

How far can Man go?

Isabel Fraga Alves, Laurens de Haan and Cláudia Neves

Abstract In this paper we address the question of “What is the Largest Jump at Man’s reach, given today’s state of the art?” To answer that question it will be used the *best from the best*, i.e. , the data will be collected from the best “jumpers” from World Athletics Competitions – “Long Jump Men Outdoors” event. Our approach to the problem is based on the probability theory of extreme values (EVT) and the corresponding statistical techniques. We shall only use the top performances of World top lists. Our estimated ultimate record, i.e. , the right endpoint of the jumping event, tells us what is possible to infer about the possible personal best mark, given today’s knowledge, sports conditions and rules.

Key words: Extreme values in sports, extreme value theory, endpoint estimation in Gumbel domain of attraction, semi-parametric approach to statistics of extremes.

1 Introduction

In this paper we address the question of

“What is the Largest Jump that Man can achieve, given today’s state of the art?”

For the purposes of our study we selected the following athletic jumping event:

- *Long Jump Men Outdoors (LJ).*

The best current mark is $8.95m$ from Mike Powell (USA) in Tokyo, 30/08/1991.

Isabel Fraga Alves

CEAUL, Department of Statistics, University of Lisbon, Portugal e-mail: isabel.alves@fc.ul.pt

Laurens de Haan

University of Tilburg, and Rotterdam, The Netherlands,

CEAUL, University of Lisbon Portugal e-mail: ldhaan@ese.eur.nl

Cláudia Neves

Department of Mathematics, University of Aveiro, Portugal e-mail: claudia.neves@ua.pt

Our estimated ultimate record, i.e. , the right endpoint of the jumping event, tells us what is possible to infer about the possible personal best mark, given today's knowledge, sports conditions and rules.

In [2], Einmahl and Magnus studied a similar problem, but LJ event could not be handled with the available methodology so far, not covering the estimation of the right endpoint of a distribution in Gumbel domain of attraction.

Related contributions for this type of athletics data are [5] and [3].

This contribution constitutes a first approach to the estimation of the right endpoint for a distribution in the Gumbel max-domain of attraction, a subject not yet addressed in the literature of extremes to our best knowledge. Asymptotic properties of endpoint estimator are under study.

The paper is organized as follows. In the following section an overview of the first and second order properties of the tail distribution and their relation to the right endpoint; therein, a subsection is devoted to the special case of finite right endpoint and Gumbel max-domain of attraction. In Section 3 the motivation for the right endpoint estimator is presented, according to asymptotic theoretical relations of Section 2. In Section 4 exact properties of the estimator will be illustrated in Reversed Fréchet model. Finally, in Section 5 we apply the theory to the LJ-data.

2 EVT framework

Let $X_1, X_2, \dots, X_n$ be an independent, identically distributed (i.i.d.) sample from an unknown distribution function (d.f.) F . If there exist normalizing constants $a_n > 0$, $b_n \in \mathbb{R}$ and a non-degenerate d.f. G such that, for all x ,

$$\lim_{n \rightarrow \infty} P \{ a_n^{-1} (\max(X_1, \dots, X_n) - b_n) \leq x \} = G(x), \quad (1)$$

then G is, up to scale and location, an *Extreme Value* d.f., dependent on a shape parameter $\gamma \in \mathbb{R}$, and given by

$$G_\gamma(x) := \begin{cases} \exp(-(1 + \gamma x)^{-1/\gamma}), & 1 + \gamma x > 0 \text{ if } \gamma \neq 0 \\ \exp(-\exp(-x)), & x \in \mathbb{R} \text{ if } \gamma = 0 \end{cases}. \quad (2)$$

The EV condition (1) is equivalent to say that F is in the max-domain of attraction of the d.f. G_γ , i.e. , $F \in \mathcal{D}_M(G_\gamma)$. For $\gamma < 0$, $\gamma = 0$ and $\gamma > 0$, the G_γ d.f. reduces to Weibull, Gumbel and Fréchet distribution functions, respectively.

