Estimating the end-point of a probability distribution using minimum-distance methods

PETER HALL[1] and JULIAN Z. WANG[1,2]

[1]*Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia*
[2]*Department of Mathematics, University of Western Sydney (Nepean), PO Box 10, Kingswood, New South Wales 2747, Australia*

A technique based on minimum distance, derived from a coefficient of determination and representable in terms of Greenwood's statistic, is used to derive an estimator of the end-point of a distribution. It is appropriate in cases where the actual sample size is very large and perhaps unknown. The minimum-distance estimator is compared with a competitor based on maximum likelihood and shown to enjoy lower asymptotic variance for a range of values of the extremal exponent. When only a small number of extremes is available, it is well defined much more frequently than the maximum-likelihood estimator. The minimum-distance method allows exact interval estimation, since the version of Greenwood's statistic on which it is based does not depend on nuisance parameters.

Keywords: central limit theorem; coefficient of determination; domain of attraction; extreme value theory; goodness of fit; Greenwood's statistic; least-squares maximum-likelihood order statistic; Pareto distribution; sporting records; Weibull distribution

1. Introduction

In some very large data sets, only extreme values are readily accessible. For example, in the case of data on athletics or similar sporting events, it can be the case that only the fastest times, in major meetings such as the Olympic Games, are straightforward to obtain. Likewise, with machine-recorded data only the k largest extreme values may be retained, in order to minimize the amount of data storage. There, the recorded data are updated if and only if a new record is observed. Meteorological data recorded automatically in remote locations are sometimes of this type.

In such cases it is often of interest to estimate the end-point or threshold of the distribution from which the data came, e.g. the fastest possible time in an athletic event, or the highest possible wind velocity in the case of meteorological data. Although the number n of potential trials is usually unknown, it would be very large. Hence, if the sampling distribution were in the domain of attraction of an extreme value type, then the joint

distribution of the k largest extremes would probably be very close to its limit as $n \to \infty$. The most widely applicable extreme value type, in the context of approach to a threshold, is arguably the Weibull.

In the present paper we discuss two approaches to inference in this context. The first, based on maximum likelihood, is not infrequently outperformed by the second, which is founded on minimum-distance methods. We use a measure of distance related to a quadratic form for order statistics, proposed by Hartley and Pfaffenberger (1972). It is based on Greenwood's statistic, which measures the distance of an empirical distribution from uniformity, and is more commonly used for spacings-based tests of the hypothesis that a sampling distribution is uniform. The minimum-distance estimator of the end-point enjoys lower asymptotic variance, for a range of values of the extremal exponent, than its counterpart derived using maximum likelihood. In addition, in small samples it is well defined much more often than the maximum-likelihood estimator.

If data are recorded over an extended period, and one wishes to avoid having to model the manner in which the sampling distribution evolves in time, then it can be necessary to restrict attention to relatively short sequences of extremes. For example, in the case of times in a sporting event it is to be expected that improvements in training methods or changes in the way rules are enforced will produce changes in distribution. This means that the fastest possible time in one period may be rather different from that in the next and implies that a relatively small value of k should be employed. In this context the minimum-distance estimator is well defined (in the sense of taking a finite value) more often than its competitor based on maximum likelihood.

The poorer performance of the maximum-likelihood estimator, relative to the minimum-distance estimator, derives from the greater reliance of the former on estimation of the extremal exponent. At one point in the derivation of the maximum-likelihood estimator, the extremal exponent is replaced by its maximum-likelihood estimate, given the end-point. However, the maximum-likelihood estimate of the extremal exponent has much larger variance, for large values of the exponent, when the end-point is unknown than it does when the end-point is known. By way of comparison, the minimum-distance estimator is constructed using an argument which is invariant under alterations to the extremal exponent. This is reflected in the fact that the distribution of the minimum distance criterion ($G(\theta)$ in Section 2.2) does not depend on the extremal exponent, whereas the distribution of the likelihood criterion ($l(\theta)$ in Section 2.1) depends on the exponent.

Early work on estimation of an end-point includes that of Peto and Lee (1973), who treated the problem in the context of fitting a Weibull distribution. Weissman (1978) conducted inference based on the asymptotic joint distribution of the k largest order statistics; Weissman (1981) also addressed interval estimation when the shape parameter is unknown. In our work the shape is also unknown and is regarded as a nuisance parameter. Cooke (1979, 1980) proposed estimators based on linear combinations of a fixed number of extreme values. Maximum-likelihood methods in related settings, where the effective sample size is generally assumed known, go back at least to the work of Hall (1982) and Smith (1985, 1987). Smith and Weissman (1985) suggested a local maximum-likelihood estimator, and discussed the existence of solutions of the likelihood equations. Related extreme value methods have been suggested by de Haan and Resnick (1994) for estimating the "home

range" of an animal. Plotting methods for extreme value data, among other matters, have been discussed in a book by Embrechts *et al*. (1997).

