Nombre de participants

F. Kany. ISEN-Brest & La Croix-Rouge

Présentation

Soit un événement (sportif par exemple) avec N participants : chaque participant a un numéro de carte d'invitation (ou un numéro de dossard). On prend un échantillon de $n < N$ personnes et on relève leur numéro x_i (avec $i \in [1, n]$). On cherche à estimer N à partir des valeurs des x_i.

On peut montrer que l'espérance du maximum d'un échantillon est : $E(\max x_i) = \frac{n.(N+1)}{n+1}$ d'où : $\boxed{N = \frac{n+1}{n}.E(\max x_i) - 1.}$

Cette formule a le bon goût de vérifier deux propriétés essentielles :

— si on tire comme échantillon $[1, 2, 3, \ldots, n]$, on obtient $N = \frac{n+1}{n}.n - 1 = n$: l'estimation de N n'est pas inférieure à la plus grande valeur de l'échantillon (toutes les formules d'estimation de l'espérance ne vérifie pas cette propriété triviale !)

— si on tire comme échantillon tout l'ensemble $[1, 2, 3, \ldots, N]$, on obtient $N = \frac{N+1}{N}.N - 1 = N$.

Faire plusieurs simulations numériques en prélevant des échantillons de 2%, 5%, 10% et 20% de la population. Présenter les résultats sous forme d'histogrammes pour montrer la probabilité de l'écart (en %) entre la taille réelle de la population et son estimation.

Pour prélever (sans remplacement) un échantillon de taille n dans une population de taille N, on utilise la procédure suivante :

— On considère le premier élément de la population

— On le prélève avec une probabilité de $\frac{n}{N}$ pour le mettre dans l'échantillon

— Si l'élément est effectivement sélectionné, alors on diminue n de 1.

— On diminue N de 1

— On recommence avec l'élément suivant de la population jusqu'à ce que la taille de l'échantillon soit n.