Fréchet, Weibull and Gumbel max-Domains of Attraction The shape parameter γ is closely related to the tail heaviness of the underlying distribution F : $\gamma < 0$ refers to short tails which must have finite right endpoint $x^F := \sup\{x : F(x) < 1\}$, whereas for $\gamma > 0$ the d.f. F is heavy tailed with polynomially decaying tails, for which the right endpoint must be infinite. The intermediate case $\gamma = 0$, the Gumbel domain, has been revealed very interesting for many applied sciences where extremes play the major role, and include a great variety of distributions with finite right endpoint

or not. Then it is assumed that the extreme value index γ of the underlying d.f. equals 0, and statistical inference procedures concerning rare events on the tail of F , such as the estimation of high quantiles, small exceedance probabilities or return periods, bear on this assumption. The latter has been naturally imposed by the lack of attention to the estimation of the right endpoint x^F in the literature, under this Gumbel max-domain setup.

First order properties for U Consider the left-continuous inverse of $1/(1-F)$, defined by

$$U(t) := F^{\leftarrow}(1 - 1/t), \text{ for, } t > 1, \quad (3)$$

where $F^{\leftarrow}$ denotes the generalized inverse function $F^{\leftarrow}(x) = \inf\{y : F(y) \geq x\}$. The following *extended regular variation* property (see [6], for instance), denoted ERV_γ , is a well-known necessary and sufficient condition for $F \in \mathcal{D}_M(G_\gamma)$:

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \begin{cases} \frac{x^\gamma - 1}{\gamma} & \text{if } \gamma \neq 0 \\ \ln x & \text{if } \gamma = 0 \end{cases}, \quad (4)$$

for every $x > 0$ and some positive measurable function a .

For the three max-domains of attraction, the following properties on the tail of the distribution hold:

$\mathcal{D}_M(G_\gamma)_{\gamma > 0}$ Heavy tails – Regularly Varying (RV) tails . No finite right endpoint, i.e. , $x^F := U(\infty) = \infty$.
 $U \in RV_\gamma$, i.e. , $\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma$, for $x > 0$, with $\lim_{t \rightarrow \infty} \frac{a(t)}{U(t)} = \gamma$.

$\mathcal{D}_M(G_\gamma)_{\gamma < 0}$ Short Tails – Finite right endpoint, i.e. $x^F := U(\infty) < \infty$.
 $U(\infty) - U \in RV_\gamma$, i.e. $\lim_{t \rightarrow \infty} \frac{U(\infty) - U(tx)}{U(\infty) - U(t)} = x^\gamma$, for $x > 0$, with
 $\lim_{t \rightarrow \infty} \frac{a(t)}{U(\infty) - U(t)} = -\gamma$.

$\mathcal{D}_M(G_\gamma)_{\gamma = 0}$ Light Tails – $U \in \Pi$ -class, i.e. , $\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \ln x$, for positive $a \in RV_0$. The right endpoint $x^F := U(\infty)$ is finite or infinite.
 $U \in RV_0$, i.e. , $\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = 1$, with $\lim_{t \rightarrow \infty} \frac{a(t)}{U(t)} = 0$, for $x > 0$.
If the right endpoint is finite, i.e. $x^F := U(\infty) < \infty$, then
 $\lim_{t \rightarrow \infty} \frac{U(\infty) - U(tx)}{U(\infty) - U(t)} = 0$, $\lim_{t \rightarrow \infty} \frac{a(t)}{U(\infty) - U(t)} = 0$, with $\lim_{t \rightarrow \infty} a(t) = 0$.

Second order properties for U The rate of convergence in (4) is *controlled* by the existence of a *second order* parameter $\rho \leq 0$ and a function A , possibly not changing in sign and tending to zero as $t \rightarrow \infty$, with $|A| \in RV_\rho$, and such that for all $x > 0$

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = H_{\gamma, \rho}(x) := \frac{1}{\rho} \left(\frac{x^{\gamma+\rho} - 1}{\gamma + \rho} - \frac{x^\gamma - 1}{\gamma} \right). \quad (5)$$

We then say that the function U is of *second order extended regular variation*, and use the notation $U \in 2ERV_{\gamma, \rho}$. In (5), the cases $\gamma = 0$ and/or $\rho = 0$ are obtained by continuity arguments.