Section 2 introduces maximum-likelihood and minimum-distance methods and describes their theoretical properties. Numerical performance is studied in Section 3, in the case of both real and simulated data. We show there that minimum-distance methods lead to confidence intervals as well as point estimates, with exactly known coverage when the data enjoy an exact extreme value distribution. Technical arguments behind the work in Section 2 are outlined in Section 4. We address only the case of a lower bound to the sampling distribution, since that of an upper bound is readily treated by making the obvious transformation.

2. Methods for estimation

2.1. Maximum-likelihood estimation

Let $X_1 < \ldots < X_k$ have an extreme value distribution of Weibull type, referred to as "type II" in the notation of Galambos (1978) and David (1981), and "type III" by Smith (1985). Denote the lower limit of this distribution by θ. Then, there exist positive constants α and a such that $(X_1 - \theta)/a$ has the distribution function $1 - \exp(-x^\alpha)$ for $x \geq 0$.

We may represent the joint distribution of $\mathcal{X} = (X_1, \ldots, X_k)$ in terms of independent random variables, as follows. Put

$$Y_i = i\{\log(X_{i+1} - \theta) - \log(X_i - \theta)\}, \qquad Y = \log(X_k - \theta) - \log a. \qquad (2.1)$$

If θ is at its true value, then $Y_1, \ldots, Y_{k-1}, Y$ are independent, each αY_i has an exponential distribution with unit mean, and αY has the density function

$$f(y) = \{(k-1)!\}^{-1} \exp(ky - \exp y).$$

The likelihood L of $\mathcal{X}$ equals the product over $1 \leq i \leq k-1$ of the density of X_i given $X_{i+1}, \ldots, X_k$, multiplied by the density of X_k. Therefore,

$$L = \left(\prod_{i=1}^{k-1} (\exp[-\alpha i\{\log(X_{i+1} - \theta) - \log(X_i - \theta)\}]\alpha i(X_i - \theta)^{-1}) \right)$$

$$\times f[\alpha\{\log(X_k - \theta) - \log a\}]\alpha(X_k - \theta)^{-1}.$$

Hence, up to an additive constant the negative log-likelihood of $\mathcal{X}$ equals

$$l_1 = \alpha \sum_{i=1}^{k-1} Y_i - k\alpha Y + \exp(\alpha Y) - k \log \alpha + \sum_{i=1}^{k} \log(X_i - \theta),$$

in which $Y_1, \ldots, Y_{k-1}$ and Y have the functional form given in (2.1). This is also the log-likelihood of points $\theta < X_1 < X_2 < \ldots$ from a Poisson process with cumulative intensity $\{(x - \theta)/a\}^\alpha$, for $x > 0$.

Differentiating l_1 with respect to a and equating to zero we may express the maximum-likelihood estimator of a as a function of those of θ and α: $\alpha \log a = \alpha \log (X_k - \theta) - \log k$. With this substitution we obtain $Y = \alpha^{-1} \log k$, and, up to an additive constant, the negative log-likelihood

$$l_2 = \alpha \sum_{i=1}^{k-1} Y_i - k \log \alpha + \sum_{i=1}^{k} \log (X_i - \theta). \qquad (2.2)$$

Differentiating with respect to α and equating to zero we obtain the maximum-likelihood estimator of α, which on substituting back into the formula for l_2 shows that, up to an additive constant, $k^{-1} l_2$ equals

$$l(\theta) = \log \left(\sum_{i=1}^{k-1} Y_i \right) + k^{-1} \sum_{i=1}^{k} \log (X_i - \theta).$$

Maximum-likelihood estimation of θ proceeds by minimizing $l(\theta)$. This may be shown to be identical with the "approximate" criterion suggested by Smith and Weissman (1985).

Next we elucidate performance of the maximum-likelihood estimator. Let $\epsilon \in (0, 1)$ be arbitrary but fixed, let θ_0 be the true value of θ, and let $\hat{\theta}$ denote any value of θ in the range $|\theta - \theta_0| < \min \{(1 - \epsilon) \exp (-\gamma/\alpha), X_1 - \theta_0\}$ which produces a local minimum of $l(\theta)$, where $\gamma = 0.5772 \ldots$ is Euler's constant. Define $\sigma_1^2 = \alpha^{-1}(\alpha - 1)^2(\alpha - 2)$.

We shall assume, here and in Theorem 2.2, that $a = 1$; in the contrary case, all asymptotic variances are multiplied by a^2. Note too that the estimator $\hat{\theta}$ is not consistent for θ_0 when $\alpha < 2$.

Theorem 2.1.

