



## Tail Index Estimation Based on Survey Data

Patrice Bertail\*

MODAL X, Université Paris-Ouest, Nanterre, France [patrice.bertail@gmail.com](mailto:patrice.bertail@gmail.com)

Emilie Chautru

Mines de Paris, Fontainebleau, France [emilie.chautru@gmail.com](mailto:emilie.chautru@gmail.com)

Stephan Cléménçon

Institut Telecom - LTCI UMR Telecom ParisTech, Paris, France

[Stephan.clemencon@telecom-paristech.fr](mailto:Stephan.clemencon@telecom-paristech.fr)

This paper is devoted to tail index estimation in the context of survey data. Assuming that the super-population of interest is described by a heavy-tailed statistical model, we prove that the survey scheme plays a crucial role in the design of consistent inference methods for extremes. As can be revealed by simulation experiments, ignoring the sampling plan generally induces a significant bias, jeopardizing the accuracy of the extreme value statistics thus computed. Focus is here on the celebrated Hill method for tail index estimation; it is shown how to modify it in order to take into account the survey design. Precisely, under specific conditions on the inclusion probabilities of first and second orders, we establish the consistency of the variant of the Hill estimator we propose. Additionally, its asymptotic normality is proved in some specific sampling plans. We first prove its asymptotic normality for Poisson sampling plans, then for conditional sampling plans including rejective sampling plan. Application of this limit result for building Gaussian confidence intervals is thoroughly discussed and illustrated by numerical results.

**Keywords:** Survey sampling, tail index estimation, Hill estimator, Poisson survey scheme, rejective sampling.