

# CALAGE SUR COMPOSANTES PRINCIPALES. APPLICATION AUX DONNÉES MÉDIAMÉTRIE

Camelia Goga<sup>1</sup> & Ahmed Muhammad Shehzad<sup>1</sup> & Aurélie Vanheuverzwyn<sup>3</sup>

<sup>1</sup> *IMB, Université de Bourgogne, 9 avenue Alain Savary, DIJON, France*  
camelia.goga@u-bourgogne.fr, Muhammad-Ahmed.Shehzad@u-bourgogne.fr

<sup>2</sup> *Médiаметrie, 70 rue Rivay, 92532 LEVALLOIS CEDEX,*  
avanheuverzwyn@mediametrie.fr

L'information auxiliaire est utilisée dans les sondages pour améliorer la précision des estimateurs. Cette information peut être utilisée lors de la phase de l'échantillonnage ou de la phase de l'estimation. Le calage (Deville and Särndal, 1992) est une des méthodes très utilisée en pratique et consiste à déterminer les poids qui sont le plus proche possible (au sens d'une certaine distance) des poids de sondage et sous la contrainte que les totaux de variables de calage soient exactement estimés.

Lorsque le nombre de variables est très grand, les poids de calage peuvent être négatifs ou extrêmement larges et l'estimation basée sur ces poids devient très instable. Plus que ça, des problèmes de collinéarités entre les variables de calage peuvent apparaître.

Nous proposons dans ce travail, une nouvelle classe d'estimateurs obtenus par calage sur un petit nombre de nouvelles variables, les composantes principales de la matrice de variables de calage. Le calcul de ces variables nécessite une information complète, c'est à dire que les variables de calage soient connues pour chaque individu de la population. Nous étudions les propriétés asymptotiques de l'estimateur de calage ainsi obtenu. La connaissance des valeurs singulières de la matrice d'information auxiliaire permet de réaliser un calage sur le moment d'ordre deux des composantes principales.

Lorsque les variables de calage de départ ne sont connues que dans l'échantillon tout en connaissant leurs totaux dans la population, nous proposons d'estimer les composantes principales dans l'échantillon et déterminer les poids par un calage sur les composantes principales ainsi estimées.

Nous illustrons les bonnes propriétés des estimateurs proposés sur un jeu de données réelles issu d'une enquête Médiamétrie.

## Bibliographie

Deville J.-C., et Särndal, C.-E. (1992). Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association*, 87, 376-382.

Goga, C., et Shehzad, M.-A., et Vanheuverzwyn, A. (2011). Principal Component Regression with Survey Data. Application on the French Media Audience, *Proceedings of the 58th ISI World Statistics Congress - Dublin 2011*.