On the optimal prediction of extreme events

Abstract

The prediction of the extreme values of Y in terms of a vector of covariates X is a fundamental problem. Doing so optimally is particularly important in a regime of extremes where the events $\{Y > y_0\}$ are rare and the data are scarce. We start with a Neyman-Pearson type perspective to optimality, characterize the optimal predictors in terms of density ratios and discuss several examples. This leads us to the natural notion of *optimal extremal predictors* $h(X)$ as ones that maximize the tail-dependence coefficient

$$\lambda(Y, h(X)) := \lim_{p \to 1} \mathbb{P}[Y > F_Y^{\leftarrow}(p) \,|\, h(X) > F_{h(X)}^{\leftarrow}(p)],$$

over a class of functions $\mathcal{H} \ni h$.

Next, we focus on the special case where (Y, X) are (jointly) multivariate regularly varying and characterize the optimal predictors in the class of non-negative continuous homogeneous functions $h : \mathbb{R}^d \to \mathbb{R}_+$. This amounts to solving a calculus of variations problem for an integral functional of the spectral (i.e., angular) measure, which we solve in the general, spectrally continuous case. In the special, spectrally discrete case, we encounter an interesting *perfect asymptotic precision* phenomenon, where *for some models* $\lambda^{(\mathrm{opt})}(Y, X) := \lambda(Y, h^{(\mathrm{opt})}(X)) = 1$. Notably, in many linear models that share the same sets of heavy-tailed factors, somewhat contrary to intuition, the extreme events can be predicted with an asymptotically perfect precision.

The two characterizations lead to two types of statistical estimators of the optimal predictors, which we briefly illustrate with simulations and apply to the prediction of extreme solar flares.