(a) *If $\alpha \geqslant 2$, then with probability tending to 1 as $k \to \infty$, $\hat{\theta}$ is well defined.*

(b) *If $\alpha = 2$, then $(\log k)^{1/2}(\hat{\theta} - \theta_0)$ is asymptotically normal $N(0, 4 \exp (-\gamma))$ and, if $\alpha > 2$, then $k^{(1/2)-(1/\alpha)}(\hat{\theta} - \theta_0)$ is asymptotically normal $N(0, \sigma_1^2)$.*

2.2. Estimation via minimum distance

The coefficient of determination $D(\theta)^2$ is defined as the ratio of the regression sum of squares to the total sum of squares in the linear model $Z = M\rho + \text{error}$, where $Z = (Z_1, \ldots, Z_k)^{\mathrm{T}}$, $Z_i = \log (X_i - \theta)$, M is the $k \times 2$ matrix $(\mathrm{E}(U), \mathbf{1})$, the vector $U = (U_1, \ldots, U_k)^{\mathrm{T}}$ has the density

$$\exp \left(\sum_{i=1}^{k} u_i - \exp (u_k) \right) \qquad \text{for } -\infty < u_1 < \ldots < u_k < \infty$$

(see, for example, Weissman (1978)), $\mathbf{1}$ is the $k \times 1$ column vector of 1s, and $\rho = (\rho_1, \rho_2)^{\mathrm{T}}$ is a 2×1 vector of scalars. Writing V for the variance matrix of U, $\hat{\rho}$ for the minimizer of the quadratic form $(Z - M\rho)V^{-1}(Z - M\rho)$, $\hat{Z} = M\hat{\rho}$ and $\overline{Z} = k^{-1}\sum_{i \leqslant k} Z_i$, we may show that $D(\theta)^2$ equals

$$\frac{(\hat{Z} - \overline{Z}\mathbf{1})^\mathrm{T} V^{-1}(\hat{Z} - \overline{Z}\mathbf{1})}{(Z - \overline{Z}\mathbf{1})^\mathrm{T} V^{-1}(Z - \overline{Z}\mathbf{1})}.$$

Wang (1995) has proved that this quantity equals

$$\frac{k^2\sigma_k^2 + k - 1}{(k - 1)\{k^2\sigma_k^2 G(\theta) + 1\}},$$

where $\sigma_k^2 = \mathrm{var}(U_k)$ and $G(\theta)$ is Greenwood's statistic, here given by

$$G(\theta) = \left(\sum_{i=1}^{k-1} Y_i^2\right) \Big/ \left(\sum_{i=1}^{k-1} Y_i\right)^2,$$

with Y_i having the meaning ascribed to it in Section 2.1. Thus, θ would be estimated as the minimizer of G.

Next we describe theoretical performance of $\hat{\theta}$, defined to be that value of θ in the range $|\theta - \theta_0| < \min\{(1 - \epsilon)\exp(-\gamma/\alpha), X_1 - \theta_0\}$ which minimizes $G(\theta)$. Put

$$\sigma_2^2 = 8\alpha^{-1}(\alpha - 1)^2(\alpha - 2)(2\alpha^2 - 4\alpha + 1)^{-2} = \sigma_1^2 8(2\alpha^2 - 4\alpha + 1)^{-2}.$$

Theorem 2.2. *The results of Theorem* 2.1 *hold in the case where $\hat{\theta}$ is the minimum-distance estimator, except that σ_1^2 there should here be replaced by σ_2^2.*

Comparing Theorems 2.1 and 2.2 we see that the minimum-distance estimator has lower variance than the maximum-likelihood estimator whenever $\alpha > (2^{1/2} + \frac{1}{2})^{1/2} + 1 = 2.384$.

In the next section we shall suggest an exact interval estimator of θ, based on the statistic $G(\theta)$. Such an estimator is possible because the distribution of $G(\theta_0)$ does not depend on α; that quantity cancels from the numerator and denominator. This property is not shared by $l(\theta_0)$, which is another reason why the minimum-distance approach is preferable. Confidence intervals whose coverage is *asymptotically* correct, as $k \to \infty$, may be constructed using the maximum-likelihood method, however (Smith and Weissman 1985).

More general models may also be put into the minimum-distance and maximum-likelihood frameworks. They include regression models of the form $X_i = \psi(\theta, i) + aW_i$, where $\psi(\theta, i)$ is a known function of a vector parameter θ and the index i, a is a scalar, and $W_1 < W_2 < \ldots$ have the multivariate extreme value distribution ascribed earlier to $(X_1 - \theta)/a < (X_2 - \theta)/a < \ldots$ (at the true values of the parameters θ and a). Examples include $\psi(\theta, i) = \theta_1 + \theta_2 i$, where $\theta = (\theta_1, \theta_2)$. Maximum-likelihood and minimum-distance estimators may be derived as before, the only change being that now

$$Y_i = i[\log\{X_{i+1} - \psi(\theta, i + 1)\} - \log\{X_i - \psi(\theta, i)\}],$$

$$Y = \log\{X_k - \psi(\theta, k)\} - \log a,$$

in place of the earlier definitions (2.1). Thus, it is again necessary to minimize $l(\theta)$ or $G(\theta)$,

this time with respect to the vector parameter θ. Since the minimum-distance method again avoids explicit estimation of α it can be expected to produce improved performance for large α values, although of course the definition of "sufficiently large α" will depend on the model and on the values of non-intercept parameters.

