CHAPTER 7

CONCLUSIONS

7.1 Discussion

The systematic study of the extreme values of any specific data-set has been the central aim of this thesis. This kind of study has turned out to be of crucial importance in many fields of human life (ranging from hydrology up to finance and insurance). An area where extreme-value analysis has recently been applied to is teletraffic engineering. Since world-wide web and its services have entered our lives, extreme-value analysis proves to be a useful tool in evaluating and, consequently, improving their performance. Extreme-value techniques and tools that are discussed in this thesis are applied (in chapter 6) in such a type of data-set (size, in bytes, requested from a particular web site).

In Statistics, the organized study of any figure is accomplished through a modelling procedure, i.e. by trying to come up with a proper model that adequately describes the phenomenon under investigation. In the present thesis, we have presented several approaches adopted for modelling extreme values. These can be essentially distinguished into two main categories : parametric and semi-parametric methods. We put special emphasis on semi-parametric methods since theses methods are more easily applied and are less 'demanding of data'.

As we have repeatedly mentioned, in the semi-parametric approach, the behaviour of extreme values is expressed by a single index, the so-called extreme-value index. Our efforts are concentrated on estimating this index. In chapter 4 several such estimators are provided (Hill, Moment, Pickands, among others). In chapter 5 we also presented some modifications of these (based on smoothing and robustifying procedures), since the dependence of these estimators on the very extreme observations which can display very large deviations is one of their drawbacks.

The simulation study for the comparison of their performance led to some very interesting results. The first is that, as one could naturally expect, the performance of estimators on a specific data-set depends on the distribution of the data. So, there is not a uniformly best estimator. Nevertheless, by looking more carefully at the results, some general conclusions may be reached. More specifically, in cases of long-tailed data (with an infinite upper end-point) Moment and Moment-Ratio estimators seem to estimate more satisfactorily the non-negative extreme-value index γ. However, when it comes to upper-bounded distributions (characterized by a negative value of γ) Peng's and Moment estimators are more preferable. As far as the impact of smoothing (averaging) procedures is concerned, we deduced that it is effective (improving the performance of standard estimators) in cases where the true value of extreme-value index is non-positive. Particularly, mean-averaging procedures improve greatly the performance of Pickands estimator (in case of zero γ), while median-averaging of Moment and Peng's also leads to improved estimators (for γ negative).

7.2 Open Problems

As we have stated above there is not a uniformly best estimator of extreme-value index. On the contrary, the performance of estimators seems to depend on the distribution of data in hand. From another point of view, one could say that the performance of estimators of extreme-value index depends on the value of the index itself. So, before proceeding to the use of any estimation formula it would be useful if we could get an idea about the range of values where the true γ lies in. This can be achieved graphically via QQ and mean excess plots as we have illustrated in chapter 6. Alternatively, there exist statistical tests which tests such hypothesis. See, for example, Hosking (1984), Hasofer and Wang (1992), Alves and Gomes (1996) and Marohn (1998).

Moreover, it should be pointed out that among the averaged estimators used in the simulation study only mean-averaged Hill and Moment estimators have been theoretically explored by Resnick and Stărică (1997, 1999). As we have seen, the median averaging procedure has also displayed some interesting effectiveness, implying that it may be worthy to be also studied theoretically (with special emphasis on Moment and Peng's estimators). The same holds for mean-averaged Pickands estimator.

However, the 'Achilles heel' of semi-parametric estimators of extreme-value index is its dependence and sensitivity on the number k of upper order statistics used in the estimation. No hard and fast rule exists. Usually the scientist is subjectively deciding on the number k used, by looking at appropriate graphs. More objective ways for doing so have been presented in chapter 5 (using regression or bootstrap). The bootstrap approach is a newly suggested and promising method in the filed of extreme-value analysis. Another area of extreme-value index estimation where bootstrap methodology could turn out to be very useful is in the estimation (and, consequently, elimination) of the bias of extreme-value index estimators. The biasness is inherent in all of these estimators, but it is not easily to be assessed theoretically because it depends on second order conditions on the underlying distribution of data, which are usually unverifiable. Bootstrap procedures could approximate the bias without making any such assumptions.

Finally, we should mention that a new promising branch of extreme-value analysis is that of multivariate extreme-value analysis. One of the problems faced in extreme-value analysis is that, usually, we deal with a few number of data which leads to great uncertainty. This drawback can be, somehow, rendered by the simultaneous use of more than one source of information (variables), i.e. by applying multivariate extreme-value analysis. Such an approach is attempted by Embrechts, de Haan and Huang (1999). This technique has already been applied to the field of hydrology. See, for example, de Haan and de Ronde (1998), de Haan and Sinha (1999) and Barão and Tawn (1999).