \sum_{i=1}^{m} \ln Y_{(i)} - \ln Y_{(m+1)} .$$

Statistical theory regarding these estimators is well established. The Pareto-tail upper tail estimate $\overline{F}(y)$ is then

$$\overline{F}(y) = \frac{m}{n} \left( \frac{y}{Y_{(m+1)}} \right)^{-1/\hat{a}*(n/\mathrm{m})} \qquad y > Y_{(m+1)}.$$

The lower tail estimate can be obtained similarly.

**Application to the Men's 100 Meter Dash.** Consider the data for the men's 100 meter after January 1, 1977, when the International Association of Athletics Federations (IAAF) required fully automatic timing to the hundredth of a second. From January 1, 1977 to September 1, 2009 there are between 30 to 40 data points, some being invalidated due to drug use and some being discarded (including Gay's 9.68 time on June 2008 during the 2008 U.S. Olympic Trials) due to wind speeds exceeding the IAAF legal limit. Here the time between the two latest records set by Bolt and the time his recent record has stood to date are inordinately small. So rather than using an *m(n)* that involves these values here we use $\sqrt{n}$ instead. This is a good choice as the rate of convergence is the same for both requirements. For the data considered, the closest integer to $\sqrt{n}$ is 6 (*m(n)* =6). The top six legal times are

$$y_1 = 9.58,\ y_2 = 9.69,\ y_3 = 9.71,\ y_4 = 9.72,\ y_5 = 9.72,\ y_6 = 9.74.$$

set respectively by Bolt, Bolt, Gay, Powell, Bolt, and Powell. Using these for $y < y_7 = 9.77$ we get the tail model

$$\overline{F}(y) = 6/33 (9.77/y)^{-126.694} \ldots .$$

Using this P($Y \leq 9.5$) and P($Y \leq 9.55$) are respectively 0.0052 and 0.0102.

**Bolt's Effect.** To measure the effect Bolt has had so far and may have in the future let us calculate the same probabilities excluding his three records. The top six legal times are then

$$y_1 = 9.71,\ y_2 = 9.72,\ y_3 = 9.74,\ y_4 = 9.77,\ y_5 = 9.79,\ y_6 = 9.84.$$

set respectively by Gay, Powell, Powell, Powell, Greene, and Bailey. Using these $y < y_7 = 9.84$ we get the tail model

$$\overline{F}(y) = 6/33(9.84/y)^{-124.954} .$$

The values of $P(X \le 9.5)$ and $P(X \le 9.55)$ are now respectively 0.00225 and 0.00433. These are significantly lower than the probabilities we get using Bolt's records. Also we have $P(X \le 9.58) = 0.0064$, which demonstrates that Bolt is in a different league.

**3.2. Methods Based on Theory of Records.** The theory of records deals with values that are strictly greater than or less than all previous values. Usually the first value is counted as a record. Then a value is a record (upper record or record high) if it exceeds all previous values.

The study of record values, their frequencies, times of their occurrences, their distances from each other, etc. constitutes the theory of records. Formally the theory deals with four main random variables:

(1) The number of records in a sequence of *n* observations.

(2) The record times (indices).

(3) The waiting time between the records.

(4) The record values.

The theory of records has not yet been fully exploited to address questions regarding the prediction of sports records because the results of the theory of records for independent and identically distributed sequences are not directly applicable to sports. In fact, in most sports records occur more frequently than what the theory predicts. To account for this one may treat the problem as if participation has increased with time, or more attempts are made so that the probability of setting a new record was increased[6].

To predict the future records, Noubary [6] has developed a simple method using the following results of the theory of records for an independent and identically distributed sequence of observations:

(a) If there is an initial sequence of $n_1$ observations and a batch of $n_2$ future observations, then the probability that the additional batch contains a new record is $n_2/(n_1+n_2)$.

(b) For large *n*, $P_{r,n}$, the probability that a series of length *n* contains exactly *r* records is given by

$$P_{r,n} \sim \frac{1}{(r-1)!\,n}(ln(n)+\gamma)^{r-1}$$

where $\gamma = 0.5772$ is Euler's constant. To illustrate the method, consider the 100 meter data for the period 1912 – 2009. The number of records set is $r$ = 20. The maximum likelihood method yields $n$ = 317,884,920 attempts. This is found by maximizing $P_{r,n}$ with respect to $n$. It estimates the number of independent and identically distributed attempts required to produce twenty records.

Now suppose, for example that, $i$ is the geometric rate of increase per year due to increase in the number of attempts. This means that the number of attempts in any given year is assumed to be $i$ times the number of attempts in the year before. The value of $i$ can be found by solving the equation

$$1 + i + i^2 + \cdots + i^{97} = 317884920,$$

where $i^j$ represents the number of attempts in year $j$. We get $i$ = 1.2013, which means 20.13% more attempts per year. To predict records for the future one and five years (in this case year 2010 and the period 2010 – 2015) we replace $n_1$ = 317884920 and $n_2$ = $(1.2013)^{98}$ = 63950619 for the next year, and $n_1$ = 317884920 and $n_2$ = $(1.2013)^{98} + \ldots + (1.2013)^{102}$ = 477112800 for the next 5 years in *(a)*. This leads to probability estimates of 0.1926 and 0.6001 for a new record in the next one and five years, respectively.