3. Applications to real and simulated data

It should be noted at the outset that in a minority of cases, neither the maximum-likelihood nor the minimum-distance estimator is well defined. Smith and Weissman (1985) observe that a graph of $l(\theta)$ can assume one of three forms, illustrated in their Figure 1. (Note that the figure depicts $-l$, in our notation.) In each case, the global minimum occurs at the smallest order statistic X_1 where it equals $-\infty$. Only for the first of Smith and Weissman's three types of curve does a local minimum occur at a θ value strictly less than X_1; in the third, a minimum occurs at $-\infty$. Following Smith and Weissman we shall say that the maximum-likelihood estimator "does not exist" unless the data produce a curve of the first type. Smith and Weissman (1985) point out that, when the maximum-likelihood estimator of θ is taken equal to X_1, inconsistent estimation of the shape parameter α results.

On the other hand, a graph of the function G seems always to be one of only two types, illustrated in Figure 1. The first has a well-defined global minimum at a value strictly less than X_1, and a global maximum at X_1. The second is monotonic increasing on $(-\infty, X_1)$ and so has its minimum at $-\infty$. In this case we shall say that the minimum-distance estimator does not exist. Note particularly that X_1 always produces a global maximum of G and a global minimum of l.
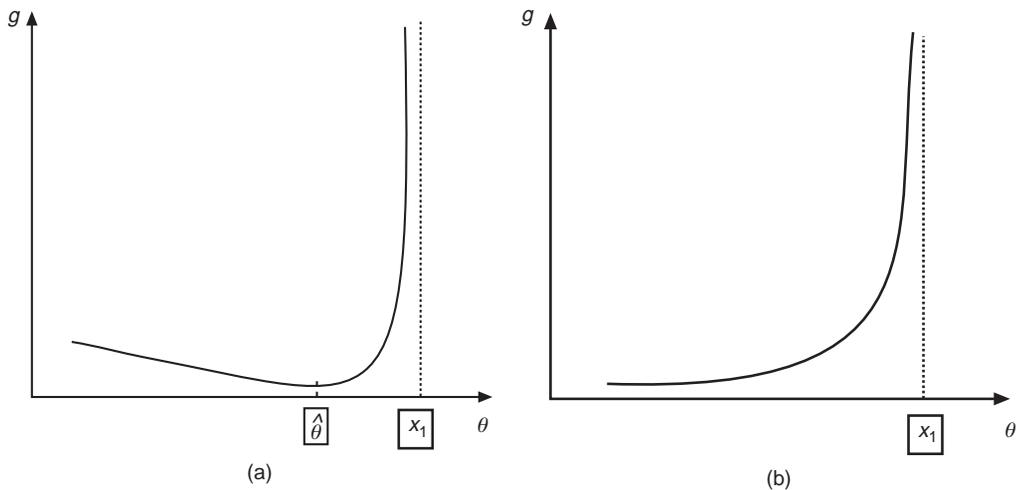

Figure 1. Two different forms of the graph of G, (a) one where the minimum is unique and finite, and (b) the other where there are no local minima and the global minimum occurs at $-\infty$.

Table 1 compares performance of the maximum-likelihood and minimum-distance estimators for small values of k. The data were generated from the limiting extremal distribution, and thus the end-point is $\theta_0 = 0$ (Weissman 1981). In the case of the maximum-likelihood estimator, again following Smith and Weissman (1985), biases and mean squared errors are given only for those cases where a local minimum of l occurs on $(-\infty, X_1)$ and, for the minimum-distance estimator, only in cases where the minimum of G is finite. Note particularly that the minimum-distance estimator is well defined more often than the maximum-likelihood estimator is and that, in those cases where it is well defined, it generally has less bias and less variance. Each value in the table was computed as the average over 1000 samples, and in particular the values of N are the numbers of the times out of 1000 that the respective estimator was well defined. We considered a range of values of α less than the minimum, 2, for which the estimators are consistent, in order to check