2.1 Gumbel Domain of Attraction and finite right endpoint

In the particular case of $F \in \mathcal{D}_M(G_0)$ with a finite right endpoint x^F , then $\rho = 0$ (see Theorem B.3.6 in [6]); consequently, (5) becomes $H_{\gamma, \rho}(x) = H_{0,0}(x) = \frac{\ln^2 x}{2}$.

Moreover, in case $\rho = \gamma = 0$, there exists a function a_0 asymptotically equivalent to a , i.e., $a_0(t) \sim a(t)$, as $t \rightarrow \infty$, such that (see [6], pages 382-3)

$$U(t) = a_0(t) + \int_0^t \frac{a_0(s)}{s} ds, \text{ with } a_0 \in \Pi. \quad (6)$$

Consequently, by (6)

$$U(\infty) - U(t) \sim \int_t^\infty \frac{a(s)}{s} ds, \text{ as } t \rightarrow \infty,$$

which is slowly varying, since $U(\infty) - U(t) \in RV_0$. Then

$$U(\infty) = U(t) + \int_t^\infty \frac{a(s)}{s} ds + o(a(t)), \text{ as } t \rightarrow \infty,$$

and the following asymptotical representation holds for the right endpoint:

$$U(\infty) \sim U(t) + \tilde{a}(t), \text{ as } t \rightarrow \infty, \quad (7)$$

where the positive function $\tilde{a}$ is defined by

$$\begin{aligned} \tilde{a}(t) &:= \int_t^\infty \frac{a(s)}{s} ds \\ &= \int_1^\infty \frac{a(ts)}{s} ds \\ &= \int_0^1 \frac{a(t/s)}{s} ds. \end{aligned}$$

Note that $\lim_{t \rightarrow \infty} \tilde{a}(t) = 0$ and $a(t) = o(\tilde{a}(t))$, as $t \rightarrow \infty$. Moreover, for $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{\int_t^\infty (U(sx) - U(s)) \frac{ds}{s}}{\int_t^\infty a(s) \frac{ds}{s}} = \lim_{t \rightarrow \infty} \frac{(U(tx) - U(t))/t}{a(t)/t} = \ln x,$$

which makes possible to express $\tilde{a}(t)$ in terms of the quantile function U , which will be revealed useful for estimation purposes.

This means that, as $t \rightarrow \infty$ and for $x > 0$,

$$\tilde{a}(t) \sim \frac{1}{\ln x} \int_0^1 \left(U\left(\frac{t}{s}x\right) - U\left(\frac{t}{s}\right) \right) \frac{ds}{s}. \quad (8)$$

Example – Reversed Fréchet distribution

The Reversed Fréchet model is defined for $x < x^F := U(\infty) \in \mathbb{R}$; for scale and shape parameters $\delta, \alpha > 0$ the d.f. F and f.d.p f , are given respectively by

$$\begin{aligned} F(x) &\equiv F(x; x^F, \alpha, \delta) = 1 - \exp \left\{ - \left(\frac{x^F - x}{\delta} \right)^{-\alpha} \right\}, \\ f(x) &\equiv f(x; x^F, \alpha, \delta) = \frac{\alpha}{\delta} \left(\frac{x^F - x}{\delta} \right)^{-\alpha-1} \exp \left\{ - \left(\frac{x^F - x}{\delta} \right)^{-\alpha} \right\}. \end{aligned} \quad (9)$$

The following results are easily checked:

- $F^{\leftarrow}(x) = x^F - \delta [-\ln(1-x)]^{-1/\alpha}$, for $x \in (0, 1)$;
- $U(t) = x^F - \delta (\ln t)^{-1/\alpha}$, for $t > 1$;
- $U \in \Pi(a)$, with $a(t) = \frac{\delta}{\alpha} (\ln t)^{-1/\alpha-1}$, $t > 1$;
- $a \in RV_0$, $\lim_{t \rightarrow \infty} a(t) = 0$ and $\lim_{t \rightarrow \infty} \frac{a(t)}{U(\infty) - U(t)} = 0$.

We conclude that $F \in \mathcal{D}_M(G_0)$, i.e., the Reversed Fréchet d.f. belongs to Gumbel max-domain of attraction with finite right endpoint x^F .

