

Echantillonnage multidimensionnel (= de plusieurs échantillons à la fois) à entropie maximum : définition, propriétés, algorithmes et programmes.

Jean-Claude DEVILLE¹ et Lionel QUALITE²

¹CREST/ENSAI

²CREST/ENSAI puis Université de Neuchâtel

On désire tirer dans une même population de taille N , Q échantillons disjoints en même temps (et non pas successivement !) ayant des probabilités d'inclusion π_k^q ($q=1$ à Q , $k=1$ à N) fixées et de tailles fixées n_q . Même dans le cas $Q=2$, les choses sont loin d'être évidentes si on désire conserver une certaine symétrie au problème.

Il se trouve que dans le cas où on cherche un échantillon Q -dimensionnel vérifiant la propriété d'entropie maximum ($-\sum_{s_q \text{ disjoint}} p(s_1, \dots, s_Q) \log(p(s_1, \dots, s_Q))$), le problème peut

être vu comme un échantillonnage de Poisson multidimensionnel conditionnel, et que de bonnes propriétés conditionnelles apparaissent. En particulier, il est possible de calculer par des récurrences multiples les probabilités d'inclusion aux ordres 1 et 2 conditionnellement à des contraintes de domaine et de tailles. Le calcul de paramètres des échantillons de Poisson 'sous-jacents' est possible avec des charges de calcul acceptables si les dimensions du problème ne sont pas trop excessives.

On décrit les algorithmes et les programmes (MATLAB) permettant ces calculs.

Une application possible est l'imputation d'une variable qualitative .