Table 1. Comparison of performance of maximum-likelihood and minimum-distance estimators, where N denotes the number of times, out of 1000, for which the respective estimator is well defined (see text for details)

k	α	Minimum-distance estimator				Maximum-likelihood estimator			
		Mean $(\hat{\theta})$	Standard deviation $(\hat{\theta})$	Mean squared error $(\hat{\theta})$	N	Mean $(\hat{\theta})$	Standard deviation $(\hat{\theta})$	Mean squared error $(\hat{\theta})$	N
5	5	0.15	2.01	2.02^2	594	-0.20	3.39	3.40^2	94
	4	0.07	2.11	2.11^2	601	-0.69	3.97	4.03^2	81
	3	-0.10	2.12	2.12^2	637	-1.71	4.56	4.87^2	72
	2	-0.04	2.19	2.19^2	682	-2.04	4.73	5.15^2	66
	1.25	0.20	1.97	1.97^2	726	-1.74	5.63	5.89^2	61
10	5	0.11	1.85	1.85^2	661	-0.23	2.94	2.95^2	402
	4	0.12	1.91	1.91^2	703	-0.75	3.11	3.20^2	385
	3	0.09	1.96	1.96^2	755	-1.11	3.56	3.73^2	356
	2	0.05	1.99	1.99^2	797	-1.24	4.07	4.25^2	296
	1.25	0.02	1.85	1.85^2	834	-2.52	5.25	5.82^2	146
15	5	0.19	1.59	1.60^2	682	-0.26	2.81	2.82^2	625
	4	0.08	1.69	1.69^2	781	-0.45	3.00	3.03^2	613
	3	-0.09	1.81	1.81^2	840	-0.51	3.05	3.09^2	591
	2	-0.18	1.88	1.89^2	855	-0.71	3.20	3.28^2	554
	1.25	0.08	1.77	1.77^2	898	-0.95	3.04	3.18^2	242
25	5	0.04	1.59	1.59^2	752	-0.29	2.75	2.77^2	828
	4	0.05	1.60	1.60^2	835	-0.25	1.84	1.86^2	817
	3	-0.11	1.77	1.77^2	891	-0.32	1.95	1.97^2	805
	2	-0.12	1.83	1.83^2	900	-0.08	1.81	1.81^2	798
	1.25	0.18	1.51	1.53^2	921	0.10	1.52	1.52^2	383

that neither method performs in an aberrant way. They do not, and the case $\alpha = 1.25$ illustrated in the table is typical.

Comparison of maximum-likelihood and minimum-distance methods on the basis of interval estimates, rather than point estimates, is awkward for at least two reasons. First, maximum-likelihood interval estimates are cumbersome to compute, because the distribution of $l(\theta)$ at the true parameter value θ_0 depends on the unknown α. The distribution could be approximated by simulation, with α set equal to its maximum-likelihood estimate, and an interval estimate computed by reference to this Monte Carlo approximation. However, the coverage of such an interval would be only approximately correct. As a result, the relative performance (e.g. in terms of length) of minimum-distance and maximum-likelihood intervals would be confounded by the level inaccuracy of the latter. The second difficulty derives from the relative complexity of the function $l(\theta)$, noted in the first paragraph of this section. In particular, confidence regions of the form $\{\theta: l(\theta) \leqslant \text{constant}\}$ are generally not connected. This problem can be overcome by imposing restrictions on the nature of the region, but that approach is not attractive.

Our next example uses data from Kimball (1960), also employed by Smith and Weissman (1985) and given there in enough detail for the analyses reported in that paper and ours. The data are of the survival times of 208 mice exposed to radiation, and our aim is to estimate the threshold, or lower end-point, of the distribution of survival time. When maximum likelihood is employed, the estimator based on the k shortest survival times is well defined only for $k \geqslant 8$ (although not including $k = 9$), but for the minimum-distance estimator it is well defined whenever $k \geqslant 4$. The minimum-distance estimator is also less variable than that based on maximum likelihood, as evidenced by the comparison in Table 2. The estimator of α obtained by either method lies between 2 and 3, depending on

Table 2. Comparison of maximum-likelihood and minimum-distance estimators of θ based on the k smallest survival times, for the data of Kimball (1960), where asterisks denote cases where the estimator is not well defined

k	Minimum-distance estimator	Maximum-likelihood estimator
4	22.4	*
6	24.9	*
8	26.0	37.0
10	33.6	36.5
12	32.2	35.3
15	31.6	31.9
20	34.8	35.5
25	34.2	34.5
30	33.6	26.3
35	33.0	24.0
40	33.8	33.1

the value of k. (An estimator of α based on minimum-distance methods may be computed from $\hat{\rho}$, defined in Section 2.2 (Weissman 1978).)

Our final example is included partly to demonstrate that the minimum-distance estimator is well defined even in cases where the estimate of α is a less than 1. By way of comparison, the maximum likelihood estimator is not well defined in this context. Table 3 lists the men's 100 m records, in seconds, in the 1988 and 1992 Olympic Games (Estadi Olímpic 1992, Wallechinsky 1992). We consider only the best performance of each individual. For example, Christie's best time of 9.96 s is the fastest of his 1988 and 1992 times of 9.97 s and 9.96 s, respectively. We suggest that it is unwise to include more than two consecutive Olympic Games, owing to the tendency for distributions to change with time. Furthermore, the two previous Olympics, in 1980 and 1984, were subject to boycotts.