Moreover, the function $\tilde{a}$ can be chosen as $\tilde{a}(t) = \delta (\ln t)^{-1/\alpha}$, $t > 1$, and the equality $x^F = U(t) + \tilde{a}(t)$ holds.

3 Estimation of the right endpoint in Gumbel Domain

Consider a random sample $X_1, X_2, \dots, X_n$ of i.i.d. random variables with common d.f. $F \in \mathcal{D}_M(G_0)$, such that the right endpoint x^F is finite. Let $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ be the associated ascending order statistics (o.s.). The following distributional identification will be the starting point for the estimation development:

$$\{X_{i,n}\}_{i=1}^n \stackrel{d}{=} \{U(Y_{i,n})\}_{i=1}^n, \quad (10)$$

denoting by $\{Y_{i,n}\}_{i=1}^n$ the o.s. pertaining to the independent random variables (r.v.) $\{Y_i\}_{i=1}^n$ with common standard Pareto d.f. $F(y) = 1 - y^{-1}$, $y \geq 1$.

Let $k = k_n$ be an intermediate sequence, i.e., a sequence of positive integers such that

$$k_n \rightarrow \infty \quad \text{with} \quad k_n = o(n), \quad \text{as } n \rightarrow \infty. \quad (11)$$

Then, as the sample size $n \rightarrow \infty$, the intermediate sequence $Y_{n-k,n}$ converges in probability to infinity with the same rate as the ratio n/k , i.e., $(k/n)Y_{n-k,n} \xrightarrow[n \rightarrow \infty]{P} 1$, which is also equivalently denoted by the

$$Y_{n-k,n} \overset{P}{\sim} \frac{n}{k}, \quad \text{as } n \rightarrow \infty. \quad (12)$$

3.1 Motivation for the estimator

Consider at this step the asymptotical relation (7), with n/k playing the role of t , according to (11); the estimator of the right endpoint $x^F = U(\infty)$ will be motivated by (7), through its empirical counterpart, for convenient estimators of the components:

$$\widehat{U(\infty)} := \widehat{U\left(\frac{n}{k}\right)} + \widehat{\tilde{a}\left(\frac{n}{k}\right)}. \quad (13)$$

From representation (10) together with (12), the estimation of the first component is immediately suggested as

$$\widehat{U\left(\frac{n}{k}\right)} := X_{n-k,n}.$$

For the estimation of $\tilde{a}(n/k)$, the asymptotic result (8) will be considered for the particular value $x = 1/2$,

$$\tilde{a}\left(\frac{n}{k}\right) \sim \frac{1}{-\ln 2} \int_0^1 \left(U\left(\frac{n}{2ks}\right) - U\left(\frac{n}{ks}\right) \right) \frac{ds}{s}. \quad (14)$$

This suggests for the estimator of $\tilde{a}\left(\frac{n}{k}\right)$ the empirical counterpart of (14)

$$\widehat{\tilde{a}\left(\frac{n}{k}\right)} := \frac{1}{-\ln 2} \int_0^1 (X_{n-[2ks],n} - X_{n-[ks],n}) \frac{ds}{s}, \quad (15)$$

where $[z]$ denotes the integer part of z .

Since $X_{n-[ks],n}$ and $X_{n-[2ks],n}$ are identical to $X_{n,n}$ when $s \in (0, (2k)^{-1}]$, the integral $I_0 := \int_0^{1/(2k)} (X_{n-[2ks],n} - X_{n-[ks],n}) \frac{ds}{s} = 0$ and is possible to rewrite the integral in (15) as the sum of $2k - 1$ integrals

$$\sum_{i=1}^{2k-1} I_i := \sum_{i=1}^{2k-1} \int_{i/(2k)}^{(i+1)/(2k)} (X_{n-[2ks],n} - X_{n-[ks],n}) \frac{ds}{s}.$$

