

# PLAN DE SONDAGE INFORMATIF, DISTRIBUTION PONDÉRÉE ET MAXIMUM DE VRAISEMBLANCE

Daniel Bonn ery <sup>1</sup> & Jay Breidt <sup>2</sup> & Fran ois Coquet <sup>1</sup>

<sup>1</sup> *Ensaï, rue Blaise Pascal, 35172 BRUZ*

*daniel.bonnery@ensai.fr, fcoquet@ensai.fr*

<sup>2</sup> *Colorado State University, 102 Statistics Building Ft. Collins, CO 80523, USA*

*jbreidt@stat.colostate.edu*

Consid erons le tirage d’un  chantillon d’une population finie de taille  $N$  selon un plan de sondage informatif, et le mod le de superpopulation suivant :   chaque  l ment  $k$  de la population, correspond la r alisation d’une variable al atoire, not e  $Y_k$ , les r alisations sur la population sont suppos es ind ependantes et identiquement distribu es (iid) selon une loi qui admet une densit   $f_\theta$  par rapport   la mesure de Lebesgue. Le plan de sondage est informatif dans le sens o  le vecteur des r alisations qui correspondent aux  l ments de l’ chantillon n’est pas un vecteur de variables al atoires ind ependantes, et la loi d’une r alisation conditionnelle   la s lection de l’ l ment correspondant n’est pas  gale   la loi initiale des r alisations sur la population. Une loi de probabilit  limite et une densit  de probabilit  limite des r alisations sur l’ chantillon sont d finies ; elles correspondent   la limite de la distribution d’une r alisation sur l’ chantillon lorsque les tailles de la population et de l’ chantillon tendent vers l’infini. La densit  de la distribution limite de l’ chantillon est une version pond r e, not e  $\rho_\theta f_\theta$  de la densit  initiale  $f_\theta$ . Si on note  $J_k$  le nombre de s lections de l’individu  $k$ , on a :

$$\rho_\theta(y) = \lim_{N \rightarrow \infty} \frac{\mathbb{E}_\theta [J_k \mid Y_k = y]}{\mathbb{E}_\theta [J_k]}.$$

En g n ral, le processus de s lection induit une d pendance entre les r alisations correspondant aux  l ments s lectionn s. Un cadre asymptotique et des conditions faibles sur le processus de s lection sont donn s, qui permettent de consid rer que les r alisations sur l’ chantillon sont iid et de densit   $\rho_\theta f_\theta$  approximativement. Nous montrons les propri t s des estimateurs du maximum de la vraisemblance obtenue   partir de cette approximation :

$$\prod_{k=1}^N (\rho_\theta(y) f_\theta(y))^{J_k}.$$

## Bibliographie

Bonn ery, D., et Breidt, F. J., et Coquet, F. (2012). Uniform convergence of the empirical cumulative distribution function under informative selection from a finite population, Bernoulli,   para tre.

Pfeffermann, D., et Krieger, A. M., et Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling, Statistica Sinica, 8, 1087-1114.