In general, there is potential for training regimes and competition to lead to athletic recods that might not realistically arise from independent and identically distributed data. A test of goodness of fit is a valuable aid to assessing the significance of this problem. For the data in Table 3, the W statistic of Hasofer and Wang (1992) gives a p value of less than 0.01, supporting our particular approach to analysing these data.

Figure 2 depicts the function G for these data. It is minimized at $\theta = 9.91$. The corresponding estimate of α is 0.82. The maximum-likelihood estimator is not well defined for these data, unless we admit $X_1 = 9.92$; then, the maximum-likelihood estimator of α is 0.

An α-level confidence interval for θ_0 may be defined as the set of values of θ such that $G(\theta) \leqslant g$, where g is the upper α-level point of the distribution of $G(\theta_0)$. Critical points are available from Burrows (1979) when $4 \leqslant k \leqslant 12$. Stephens (1981) gives points for larger k. Depending on the data, the confidence interval may be either semi-infinite, of the form $(-\infty, \hat{\theta}_1)$, or bounded. In the case of the mice data, and for $k = 10$, the interval was $(-\infty, 39.9)$, or more correctly $(0, 39.9)$, taking into account the physical character of survival times. For the athletic data the confidence interval was $(9.80, 9.92)$. Regardless of

Table 3. Times in men's 100 m event at 1988 and 1992 Olympic Games

|	Competitor	Time (s)	Year, Place
(1)	C. Lewis	9.92	1988 Seoul
(2)	L. Christie	9.96	1992 Barcelona
(3)	C. Smith	9.99	1988 Seoul
(4)	F. Fredericks	10.02	1992 Barcelona
(5)	D. Mitchell	10.04	1992 Barcelona
(6)	B. Surin	10.09	1992 Barcelona
(7)	L. Burrell	10.10	1992 Barcelona
(8)	R. da Silva	10.11	1988 Seoul
(9)	D. Williams	10.11	1988 Seoul
(10)	O. Adenniken	10.12	1992 Barcelona
(11)	R. Stewart	10.22	1992 Barcelona
(12)	D. Ezinwa	10.26	1992 Barcelona
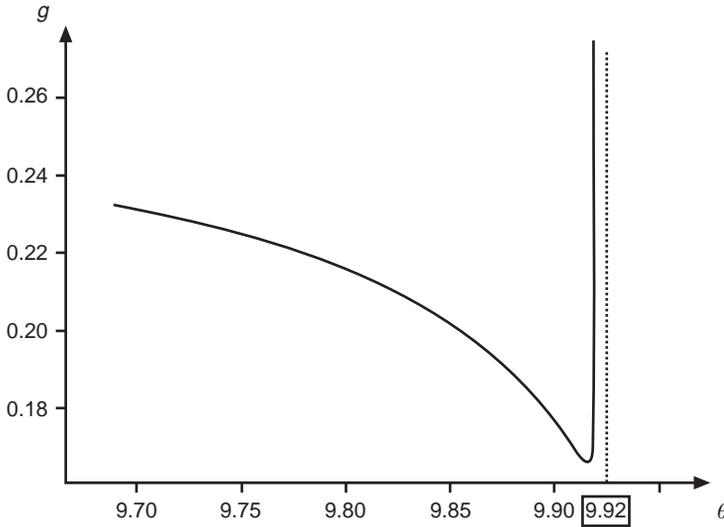
Figure 2. Graph of $G(\theta)$ for data on times in men's 100 m Olympic event.

the value of α, but assuming that the data enjoy exactly an extreme value distribution, these are exact confidence intervals for θ_0.

It might be argued that, when the maximum-likelihood estimate fails to exist, i.e. takes the value infinity, the data support the value $\theta_0 = \infty$ more strongly than they do any finite value of α. While $\theta_0 = \infty$ is often not plausible on physical grounds, this type of argument is common in statistics and can be helpful. For example, when using high-order kernel methods, one sometimes obtains a negative density estimate which, while physically implausible, may be interpreted as implying that there is not strong evidence to support the hypothesis that the true density is non-zero at the point of estimation.

4. Technical arguments

Proof of Theorem 2.2.

(i) *Taylor expansion of G.* In the definition of Y_i at (2.1), interpret θ as an algebraic variable, and let the true value of θ be $\theta_0 = 0$. For $j \geqslant 1$ put $d_{1ji} = (i/j)(X_i^{-j} - X_{i+1}^{-j})$, $Y_{0i} = i(\log X_{i+1} - \log X_i)$, $A_l = \sum_{1 \leqslant i \leqslant k-1} Y_{0i}^l$,

$$d_{2ji} = \sum_{(j_1, j_2): j_1 + j_2 = j} \sum d_{1j_1 i} d_{1j_2 i} + 2 Y_{0i} d_{1ji}, \qquad a_{lj} = \sum_{i=1}^{k-1} d_{lji},$$

for $l = 1, 2$. In this notation we have, by Taylor expansion and provided that $|\theta| < X_1$,

