

# ESTIMATEURS MODEL-BASED NON PARAMÉTRIQUES DE LA FONCTION DE RÉPARTITION D'UNE VARIABLE CENSURÉE À DROITE SUR PETITS DOMAINES

Sandrine Casanova <sup>1</sup> & Ève Leconte <sup>1</sup>

<sup>1</sup> *TSE (Gremaq), Université Toulouse 1 Capitole,  
21, allée de Brienne, 31000 TOULOUSE, France*  
sandrine.casanova@univ-tlse1.fr, eve.leconte@tse-fr.eu

L'estimation de la fonction de répartition (f.d.r.) en population finie est très utile pour déduire des estimateurs de paramètres complexes tels que les quantiles. Nous considérons ici le cas où la variable d'intérêt est censurée à droite, cas qui, à notre connaissance, n'a jamais été étudié dans le cadre des sondages. Dans ce contexte, disposant d'une information auxiliaire fournie par une covariable, nous proposons de nouveaux estimateurs *model-based* non paramétriques de la f.d.r. sur la population : la variable d'intérêt d'un individu hors échantillon est prédite par un estimateur de la médiane conditionnelle sachant la covariable. Cet estimateur de la médiane est obtenu par inversion de l'estimateur de Kaplan-Meier généralisé lissé (Leconte *et al.*, 2002) de la f.d.r. conditionnelle.

Nous nous sommes également intéressées à l'estimation de la f.d.r. sur un domaine. Si ce domaine est de taille suffisante, l'estimation des paramètres d'intérêt est basée sur les données relatives aux individus du domaine et les estimateurs produits sont de précision acceptable. Cependant, dans la plupart des applications, les tailles d'échantillons correspondant à des petits domaines ne sont pas suffisantes. De l'information doit alors être "empruntée" aux autres domaines pour améliorer la précision. Pour cela, nous avons adapté au cas censuré la technique de Casanova (2012) : la variable d'intérêt d'un individu hors échantillon est prédite en estimant un quantile conditionnel dont l'ordre doit être préalablement estimé. Les estimations de ces deux étapes utilisent les échantillons de tous les domaines afin d'emprunter de la force aux voisins.

Pour obtenir des estimations du biais et de la variance de l'erreur de prédiction des nouveaux estimateurs, nous avons généralisé au cas censuré la méthode de rééchantillonnage par bootstrap proposée par Lombardia *et al.* (2004). Nous obtenons également par cette méthode des intervalles de confiance de la f.d.r.

Des simulations montrent le gain apporté par les nouveaux estimateurs par rapport à l'estimateur de Kaplan-Meier calculé sur les points échantillonnés. Un exemple d'application à des données de durées de chômage illustre la méthode.

## Bibliographie

Casanova, S. (2012). Using Nonparametric Conditional M-quantiles to Estimate a Cumulative Distribution Function in a Domain, *Annales d'Economie et de Statistique*, à paraître.

Leconte, E., Poiraud-Casanova, S. et Thomas-Agnan, C. (2002). Smooth Conditional Distribution Function and Quantiles under Random Censorship, *Lifetime Data Analysis*, 8, 229-246.

Lombardía, M. J., González-Manteiga, W., Prada-Sánchez, J. M. (2004). Bootstrapping the Dorfman-Hall-Chambers-Dunstan estimator of a finite population distribution function, *Journal of Nonparametric Statistics*, 16, 63-90.