

PRÉDICTEUR EMPIRIQUE DE BAYES BASÉ SUR DES COPULES ARCHIMÉDIENNES POUR DES PROPORTIONS RÉGIONALES

Fodé Tounkara ¹ & Louis-Paul Rivest ²

¹ *fode.tounkara.1@ulaval.ca*

² *Louis-Paul.Rivest@mat.ulaval.ca*

Résumé.

Les données d'enquête par sondage sont utilisées pour obtenir des estimations des caractéristiques telles que la moyenne et la proportion liées à la population, mais aussi à des sous populations appelées petits domaines.

Il y a une demande mondiale de plus en plus croissante d'estimations fiables pour des petits domaines. Le besoin se fait sentir aussi bien dans les secteurs publics que dans les secteurs privés. Il existe de multiples utilisations, par exemple, l'affectation des fonds publics, les stratégies de planification régionale, et les décisions de marketing au niveau local. Quelques exemples importants comprennent le taux de chômage, la proportion de personnes sans assurance maladie pour les petits domaines, voir Rao (2003) pour plus d'exemples.

Les estimateurs directs, basés uniquement sur les données d'échantillons spécifiques aux domaines peuvent être adéquats pour les grands domaines, mais ils sont généralement accompagnés de grandes variances et de grands coefficients de variation, en raison de la faible taille de l'échantillon qui recoupe ces domaines. La principale raison est que les enquêtes originales ont été conçues pour atteindre un niveau précis de précision pour les régions ou domaines plus importants, et malheureusement on ne dispose pas des ressources nécessaires pour mener à bien de nouvelles enquêtes afin d'atteindre le niveau de précision souhaité pour ces petites régions. Il est alors nécessaire d'utiliser des modèles qui permettront de connecter différents petits domaines en utilisant des informations auxiliaires pertinentes venant de sources exhaustives ou issues d'enquêtes par sondage de grande taille.

Ici, nous considérons la situation où la variable réponse mesuré sur l'unité j du domaine i , soit Y_{ij} , est binaire prenant les valeurs 1 (succès) et 0 (échec), et que la caractéristique d'intérêt pour le petit domaine est une proportion.

Dans ce contexte, nous proposons une nouvelle classe de modèles probabilistes pour des données de Bernoulli échangeables. Dans ces modèles, la probabilité conditionnelle de succès est une fonction de la probabilité marginale de succès et d'un effet aléatoire positif spécifique à chaque grappe. Les probabilités marginales de succès sont modélisées en utilisant les fonctions de lien Logit et Log-Log complémentaire. La distribution de l'effet aléatoire contient un paramètre d'association qui est estimé pour donner une mesure de

la force de la dépendance résiduelle ignorée par les marges. Nous distinguons deux types de modèles :

(i) Modèle 1 : le paramètre de Bernoulli est sensible aux domaines :

$$p_i = e^{-a_i \psi_\alpha^{-1}(\pi_i)}, \quad \pi_i = \frac{e^{x_i \beta}}{1 + e^{x_i \beta}},$$

(ii) Modèle 2 : le paramètre de Bernoulli est sensible aux individus dans les domaines :

$$p_{ij} = e^{-a_i \psi_\alpha^{-1}(\pi_{ij})}, \quad \pi_{ij} = \frac{e^{x_{ij} \beta}}{1 + e^{x_{ij} \beta}},$$

où a_i est une variable aléatoire positive admettant une transformée de Laplace ψ_α , et ψ_α^{-1} dénote l'inverse de ψ_α ; α est le paramètre de dépendance résiduelle et β est le vecteur des paramètres de régression.

Nous montrons que la transformée de Laplace de l'effet aléatoire a_i est liée au générateur des modèles de copules Archimédiennes, Nelson (2006) et Mai et Scherrer (2012), voir aussi Tounkara et Rivest (2014) et Tounkara et Rivest (2015), pour plus de discussions concernant les copules Archimédiennes sous-jacentes aux modèles proposés.

Nous utilisons des techniques de Bayes basées sur ces modèles pour estimer des proportions régionales. L'inférence Bayésienne que nous proposons consiste à trouver la distribution a posteriori de l'effet aléatoire et sa transformée de Laplace sachant les données et les paramètres du modèle. Cette transformée de Laplace est ensuite utilisée pour trouver des estimateurs de Bayes et leurs variances a posteriori pour les vraies proportions. Nous développons une étude de comparaison entre le meilleur prédicteur de Bayes (BP) et le meilleur prédicteur linéaire sans biais (BLUP). Nous avons également étudié l'efficacité du BP obtenu sous nos modèles relativement au BLUP. Les paramètres du modèle sont estimés en utilisant la méthode du maximum de vraisemblance. Nous obtenons un prédicteur empirique de Bayes (EBP) en remplaçant les paramètres par leurs estimateurs dans l'expression du BP. Nous utilisons le critère d'information d'Akaike (AIC) pour la sélection d'un modèle de copule. Nous utilisons la méthode du Bootstrap paramétrique pour estimer l'erreur quadratique moyenne des prédicteurs empiriques. Des résultats empiriques obtenus à partir de données simulées et réelles seront également présentés.

Mots-clés. petit domaine, proportions, copules Archimédiennes multidimensionnelle, techniques de Bayes empiriques, Bootstrap paramétrique.

Bibliographie

[1] Mai, J.-M. et Scherer, M. (2012), *Simulating copulas; Stochastic Models, Sampling Algorithms and Application. Series in Quantitative Finance : vol. 4.*, World Scientific Publishing company, London.

- [2] Nelson, R., B. (2003), *An introduction to copula*, Springer-Verlag, New York.
- [3] Rao, J., N., K. (2003), *Small Area Estimation*, Wiley Series in Survey Methodology, New York.
- [4] Tounkara, F. et Rivest, L.-P. (2014), *Some random effect models for correlation binary response*, Dependence Modeling, 2, 73–87.
- [5] Tounkara, F. et Rivest, L.-P. (2015), *Mixture Regression Models for Closed Population Capture-Recapture Data*, Biometrics, 71, 721–730.