Testing the Gumbel hypothesis by Galton’s ratio

JOHAN SEGERS
Departement Wiskunde, Katholieke Universiteit Leuven, Celestijnenlaan 200B, B-3001 Heverlee, Belgium
E-mail: johan.segers@wis.kuleuven.ac.be

JEF TEUGELS
Departement Wiskunde, Katholieke Universiteit Leuven, Celestijnenlaan 200B, B-3001 Heverlee, Belgium
E-mail: jef.teugels@wis.kuleuven.ac.be

[Received, Revised and Accepted March 8, 2001]

Abstract. We study a generalization of a ratio of spacings introduced by Galton in 1902. The ratio proves to be an important building block in the construction of a large sample test for the hypothesis that a distribution from an extremal domain of attraction belongs to the domain of attraction of the Gumbel law.

Key words. extremal law, domain of extremal attraction, Gumbel law, hypothesis test, Galton’s ratio

AMS 1991 Subject Classification. MSC 2000: 62G32

1. Introduction

Let $X_1, X_2, \ldots, X_n$ be independent observations from the distribution function (d.f.) $F$. We order the sample values to obtain the order statistics

$$X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$$

that will play a predominant role in what follows. Although this is not strictly necessary, we will sometimes assume that $F$ is continuous. This assumption could be removed at the cost of some extra notational complexity.

The ratio

$$G_n := \frac{X_{n,n} - X_{n-2,n}}{X_{n-1,n} - X_{n-2,n}}$$

has been introduced by Galton (1902), and is treated in Gumbel (1958) for the case of the exponential distribution. There is no real problem to deal with the more general case that we still will coin Galton’s ratio, i.e. for $1 \leq s \leq n - 2$ we define
\[ G_n(s) := \frac{X_{n-s+1,n} - X_{n-s-1,n}}{X_{n-s,n} - X_{n-s-1,n}}. \] (1.1)

In order to study the asymptotic behavior of \( G_n(s) \) for \( n \) large and \( s \) fixed, it is customary to assume that \( F \) belongs to an extremal domain of attraction, by which we mean that for appropriate positive constants \( a_n \) and real constants \( b_n, n = 1, 2, \ldots \), the normalized maximum \( (X_{n,n} - b_n)/a_n \) converges in distribution to a non-degenerate limit law as \( n \to \infty \). This can be rephrased in terms of the tail quantile function \( U(y) := \inf \{ x : F(x) \geq 1 - 1/y \}, y > 1 \). The d.f. \( F \) belongs to an extremal domain of attraction if and only if there exists a positive, measurable auxiliary function \( g \) and a real tail index \( \gamma \) such that for all \( \lambda > 0 \) the condition

\[ \lim_{x \to \infty} \frac{U(\lambda x) - U(x)}{g(x)} = \int_1^\infty w^{\gamma-1}dw = h_\gamma(\lambda) \] (1.2)

holds see, for example, de Haan (1993). Note that in particular \( h_0(\lambda) = \log \lambda \). It then follows that the constants can be taken as \( b_n = U(n) \) and \( a_n = g(n) \) and that the possible limit laws for \( (X_{n,n} - U(n))/g(n) \) are given by the extreme value distributions

\[ G_\gamma(x) := \exp\{ - (1 + \gamma x)^{-1/\gamma} \}, \quad 1 + \gamma x > 0, \] (1.3)

where \( (1 + \gamma x)^{-1/\gamma} \) has to be understood as \( e^{-x} \) in case \( \gamma \) equals 0. We will denote (1.2) by \( F \in D(G_\gamma) \). A possible choice for the auxiliary function \( g \) is \( g(x) = \{ U(ex) - U(x) \}/h_\gamma(e) \), \( x > 1 \); it always satisfies the regular variation property

\[ \lim_{x \to \infty} \frac{g(\lambda x)}{g(x)} = \lambda^\gamma, \quad \lambda > 0. \] (1.4)

The literature on extreme value theory and on regular variation is enormous. See, for example, Beirlant, Vynckier and Teugels (1996), Bingham, Goldie and Teugels (1987), de Haan (1970), Gumbel (1958), Leadbetter, Lindgren and Rootzén (1983) and Reiss (1989) for the necessary background and references.

The tail index \( \gamma \) reveals a great deal about the weight in the right tail of \( F \). For \( \gamma > 0 \), the tail function looks like \( 1 - F(x) \approx x^{-1/\gamma} \ell(x) \) for some slowly varying function \( \ell \), and thus \( F \) is heavy-tailed. If \( \gamma < 0 \), then \( F \) has a finite right endpoint and \( 1 - F \) vanishes at an algebraic rate near this endpoint. The intermediate case \( \gamma = 0 \) covers a wide range of distributions, with finite or infinite right endpoint, and including the normal, gamma (in particular exponential) and lognormal distribution. The corresponding extreme value distribution

\[ G_0(x) = \exp(-x), \quad x \in \mathbb{R}, \]