

COMMENT CONSERVER L'ÉQUILIBRE DANS UN SONDAGE : LA QUÊTE DU GRAAL ET LA SUITE

Jean-Claude Deville ¹

¹ *Ensayi, LSE/Crest, Campus de Ker-Lann, 35170 Bruz, deville@ensai.fr*

Je vais vous raconter quelques souvenirs sur la façon dont j'ai vécu avec divers partenaires (dont certains uniquement livresques) cette quête du Graal qu'est l'échantillonnage équilibré.

Elle commence en 1981 avec une perspective d'avoir à m'intéresser aux sondages à l'Insee et la publication du livre de Hájek (1981). Elle se poursuit par plusieurs épisodes ponctués par quelques publications, des tas de brouillons jamais finis et des rencontres. Avec aussi des défis comme la mise au point de l'échantillonnage au vingtième pour le recensement de la population de 1990.

Elle culmine avec la découverte de la méthode du cube, en 1997, son utilisation pour le recensement rénové de la population et pratiquement toutes les enquêtes de l'Insee depuis le début des années 2000, et la publication des articles de références en 2004 et 2005. La méthode "fast-cube" de Chauvet-Tillé (2006) apportait une amélioration majeure tant sur le plan pratique que théorique.

Restait (au moins) un os : on ne savait pas dans quels cas on pouvait arriver à un équilibrage exact. Cela avait même servi de prétexte à un arbitre resté anonyme pour refuser la publication d'un article qui a bien fini par être accepté ailleurs, mais ça a pris des années ! Bien que j'aie annoncé le résultat depuis longtemps, j'ai buté pendant 10 ans sur une difficulté de démonstration. Comme ma dernière idée (du 10 septembre) a vraiment l'air de marcher mieux que les autres, j'en parlerai un peu. Elle clôt le problème initial, mais ouvre des perspectives étonnantes en statistique, en théorie des codes, et peut-être en théorie des groupes discrets et en physique théorique.

Bibliographie

- Chauvet, G., et Tillé, Y. (2006). A fast algorithm for balanced sampling, *Comp. Statistics*, 21, 53-62.
- Deville, J. C. (1991). Une théorie des enquêtes par quota, *Techniques d'Enquête*, 17, 163-181.
- Deville, J.-C. (1992). Constrained samples, conditional inference, weighting : three aspects of the utilization of auxiliary information. In *Proceedings of the Workshop on the Use of Auxiliary Information in Surveys*, 21-40, Örebro, Sweden : Statistics Sweden.
- Deville, J.-C., et Grosbras, J.-M., et Roth, N. (1988). Efficient sampling algorithms and balanced sample, *COMPSTAT, Proceeding in Computational Statistics*, Ed.R. Payne and P. Green, 255-66, Heidelberg. Physica Verlag.
- Deville, J. C., et Tillé, Y. (1988). Unequal Probability Sampling without Replacement through a Splitting Method, *Biometrika*, 85, 89-101.
- Deville, J. C., et Tillé, Y. (1999). Balanced Sampling : Geometric Representation & Variance and Variance Estimation, 51ème session de l'Institut International de Statistique, Helsinki.
- Deville, J. C., et Tillé, Y. (2004). Efficient balanced sampling : The cube method, *Biometrika*, 91, 893-912.
- Deville, J. C., et Tillé, Y. (2005). Variance approximation under balanced sampling, *Journal of Statistical Planning and Inference*, 128, 411-425.
- Hájek, J. (1981). *Sampling from a finite population*, New York : Marcel Dekker.
- Kiaer, A. (1896). Observations et expériences concernant des dénombrements représentatifs. *Bull. Inst. Int. Statist.* 9, 176-83.
- Royall, R., et Herson, J. (1973). Robust estimation in finite populations I, *Journal of the American Statistical Association*, 68, 880-889.
- Thionnet, P.(1953). *La Théorie des Sondages*, Paris : INSEE, Imprimerie Nationale.