Let $X_1, X_2, \ldots, X_n$ be random variables with common distribution function $F(x)$. Put $Z_n = \max (X_1, X_2, \ldots, X_n)$. Assume that the interdependence of the $X_j$ is such that, with suitable constants $a_n$ and $b_n > 0$, $P(Z_n < a_n + b_n x)$ is asymptotically $F_n^n(a_n + b_n x)$; see Chapter 3 in the book by the present author.
J. Galambos (1978) for models with this property. Hence, for most functions $F(x)$ occuring in practice, $(Z_n - n) / \delta_n$ has an asymptotic distribution $H(x)$ as $n \to + \infty$. This asymptotic distribution is necessarily one of the three classical types obtained in the 1920s by Fisher, Tippett and Frechet. However, if $F(x)$ is not known (which is the case of most practical situations), then neither the normalizing constants $a_n$ and $b_n$, nor the actual type of the limit distribution $H(x)$ can be determined. Since goodness of fit tests are not sensitive for the tail of the distribution $F(x)$, and since an estimate for $F(x)$ is not suitable in extreme value theory (see Galambos (1978), Example 2.6.3), some other techniques are needed which are free from $F(x)$. The paper is devoted to such a technique, which is an extension of the approach by J. Pickands III (1975). In such a method, one uses a number of upper extremes and the parametric form of the three classical types for $H(x)$, known as the von Mises form.
Statistical aspects of extreme value theory

by

Janos Galambos

Department of Mathematics
Temple University

Preliminary Report. Presented at the
Australian Statistical Conference,
11-14 July, 1978, Canberra, Australia.

Research supported by a Grant (#78-3504) from AFOSR
to Temple University.
Abstract

Let $X_1, X_2, \ldots, X_n$ be random variables with common distribution function $F(x)$. Put $Z_n = \max(X_1, X_2, \ldots, X_n)$. Assume that the interdependence of the $X_j$ is such that, with suitable constants $a_n$ and $b_n > 0$, $P(Z_n < a_n + b_n x)$ is asymptotically $F^n(a_n + b_n x)$; see Chapter 3 in the book by the present author, J. Galambos (1978) for models with this property. Hence, for most functions $F(x)$ occurring in practice, $(Z_n - a_n)/b_n$ has an asymptotic distribution $H(x)$ as $n \to +\infty$. This asymptotic distribution is necessarily one of the three classical types obtained in the 1920s by Fisher, Tippett and Frechet. However, if $F(x)$ is not known (which is the case of most practical situations), then neither the normalizing constants $a_n$ and $b_n$, nor the actual type of the limit distribution $H(x)$ can be determined. Since goodness of fit tests are not sensitive for the tail of the distribution $F(x)$, and since an estimate for $F(x)$ is not suitable in extreme value theory (see Galambos (1978), Example 2.6.3), some other techniques are needed which are free from $F(x)$. The paper is devoted to such a technique, which is an extension of the approach by J. Pickands III (1975). In such a method, one uses a number of upper extremes and the parametric form of the three classical types for $H(x)$, known as the von Mises form.

References


A new statistical method

Let $X_1, X_2, \ldots, X_n$ be identically distributed random variables. Put $Z_n = \max(X_1, X_2, \ldots, X_n)$. Assume that there are constants $a_n$ and $b_n > 0$ such that $P(Z_n < a_n + b_n x)$ can be approximated by $F_n(a_n + b_n x)$, where $F$ is usually the common distribution function of the $X$'s. This implies that if $F$ has a smooth enough tail, then $(Z_n - a_n)/b_n$ has an asymptotic distribution $H(x)$. Our aim is to determine the form of $H(x)$ in the case when $F$ is not known. The approximation of $F(x)$ by the empirical distribution function, or by other classical methods, is inefficient in extreme value theory; see Galambos (1978a) for detailed discussion. Hence a direct method is required to get information on $H(x)$ without trying to determine $F$.

There is only one work on this line in the literature by J. Pickands III (1975). Here we consider an alternate model. While Pickands develops a method, in which data will automatically separate the form of $H(x)$ out of the three possibilities, our method is a hypothesis test, in which we make a decision on the mentioned form. It remains an unsolved problem to compare the two methods on a theoretical basis. Because the nature of the two approaches is different, a clear comparison is difficult. However, advantages and disadvantages of one method or the other can be clarified from several points of view. As a matter of opinion, we can state that the method by Pickands would be preferable to a test if the accuracy of his method were available. Since the Pickands method involves three different passages to limit for supporting its conclusions, an estimate of accuracy is quite important. On the other hand, in some cases, one would need the actual form of $H(x)$, for which a test is suitable, while the method of Pickands leaves this undecided.

The method described below is in an early stage of investigation. However, initial results and some numerical computations suggest that it is a useful attempt and it thus deserves a thorough investigation.
Basic assumptions: 1. The $X_j$ are identically distributed with common distribution function $F(x)$; 2. The dependence of the $X_j$ is such that the distribution of $Z_n$ can be approximated by the $n$-th power of $F(x)$ for large values of $x$; 3. $F(x) < 1$ for all $x$; 4. $E(X_j)$ is finite; and 5. There are constants $a_n$ and $b_n > 0$ such that $(Z_n - a_n)/b_n$ has an asymptotic distribution $H(x)$.

The above assumptions imply that there are constants $A$ and $B > 0$ such that $H(A+Bx)$ equals either $H_{3,0}(x) = \exp(-e^{-x})$ or $H_{1,c}(x) = \exp(-x^{-c})$, $x > 0$, $c > 0$. Our aim is to decide by a statistical method based on the $X_j$ that which of the two forms mentioned applies.

A test for the form of $H(x)$. Notice that the $X_j$ provide a single observation on $Z_n$. Yet a number of extremes can be used to provide information on $H(x)$. The actual method is as follows. Choose a number $t$. Select those $X_j$ that exceed $t$. Transform such $X_j$s to

$$Y_j = (X_j - t)/E(X_j - t | X_j > t).$$

Test that the distribution of these $Y_j$ is unit exponential against the composite hypothesis that this distribution belongs to the Pareto family. Accepting the hypothesis means acceptance of $H(A+Bx) = H_{3,0}(x)$, while rejection of the null hypothesis leads to the acceptance of $H_{1,c}(x)$ as the limiting distribution of the properly normalized maximum. The power of the test depends on the choice of $t$, which choice is yet to be investigated in terms of some optimality criterion.

The theoretical justification of the above procedure is a classical criterion, combined with a result of L. de Haan (1970), for distributions belonging to given domains of attraction for the maxima. Details will be given in an expanded form of this present paper.
3. References