After integrating (15) all over $\left(\frac{i}{2k}, \frac{i+1}{2k}\right]$, $i = 1, \dots, 2k-1$ the estimator for $\tilde{a}\left(\frac{n}{k}\right)$ is finally obtained as

$$\begin{aligned} \widehat{\tilde{a}\left(\frac{n}{k}\right)} &= \frac{1}{\ln 2} \sum_{i=0}^{k-1} (X_{n-2i,n} - X_{n-2i-1,n}) \ln \left(1 + \frac{1}{2i+1}\right) \\ &+ \frac{1}{\ln 2} \sum_{i=1}^{k-1} (X_{n-i,n} - X_{n-2i,n}) \ln \left(1 + \frac{1}{i}\right), \end{aligned} \quad (16)$$

which together with (13) leads to the endpoint estimator

$$\begin{aligned} \hat{x}^F &:= X_{n-k,n} + \frac{1}{\ln 2} \sum_{i=0}^{k-1} (X_{n-2i,n} - X_{n-2i-1,n}) \ln \left(1 + \frac{1}{2i+1}\right) \\ &+ \frac{1}{\ln 2} \sum_{i=1}^{k-1} (X_{n-i,n} - X_{n-2i,n}) \ln \left(1 + \frac{1}{i}\right). \end{aligned} \quad (17)$$

Denoting the descending j -th order statistics by $\{X_{(j)} := X_{n-j+1,n}\}_{j=1}^n$,

$$\begin{aligned} \hat{x}^F &:= X_{(k+1)} + \frac{1}{\ln 2} \sum_{i=1}^k (X_{(2i-1)} - X_{(2i)}) \ln \left(1 + \frac{1}{2i-1}\right) \\ &+ \frac{1}{\ln 2} \sum_{i=1}^{k-1} (X_{(i+1)} - X_{(2i+1)}) \ln \left(1 + \frac{1}{i}\right), \end{aligned} \quad (18)$$

and note that $2k$ top observations are involved in the estimation of $\tilde{a}\left(\frac{n}{k}\right)$.

4 Simulations – Reversed Fréchet Model

In order to illustrate distributional properties of the estimators in (13) and (15), it was performed a simulation study for samples from the Reversed Fréchet distribution defined in (9); for the particular values of $\alpha = \delta = 0.5$ and $x^F = 10$, with 1000 runs of samples of size $n = 1000$, the exact properties of $\widehat{\tilde{a}\left(\frac{n}{k}\right)}$ and $\hat{x}^F$ defined in (16) and (17), respectively, are depicted in Fig.1 and Fig.2, against the top k order statistics involved, $k \leq 500$.

In the graphics the gray lines represent the empirical bandwidths for the mean value of the estimators depicted in black, together with the target true values, in red.

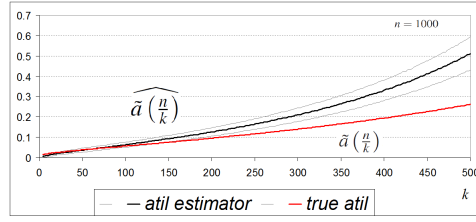


Fig. 1 Estimation of $\tilde{a}(\frac{n}{k})$: $N = 1000$ replicas with sample size $n = 1000$ from Reversed Fréchet d.f. $F(x; x^F, \alpha, \delta)$, $\alpha = \delta = 0.5$ and $x^F = 10$.

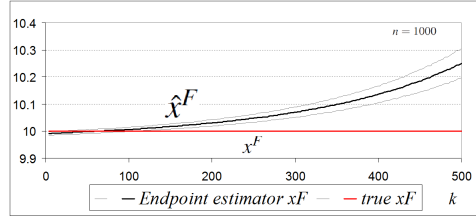


Fig. 2 Estimation of x^F : $N = 1000$ replicas with sample size $n = 1000$ from Reversed Fréchet d.f. $F(x; x^F, \alpha, \delta)$, $\alpha = \delta = 0.5$ and $x^F = 10$.

Note that since the asymptotic results require the intermediate condition (11) for k along with n , and for a sample size the n is fixed, the best behaviour is clearly for the top k observations, with k small, as it is expected beforehand.

5 Long Jump Men Outdoors – How far?

For the purposes of our study we selected the following athletic jumping event:

- *Long Jump Men Outdoors (LJ)*.

