Estimation of (conditional) extreme probabilities and quantiles with application to the measurement of the activity of bivalves

by Ion Grama
Laboratoire de Mathématiques de Bretagne Atlantique (LMBA)
Université de Bretagne-Sud & UMR CNRS 6205

Let \( X_1, \ldots, X_n \), be i.i.d. observations with continuous d.f. \( F \) supported on the interval \([x_0, \infty)\), \( x_0 \geq 0 \). Assume that d.f. \( F \) is "heavy tailed", i.e. that \( F \) belongs to the domain of attraction of the Fréchet law with parameter \( 1/\gamma \). By Fisher-Tippett-Gnedenko theorem this is equivalent to saying that for any \( x \geq 1 \),

\[
F_t(x) \rightarrow P_\gamma(x) \quad \text{as } t \rightarrow \infty,
\]

where \( F_t(x) \) is the excess d.f. over the threshold \( t > x_0 \) defined by \( F_t(x) = 1 - \frac{1-F(x)}{1-F(t)} \), \( x \geq 1 \) and \( P_\theta(x) = 1 - x^{-1/\theta} \), \( x \geq 1 \) is the standard Pareto d.f. with parameter \( \theta > 0 \). Relation (1) suggests using \( P_\gamma(x) \) with estimated \( \gamma \) as an approximation of \( F_t(x) \) for a given \( x \) and large \( t \). However, (1) can be misleading in cases when the convergence to the limit distribution is too slow. This is easily seen by inspecting the trajectories of the Hill estimator \( \hat{h}_{n,k}, k = 1, \ldots, n \), computed from samples drawn from the loggamma distribution \( F(x) \), see Figure 1. It is sometimes called the Hill horror plot, because of the important discrepancy between \( \hat{h}_{n,k} \) and the estimated parameter \( \gamma \). The explanation lies in the fact that the Hill estimator merely fits a Pareto distribution to the data thereby providing an approximation of the excess d.f. \( F_t \) rather than for \( \gamma \) itself. Despite these evidences the problem of estimating the excess d.f. \( F_t \) regardless of the limit \( P_\gamma \) is less studied in the literature.

Figure 1: 1 – The Hill estimator for loggamma d.f. with rate parameter 1 and shape parameter 2. 2 – Index of regular variation \( \gamma = 1 \) which is expected to be estimated. 3 – The fitted Pareto parameter \( \theta_t(F) \). Left: 1 realization; Right: 100 realizations.

In the first part of the presentation we shall consider the problem of recovering the excess d.f. \( F_t \) from the data \( X_1, \ldots, X_n \) directly, circumventing estimation of the tail index \( \gamma \). We shall also propose a new data driven choice of the location of the tail \( t \) based on a sequential lack-of-fit testing procedure.

In the second part we will give an extension of this approach to the problem of estimation of extreme conditional quantiles and give an application to the measurement of the activity of bivalves.