I. Statistical Framework

Let \( x_1, \ldots, x_n \) be a sample of independent and identically distributed random variables drawn from a distribution function \( F \). Let \( X_{(n)} \leq \cdots \leq X_{(1)} \) denote the order statistics associated to this sample.

We want to estimate the extreme quantile \( x_\tau \) of order \( \tau \in [0, 1] \) defined by:
\[
x_\tau = F^{-1}(\tau) = \inf \{ x : F(x) \leq \tau \},
\]
with \( \tau \to 0 \) when \( n \to \infty \). The function \( F^{-1} \) is the generalized inverse of the non-increasing function \( F \).

In [2], a family of distributions is introduced, which encompasses the whole Fréchet maximum domain of attraction as well as Weibull tail-distributions. These distributions depend on two parameters \( \tau \in [0, 1] \) and \( \theta > 0 \).

II. Model: Gumbel/Fréchet

The asymptotic distributions of \( \hat{\theta}_n \) and \( \hat{\tau}_n \) have been established in [2] under a second-order condition on \( \ell \).

There exist \( \rho \in \mathbb{R} \) a function \( b \) satisfying \( b(x) \to 0 \) and \( |b(x)| \) asymptotically decreasing such that uniformly locally on \( x > 0 \):
\[
\log \left( \frac{f(x)}{f(\ell)} \right) \sim b(x), \quad x \to \infty.
\]

The main goal of this work is to propose an estimator for \( \tau \) independent from \( \theta \). This parameter controls the behavior of the tail-distribution: the larger the value of \( \tau \), the heavier the tail.

III. Estimators depending on \( \tau \)

Denoting by \( (k_n) \) an intermediate sequence of integers, the following estimator of \( \theta \) is considered in [2]:
\[
\hat{\theta}_n = \frac{H_n(k_n)}{\sum_{i=1}^{k_n} \log(X_{(i+1)} - X_{(i)}) - \log(X_{(k_n+1)})},
\]
where \( H_n(k_n) \) is the Hill estimator [3],
\[
H_n(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \log(X_{(i+1)} - X_{(i)}) - \log(X_{(k_n+1)}).
\]

An estimator of the extreme quantile \( x_\hat{\tau} \) proposed in [2]:
\[
\hat{x}_{\hat{\tau}} = X_{(k_n+1)} \exp \left( \frac{\hat{\theta}_n(k_n)}{\hat{\theta}_n(1)} - K_n(\log(1/p_0)) - K_n(\log(n/\hat{\theta}_n)) \right).
\]

IV. Main Goal

Replacing \( \tau \) by \( \hat{\tau}_n \), we obtain:
\[
\hat{\theta}_n \sim H_n(k_n)/\sum_{i=1}^{k_n} \log(X_{(i+1)} - X_{(i)}) - \log(X_{(k_n+1)}),
\]
\[
\text{an estimator of } x_{\hat{\tau}_n}:
\]
\[
\hat{x}_{\hat{\tau}_n} = X_{(k_n+1)} \exp \left( \frac{\hat{\theta}_n(k_n)}{\hat{\theta}_n(1)} - K_n(\log(1/p_0)) - K_n(\log(n/\hat{\theta}_n)) \right).
\]

Under some assumptions on \( (k_n) \) and \( (\hat{\theta}_n) \) we establish the asymptotic normality of \( \hat{\theta}_n \) and \( \hat{x}_{\hat{\tau}_n} \). In particular:
\[
\frac{\sqrt{n} \log(n/\hat{\theta}_n)}{\hat{\theta}_n} \sim N(0, 1).
\]

V. Estimator of \( \tau \)

Let us consider for \( t > t' \):
\[
\psi(t, t') = \frac{\exp(t') - 1}{\exp(t) - 1}.
\]

Denoting by \( (k_n) \) and \( (\ell_n) \) two intermediate sequences of integers such that \( \ell_n > k_n \), the following estimator of \( \tau \) is considered:
\[
\hat{\tau}_n = \psi \left( \ell_n, k_n \right) \frac{n}{\log(n/\ell_n), \log(n/k_n)}
\]
\[\text{if } \frac{\ell_n}{k_n} = \frac{\log(n/\ell_n)}{\log(n/k_n)} > 1/n,\]
\[\text{where } n \in [0, 1].\]

\( \hat{\tau}_n \) exists because \( \psi(x, y) \) is a bijection.

VI. Asymptotic distribution

The estimator \( \hat{x}_{\hat{\tau}_n} \) is computed on \( N = 100 \) samples of size \( n = 500 \) for \( k_n = 2, \ldots, 109 \) and \( \ell_n = 495, \ldots, 500 \) when \( p_0 = 10^{-3} \).

The associated deciles of the empirical Mean-Squared Error \( \text{MSE} \) are plotted


VII. Numerical experiments on simulated data

**Gamma distribution** \( \mathcal{G}(\text{Gumbel}) \)

**Pareto distribution** \( \mathcal{G}(\text{Fréchet}) \)

**MSE** Dekkers et al.

**MSE** Dekkers et al.

**MSE** Dekkers et al.

Bibliography