The best current mark is $8.95m$ from Mike Powell (USA) in Tokyo, 30/08/1991.

Our estimated ultimate record, i.e., the right endpoint of the jumping event, tells us what is possible to infer about the possible personal best mark, given today's knowledge, sports conditions and rules.

LJ-Data

For the LJ event the collected data has been of “the personal best” of as many of the top athletes in LJ event as it was possible. Each athlete appears only once in the data set, namely with his best own mark. Our observation period ends on December 31, 2009. The LJ data are obtained from two websites, namely a Swedish website compiled by Hans-Erik Pettersson, for the period up to August 2001 (source: <http://hem.bredband.net/athletics/atb-m27.htm>),

which has been merged with the data for the period 1999-2009, from the official website of the International Association of Athletics Federations (IAAF) (*source*: <http://www.iaaf.org/statistics/toplists/>). Similar to [2], only marks greater than $8m$ have been collected, with an obtained top list of 720 best personal marks.

These 720 personal bests are considered as i.i.d. observations X_i from some distribution function F . Let $X_1, X_2, \dots, X_n$ be an iid sample F and $X_{n-720+1,n} \leq \dots \leq X_{n-k,n} \leq X_{n-k+1,n} \leq \dots \leq X_{n,n}$ be the associated top retained 720 order statistics, so that $x_{n-720+1,n} = 8$ and $x_{n,n}$ denotes the world record of 8.95, obtained by Powell.

This athletics event has been previously detected by Einmahl and Magnus in [2], as a case where the Weibull Domain of attraction is questionable, since the estimated γ achieved was not clearly negative.

LJ-Data Analysis

Similar to the endpoint estimation the inference semi-parametric procedures considered in LJ-Data Analysis do not involve the explicitly the value n and depend on the number k of top o.s.'s.

The first LJ-Data Analysis involved the estimation of γ , which has been done according to three specific estimators, namely the Pickands, Moment and Mixed Moment estimators:

$$\hat{\gamma}^P := (\ln 2)^{-1} \ln \frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}}; \quad (19)$$

$$\hat{\gamma}^M := M^{(1)} + \frac{1}{2} \left\{ 1 - \left(\frac{M^{(2)}}{[M^{(1)}]^2} - 1 \right)^{-1} \right\}, \quad (20)$$

with $M^{(j)} := \frac{1}{k} \sum_{i=1}^k \{\ln X_{n-j+1,n} - \ln X_{n-k,n}\}^j$, $j = 1, 2$;

$$\hat{\gamma}^{MM} := \frac{\hat{\phi} - 1}{1 + 2 \min(\hat{\phi} - 1, 0)}, \quad (21)$$

with $\hat{\phi} := \frac{M^{(1)} - L^{(1)}}{(L^{(1)})^2}$, where $L^{(1)} := \frac{1}{k} \sum_{i=1}^k \{1 - X_{n-k,n}/X_{n-j+1,n}\}$.

For details about these estimation procedures see [11], [1] and [4].

In Fig.3 are depicted the sample paths of the three γ -estimators, $\hat{\gamma}^P$, $\hat{\gamma}^M$ and $\hat{\gamma}^{MM}$, associated with the LJ-Data against the same number k of o.s.'s involved in the estimation.

The second LJ-Data analysis consisted in performing the test for max-Gumbel domain of attraction, i.e. ,

$$H_0 : F \in \mathcal{D}_M(G_0) \quad \text{versus} \quad H_1 : F \in \mathcal{D}_M(G_\gamma)_{\gamma \neq 0}. \quad (22)$$

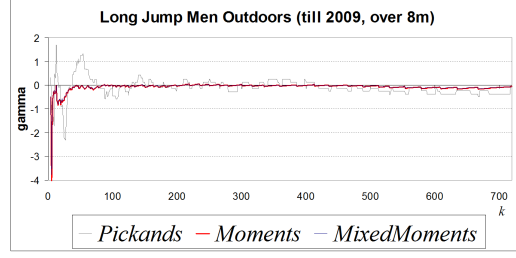


Fig. 3 Long Jump Men Outdoors (till 2009, over 8m) – γ –Estimation. Sample path of γ -estimators against $k \leq 720$, k the number of top o.s.'s involved.

