

Une nouvelle approche du sondage aléatoire simple

Martin Körnig
m.kornig@wanadoo.fr

Résumé

Une approche bayésienne du sondage aléatoire simple offre des solutions simples, pratiques et relativement faciles à exploiter numériquement. Il s'agit de solutions analytiques (loi de probabilité, fonction de répartition, espérance, variance) permettant de prédire, pour une classe donnée, le nombre de représentants dans la population.

Mots-clés

sondage aléatoire simple, statistique bayésienne

Abstract

A bayesian approach to statistical surveys using random sampling leads to simple and practical solutions that are relatively easy to numerically implement. The given analytical solutions (probability function, cumulative distribution function, expected value, variance) enable the prediction of the number of individuals in the population that belong to a given class.

Key words

statistical survey, random sampling, bayesian statistics

1. Introduction

Je reprends un problème « classique » : le sondage aléatoire simple (voir par exemple [1]). Sondage où

- les individus sont classés par rapport à un caractère (qualitatif) en plusieurs classes distinctes.
- l'objectif est de prédire, pour chaque classe, le nombre de représentants dans la population en analysant un échantillon aléatoire.

Pour un tel sondage une théorie bayésienne existe [2] (voir également [3] et [4] qui ne traitent que le cas de deux classes). Dans le cas de trois ou davantage de classes, l'application de cette théorie demandait jusqu'ici une intégration numérique importante. Le présent article enlève cette difficulté en donnant des solutions analytiques à la section 5.

Le Tab. 1 rassemble les notations principales adoptées dans cet article.

Pour une classe i quelconque, le nombre n_i (voir Tab. 1) n'est généralement pas connu avec exactitude - ni avant ni après le sondage. Il s'agit d'une variable aléatoire, dont on cherche à spécifier la loi de probabilité (réponse à la section 5.1).

Dans le cas théorique $K = N$, il n'y a pas vraiment de problème de prédiction puisque la population entière a été échantillonnée. C'est le seul cas, d'ailleurs, où les nombres n_1, n_2, \dots, n_I sont connus exactement. L'ensemble de la théorie exposée dans les sections 3 et 5 reste néanmoins valable dans ce cas là.

L'article continue avec une discussion sur l'importance du nombre de classes I (section 6) et une suggestion concernant l'information a priori (section 7).