## 4. ULTIMATE RECORD.

Prediction of ultimate records can be carried out using a model with an exponential-decay trend. It is also possible to consider simpler trends such as

$$y = (b + ct)/(1 + t) \quad \text{or} \quad y = (a + bt + ct^2)/(1 + t + t^2)$$

and use the fact that as $t$ ◇ ×, $y$ ◇ c. However, the application of such models usually results in predictions with large standard errors that are not useful or even acceptable[8]. This section describes a method based on tail modeling using a certain number of exceptional performances. This avoids the above mentioned problem and provides a confidence interval for the ultimate record based on the most recent performances or records.

Let $Y$ and $Y_1, Y_2, \ldots, Y_n$ represent, respectively, a performance measure and a set of observed values for a given sport such as 100 meter dash where

$$Y_1 \le Y_2 \le \ldots \le Y_n .$$

Assuming that the distribution function *F(y)* has a lower endpoint and certain conditions are satisfied, a level (1 – $p$) confidence interval for the minimum of $Y$ is (De Haan [3])

$$\{Y_1 - (Y_2 - Y_1)/[(1-p)^{-k} - 1],\ Y_1\} \ .$$

De Haan [3] has also shown that

$$\frac{\ln m(n)}{\ln[(y_{m(n)} - y_3)/(y_3 - y_2)]}$$

is a good estimate for *k*.

For example, suppose that we wish to estimate the ultimate record for the 100 meter dash. We consider the same data as in Section 3.1. As in Section 3.1, we need to choose an integer *m (n)* depending on *n* such that $m(n) \to \infty$ and $m(n)/n \to 0$ as $n \to \infty$. Again we choose $\sqrt{n}$ as it has the same rate of convergence for both requirements. For the data used the closest integer to $\sqrt{n}$ is 6 (*m (n)* =6) and the top six legal times are:

$$y_1 = 9.58,\ y_2 = 9.69,\ y_3 = 9.71,\ y_4 = 9.72,\ y_5 = 9.72,\ y_6 = 9.74.$$

Using the above formula we get *k* = 4.419 and a 90% one-sided prediction interval for the ultimate record is (9.40, 9.58).

As mentioned earlier Bolt believes that the world record will not go below 9.40 and the British bookmakers are betting that Bolt will get there.

**Bolt's Effect.** Now to measure the effect Bolt has had so far and may have in future let us calculate the prediction interval by excluding his three records. The top six legal times are then

$$y_1 = 9.71,\ y_2 = 9.72,\ y_3 = 9.74,\ y_4 = 9.77,\ y_5 = 9.79,\ y_6 = 9.84.$$

set respectively by Gay, Powell, Powell, Powell, Greene, and Bailey. Using the formula we get *k* = 1.113 and a 90% one-sided prediction interval for the ultimate record is (9.62, 9.71). The lower bound is greater than the present record set by Bolt, so he has already exceeded what would have been predicted to be the limit of human performance. This shows how remarkable Bolt's performance has been and how this one runner can change our perception of human capabilities.

**ACKNOWLEDGMENTS**. I would like to thank Professor Joe Gallian for providing me with help and direction.

REFERENCES

1. S.M. Berry, A statistician reads the sports pages, *Chance.* **15** (2002) 49-53.
2. R. Davis, and S. Resnick, Tail estimates motivated by extreme-value theory. *Annals of Statistics.* **12** (1984) 1467-1487.
3. L. DeHaan, Estimation of the minimum of a function using order statistics. *Journal of the American Statistical Association.* **76** (1981) 467- 469.
4. J.H.J. Einmahl, and J.R. Magnus, Records in athletics through extreme-value theory, *Journal of the American Statistical Association.* **103** (2008) 484-494.
5. B. M. Hill, A Simple general approach to inference about the tail of a distribution,

*Annals of Statistics.* **3** (1975) 1163-1174.
6. R. Noubary, A procedure for prediction of sports records, *J. of Quantitative Analysis in Sports*. **1** (2005) 1-12.
7. J. Pickands, Statistical inference using extreme order statistics, *Annals of Statistics.* **3** (1975) 131-199.
8. R. L. Smith, Forecasting records by maximum likelihood, *J. of the Amer. Stat. Assoc*. **83** (1988) 331-338.
9. A. R. Solow and W. Smith, How surprising is a new record, *The Amer. Statistician*. **59** (2005) 153-155.
10. J. T. Terpstra and N. D. Schauer, A simple random walk model for predicting track and field World records, *J. of Quantitative Analysis in Sports*. **3** (2007) 1-16.

REZA D.NOUBARY received his B.Sc. and M.Sc. in Mathematics from Tehran University, and M.Sc. and Ph.D. in statistics from Manchester University. His research interests include time series analysis, modeling and risk analysis of natural disasters, and applications of mathematics and statistics in sports. He is a fellow of the Alexander von Humboldt Foundation. He frequently teaches courses based on sports. His outside interests include soccer, racquetball, and tennis.
Department of Mathematics, Bloomsburg University, Bloomsburg, PA 17815, USA
rnoubary@bloomu.edu.