For testing (22) three complementary tests statistics have been used: G (*Greenwood*), HW (*Hasofer&Wang*), and R (*Ratio*), which have been studied under semi-parametric approach in [8] and [7]:

$$G := \frac{M_2}{(M_1)^2} \quad HW := \frac{1}{k(G-1)} \quad R := \frac{X_{n,n} - X_{n-k,n}}{M_1}, \quad (23)$$

with $M_j := \frac{1}{k} \sum_{i=1}^k (X_{n-i+1,n} - X_{n-k,n})^j$, $j = 1, 2$. Under the null hypothesis, the standardized versions of (23), $G^* := \sqrt{k/4}(G-2)$, $HW^* := \sqrt{k/4}(kHW-1)$ have asymptotic Standard Normal distribution, whereas $R^* := R - \ln k$ converges to the Gumbel distribution. An overview about EV-testing is available in [9]. In Fig.4 are pictured the sample paths of the three test statistics G^* , HW^* and R^* associated with the LJ-Data, together with the respective Normal and Gumbel (0.05,0.95)-quantiles.

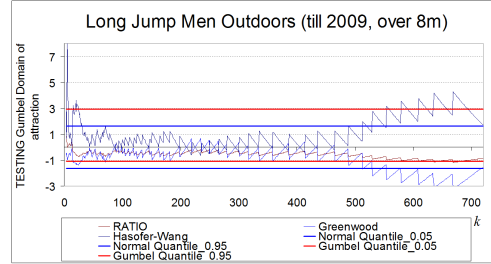


Fig. 4 Long Jump Men Outdoors (till 2009, over 8m) – $\mathcal{D}_M(G_0)$ – Testing. Sample path of the Tests Statistics against $k \leq 720$, k the number of top o.s.'s involved.

Altogether, from Fig.3 and Fig.4, it is seems reasonable to conclude that the distribution underlying the LJ-Data belongs to the Gumbel domain of attraction and not to the Weibull domain, which confirms the guess predicted in [2].

The third LJ-Data analysis has been testing the finiteness of the endpoint, in the same sense of Neves and Pereira in [10], i.e. ,

$$H_0 : F \in \mathcal{D}_M(G_0), x^F = \infty \quad \text{versus} \quad H_1 : F \in \mathcal{D}_M(G_\gamma)_{\gamma \leq 0}, x^F < \infty. \quad (24)$$

through the test statistic

$$T_1 := \frac{1}{k} \sum_{i=1}^k \frac{X_{n-i,n} - X_{n-k,n} - T}{X_{n,n} - X_{n-k,n}}, \quad \text{with } T := X_{n-k,n} \frac{M^{(1)}}{2} \left(1 - \frac{[M^{(1)}]^2}{M^{(2)}} \right)^{-1}. \quad (25)$$

Under the null hypothesis, the standardized version of (25), $T_1^* := \sqrt{k} \ln k T_1$, has asymptotic Standard Normal distribution. In Fig.5 is pictured the sample path of the test statistic T_1^* associated with the LJ-Data, together with the respective Normal (0.05,0.95)-quantiles.

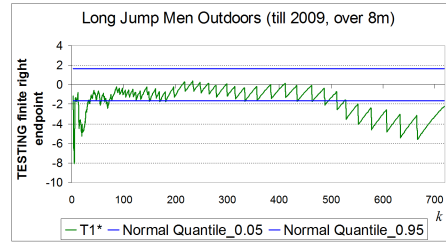


Fig. 5 Long Jump Men Outdoors (till 2009, over 8m) – Testing finiteness of right endpoint. Sample path of the Test Statistic T_1^* against $k \leq 720$, k the number of top o.s.'s involved.

From Fig.5 it is not so clear that the right endpoint is finite. This is not unexpected as T_1^* -test is conservative, as remarked in [10]. However, it yields a sample path near the lower the critical point, the normal 0.05-quantile, for several values of k lying in the most stable region of the graph that can be taken as intermediate. Moreover, it seems reasonable to consider the finiteness of the ultimate record in Long Jump Men outdoors event, by the nature of the event itself, since clearly Man has a finite limit length to attain in Jumping sports.