$$Y_i = Y_{0i} + \sum_{j=1}^{\infty} d_{1ji}\theta^j, \qquad \sum_{i=1}^{k-1} Y_i^l = A_l + \sum_{j=1}^{\infty} a_{lj}\theta^j$$

and

$$A_1^2 G(\theta) = A_2 + \theta B_1 + \theta^2 B_2 + R, \tag{4.1}$$

where

$$B_1 = a_{21} - 2A_1^{-1}A_2 a_{11}, \tag{4.2}$$

$$B_2 = a_{22} - 2A_1^{-1}A_2 a_{12} - 2A_1^{-1}a_{11}a_{21} + 3A_1^{-2}A_2 a_{11}^2, \tag{4.3}$$

and the remainder term R represents contributions from powers of θ higher than the second.

(ii) *Properties of B_1*. We may write $\alpha Y_{0i} = Z_i$, where $Z_1, \ldots, Z_k$ are independent, exponentially distributed random variables. In that notation, and with $\beta = 1/\alpha$,

$$X_i = \exp\left\{-\beta\left(\sum_{j=i}^{\infty} j^{-1}(Z_j - 1) + \gamma - \sum_{j=1}^{i-1} j^{-1}\right)\right\} = i^{\beta}\{1 + O_p(i^{-1/2})\}, \tag{4.4}$$

uniformly in i, where γ denotes Euler's constant (Hall 1978). Therefore, defining $S_j = \sum_{i \leqslant k-1}(Z_i^j - j) = O_p(k^{1/2})$, we have, by (4.2),

$$\tfrac{1}{2}\alpha^2 A_1 B_1 = (k - 1 + S_1)\sum_{i=1}^{k-1} d_{11i}(Z_i - 2) + (2S_1 - S_2)\sum_{i=1}^{k-1} d_{11i}. \tag{4.5}$$

It may be proved from (4.4) and (4.5) that, for $\beta = \tfrac{1}{2}$, $\tfrac{1}{2}\alpha^2 A_1 B_1$ is asymptotically normal $N(0, 2k^2 \log k)$ while, for $\beta < \tfrac{1}{2}$, it is asymptotically normal $N\{0, 8\beta^4(1-\beta)^{-2}(1-2\beta)^{-1}k^{3-2\beta}\}$. Combining these results and noting that $\alpha A_1/k \to 1$ in probability, we may show that

$$\begin{aligned}&\text{(a) if } \beta = \tfrac{1}{2}, \; B_1 \sim_d N(0, 2\log k),\\ &\text{(b) if } \beta < \tfrac{1}{2}, \; B_1 \sim_d N(0, 32\beta^6(1-\beta)^{-2}(1-2\beta)^{-1}k^{1-2\beta}),\end{aligned} \tag{4.6}$$

where $\sim_d$ denotes "is asymptotic in distribution to".

(iii) *Properties of B_2*. Observe that each $a_{lj} \geqslant 0$, and that $a_{11} \sim_p (1-\beta)^{-1}\beta k^{1-\beta}$, $a_{12} \sim_p \beta\sum_{1 \leqslant i \leqslant k} i^{-2\beta}$, $a_{21} \sim_p 2\beta a_{11}$, $a_{22} \sim_p 3\beta^2\sum_{1 \leqslant i \leqslant k} i^{-2\beta}$, where $\sim_p$ means that the ratio of the left- and right-hand sides converges to 1 in probability. Therefore, by (4.3) and since $\alpha^j A_j/k \to 1$ in probability for $j = 1, 2$,

$$B_2 \sim_p \beta^2\left(\sum_{i=1}^{k} i^{-2\beta} + (1-\beta)^{-2}k^{1-2\beta}\right). \tag{4.7}$$

(iv) *Properties of R*. Note that

$$d_{1ji} \leqslant \frac{i(X_{i+1} - X_i)}{X_i^j X_{i+1}}, \tag{4.8}$$

whence

$$\sum_{(j_1,j_2):j_1+j_2=j}\sum d_{1j_1i}d_{1j_2i} \leq (j-1)\left(\frac{i(X_{i+1}-X_i)}{X_{i+1}}\right)^2 X_i^{-j}.$$

Hence,

$$0 \leq a_{2j} \leq (j-1)\sum_{i=1}^{k-1}\left(\frac{i(X_{i+1}-X_i)}{X_{i+1}}\right)^2 X_i^{-j} + 2\beta\sum_{i=1}^{k-1}Z_i\frac{i(X_{i+1}-X_i)}{X_{i+1}}X_i^{-j}. \qquad (4.9)$$

More simply, by (4.8),

$$0 \leq a_{1j} \leq \sum_{i=1}^{k-1}\frac{i(X_{i+1}-X_i)}{X_{i+1}}X_i^{-j}. \qquad (4.10)$$

