

# PROCESSUS EMPIRIQUES DANS LE CADRE DES SONDAGES

Emilie Chautru <sup>1</sup> & Patrice Bertail <sup>2</sup> & Stéphan Cléménçon <sup>3</sup>

<sup>1</sup> *Institut Télécom, LTCI UMR Télécom ParisTech/CNRS No. 5141, 37/39 rue Dareau, 75014 Paris, chautru@telecom-paristech.fr*

<sup>2</sup> *MODAL X, UFR SEGMI, Université Paris X, 200 avenue de la République, 92 001 Nanterre, pbertail@u-paris10.fr*

<sup>3</sup> *Institut Télécom, LTCI UMR Télécom ParisTech/CNRS No. 5141 37/39 rue Dareau, 75014 Paris, clemenco@telecom-paristech.fr*

Le présent travail propose d'étudier l'impact des plans de sondage sur les processus empiriques. En nous plaçant dans le cadre des modèles de surpopulation, nous considérons une variable aléatoire d'intérêt dont les réalisations dans la population sont indépendantes et de même loi de probabilité. Nous nous intéressons alors aux fonctionnelles de l'estimateur de type Horvitz-Thompson de cette mesure de probabilité, et établissons un ensemble de conditions sur le plan de sondage utilisé permettant leur validité asymptotique.

Notre approche s'inspire des travaux de Saegusa et Wellner (2011), qui exhibent un ensemble d'hypothèses sous lesquelles de tels processus sont convergents dans le cas précis des plans de sondage stratifiés, potentiellement suivis d'une phase de calibration des poids individuels. Cherchant à généraliser ces résultats à d'autres plans de sondages, nous nous intéressons tout particulièrement aux échantillonnages poissonien, réjectif et séquentiel, tels qu'ils ont été étudiés par Hájek (1964) et Berger (1998). En généralisant leur approche, nous établissons des théorèmes de type Glivenko-Cantelli et Donsker pour les processus empiriques de type Horvitz-Thompson. Nous discutons ensuite l'impact d'une phase supplémentaire de calibration sur le comportement asymptotique des ces processus.

Nous présentons enfin une application de nos travaux à l'étude des valeurs extrêmes. L'objectif principal de cette branche de la statistique est d'estimer des probabilités très faibles correspondant à des événements rares, allant potentiellement au delà de l'observé. Elle utilise les propriétés asymptotiques du maximum d'un échantillon, et étudie dans le cas univarié l'épaisseur des queues de distribution. Cette dernière est déterminée par un indice, appelé indice de valeurs extrêmes, que nous souhaitons estimer. S'il existe d'ores et déjà pléthore d'estimateurs de cette quantité (cf. Beirlant *et al.*, 2004), aucun d'entre eux ne tient compte de la méthode d'échantillonnage utilisée pour la récolte des données. Nous introduisons ainsi un nouvel estimateur de l'indice de valeurs extrêmes pour les distributions à queue épaisse, intégrant les poids individuels issus du plan de sondage utilisé, et dont la validité asymptotique est garantie par les résultats obtenus sur les processus empiriques.

## Bibliographie

Beirlant, J., et Goegebeur, Y., et Segers, J., et Teugels, J. (2004). *Statistics of Extremes : Theory and Applications*, Wiley.

Berger, Y.G. (1998). Rate of convergence to normal distribution for the Horvitz-Thompson estimator, *Journal of Statistical Planning and Inference*, 67, 209-226.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population, *Annals of Mathematical Statistics*, 1491-1523.

Saegusa, T., et Wellner, J.A. (2011). Weighted likelihood estimation under two-phase sampling, rapport non publié (Arxiv preprint arXiv :1112.4951).