Finally, in Fig.6 are depicted the estimates of the right endpoint of the distribution underlying the LJ-data, against k , the number of observations required in the estimation. Note that in Fig. 6 the sample path of endpoint estimator touches insistently the $9m$.

From the LJ-Analysis we conclude that there is some space to improve the current Powell's record of $8.95m$, who also holds the longest *non-legal jump* of $8.99m$ set in 1992. In fact, from this preliminary study it seems possible to attain an ultimate record of $9m$ for the Long Jump Men outdoors.

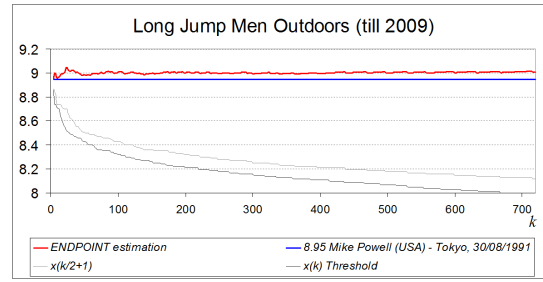


Fig. 6 Long Jump Men Outdoors (till 2009, over 8m) – $\hat{x}^F$ – Right Endpoint Estimation, i.e. "ultimate record" (different thresholds, $X_{(k/2+1)}$ and $X_{(k)}$ for estimating the first and second components in (13)).

Acknowledgements Research partially financed by FCT/POCTI, POCI2010 and PPCDT/FEDER.

References

1. Dekkers, A. L. M., Einmahl, J. H. J., deHaan, L.: A moment estimator for an extreme-value distribution. *Ann. Stat.* **17**, No.4, 1833–1855 (1989).
2. Einmahl, J. and Magnus, J.: Records in athletics through extreme-value theory. *J. Am. Stat. Assoc.*, **103**, 1382–1391 (2008)
3. Einmahl, J.H.J. and Smeets, S.: Ultimate 100M World Records through Extreme-Value Theory (July 10, 2009). Center Discussion Paper Series No. 2009-57. Available at SSRN:<http://ssrn.com/abstract=1433242>
4. Fraga Alves, M. I., Gomes, M. I., de Haan, L. and Neves, C.: Mixed moment estimator and location invariant alternatives. *Extremes* **12**, 149–185. doi:10.1007/s10687-008-0073-3.(2009)
5. Gomes, M.I., Pestana, D. and Henriques Rodrigues, L. : Athletic Events and Statistics of Extremes: Estimation of Useful Parameters. Technical Report CEAUL 06/09 (2009) <http://www.ceaul.fc.ul.pt/getfile.asp?where=notas&id=250>
6. de Haan, L. and Ferreira, A.: Extreme Value Theory: An Introduction. Springer Series in Operations Research and Financial Engineering. Springer, New York. (2006)
7. Neves, C., Picek, J. and Fraga Alves, M. I: Contribution of the maximum to the sum of excesses for testing max-domains of attraction. *J. Stat. Plan. Infer.*, **136**, Issue 4,1281–1301. doi:10.1016/j.jspi.2004.09.008 (2006)
8. Neves, C. and Fraga Alves, M. I: Cláudia Neves and M. I. Fraga Alves . Semi-parametric Approach to Hasofer-Wang and Greenwood Statistics in Extremes. *TEST*,**16**, 297–313. doi:10.1007/s11749-006-0010-1 (2008)
9. Neves, C. and Fraga Alves, M. I: Cláudia Neves and M. I. Fraga Alves . Testing extreme value conditions - an overview and recent approaches. *REVSTAT - Stat. Journal*, Special issue on "Statistics of Extremes and Related Fields" Beirlant, J. Fraga Alves, M.I. and Leadbetter, R. editors. **6**, Issue 1, 83–100.(2008)
10. Neves, C. and Pereira, A.: Detecting finiteness in the right endpoint of light-tailed distributions. *Stat. Probabil. Lett.*,**80**, Issue 5-6, 437–444. (2010)
11. Pickands III, J : Statistical inference using extreme order statistics. *Ann., Statist.*, **3**, 119-131. (1975)