For a constant $C_1 > 0$,

$$X_i^{-1} \leq (1 + C_1 i^{-1})i^{-\beta}\exp\left(\beta\sum_{j=i}^{\infty}j^{-1}(Z_j - 1)\right), \qquad (4.11)$$

and $\sum_{j\geq i}j^{-1}(Z_j - 1) \to 0$ almost surely as $i \to \infty$. It follows from (4.9)–(4.11) that

$$|R| = O_p\left\{\left(\sum_{j=3}^{\infty}\{|\theta|(1 + \tfrac{1}{2}\epsilon)\}^j\right)\sum_{i=1}^{k}i^{-2\beta}\right\} \qquad (4.12)$$

uniformly in $|\theta| \leq 1 - \epsilon$, for arbitrary fixed $\epsilon \in (0, 1)$. The theorem follows on combining (4.1), (4.6), (4.7) and (4.12). ☐

Proof of Theorem 2.1. In analogy to (4.1),

$$l(\theta) = \left(\log A_1 + k^{-1}\sum_{i=1}^{k}\log X_i\right) + B_1\theta + B_2\theta^2 + R,$$

where $B_1 = A_1^{-1}a_{11} - k^{-1}\sum_{1\leq i\leq k}X_i^{-1}$,

$$B_2 = A_1^{-1}a_{12} - \tfrac{1}{2}A_1^{-2}a_{11}^2 - \tfrac{1}{2}k^{-1}\sum_{i=1}^{k}X_i^{-2},$$

and R represents terms in θ^3, θ^4, $\ldots$. If $\beta < \tfrac{1}{2}$ then, as $k \to \infty$,

$$B_1 \sim_d N(0, \beta^2(1-\beta)^{-2}(1-2\beta)^{-1}k^{-1-2\beta}),$$

$B_2 \sim_p \tfrac{1}{2}\beta^2(1-\beta)^{-2}(1-2\beta)^{-1}k^{-2\beta}$, and R may be bounded as in the proof of Theorem 2.2 while, if $\beta = \tfrac{1}{2}$, $B_1 \sim_d N(0, k^{-2}\log k)$ and $B_2 \sim_p \tfrac{1}{2}k^{-1}\log k$, from which the theorem follows as before. ☐

Acknowledgements

References

Burrows, P. (1979) Selected percentage points of Greenwood's statistics. *J. Roy. Statist. Soc. A*, **142**, 256–258.

Cooke, P.J. (1979) Statistical inference for bounds of random variables. *Biometrika*, **66**, 367–374.

Cooke, P.J. (1980) Optimal linear estimation of bounds of random variables. *Biometrika*, **67**, 257–258.

David, H.A. (1981) *Order Statistics*, 2nd edn. New York: Wiley.

de Haan, L. and Resnick, S. (1994) Estimating the home range. *J. Appl. Probab.*, **31**, 700–720.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events*. Heidelberg: Springer-Verlag.

Estadi Olímpic (1992) *Games of the XXV Olympiad*, Barcelona, 1992. Australian Olympic Committee.

Galambos, J. (1978) *The Asymptotic Theory of Extreme Order Statistics*. New York: Wiley.

Hall, P. (1978) Representations and limit theorems for extreme value distributions. *J. Appl. Probab.*, **15**, 639–644.

Hall, P. (1982) On estimating the endpoint of a distribution. *Ann. Statist.*, **10**, 556–568.

Hartley, H.O., and Pfaffenberger, R.C. (1972) Quadratic forms in order statistics used as goodness-of-fit criteria. *Biometrika*, **59**, 605–611.

Hasofer, M. and Wang, Z. (1992) A test for extreme value domain of attraction. *J. Amer. Statist. Assoc.*, **87**, 171–177.

Kimball, A.W. (1960) Estimation of mortality intensities in animal experiments. *Biometrics*, **16**, 505–521.

Peto, R. and Lee, P. (1973) Weibull distributions for continuous-carcinogenesis experiments. *Biometrics*, **29**, 457–470.

Smith, R.L. (1985) Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, **72**, 67–90.

Smith, R.L. (1987) Estimating the tails of probability distributions. *Ann. Statist.*, **15**, 1174–1207.

Smith, R.L. and Weissman, I. (1985) Maximum likelihood estimation of the lower tail of a probability distribution. *J. Roy. Statist. Soc. B*, **47**, 285–298.

Stephens, M.A. (1981) Further percentage points for Greenwood's statistics. *J. Roy. Statist. Soc. A*, **144**, 364–366.

Wallechinsky, D. (1992) *The Complete Book of the Olympics*. London: Aurum.

Wang, J.Z. (1995) Selection of the k largest order statistics for the domain of attraction of the Gumbel distribution. *J. Amer. Statist. Assoc.*, **90**, 1055–1061.

Weissman, I. (1978) Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.*, **73**, 812–815.

Weissman, I. (1981) Confidence intervals for the threshold parameter, II: Unknown shape parameter. *Commun. Statist. Theor. Meth. A*, **11**, 2451–2474.