



HAL
open science

ÉCHANTILLON REPRÉSENTATIF (D'UNE POPULATION FINIE) : DÉFINITION STATISTIQUE ET PROPRIÉTÉS

Léo Gerville-Réache, Vincent Couallier

► **To cite this version:**

Léo Gerville-Réache, Vincent Couallier. ÉCHANTILLON REPRÉSENTATIF (D'UNE POPULATION FINIE) : DÉFINITION STATISTIQUE ET PROPRIÉTÉS. 2011. hal-00655566

HAL Id: hal-00655566

<https://hal.science/hal-00655566>

Preprint submitted on 30 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ECHANTILLON REPRESENTATIF (D'UNE POPULATION FINIE) : DEFINITION STATISTIQUE ET PROPRIETES

Léo Gerville-Réache^{1,2}, Vincent Couallier^{1,2} & Nicolas Paris³

1. Université de Bordeaux 2, Bordeaux, F-33000, France

2. CNRS, UMR 5251, Bordeaux, F-33000, France

3. Optima-europe, Mérignac, France

Résumé : La notion de représentativité d'un échantillon est sous-jacente à la plupart des études par sondage. L'idée, furtive, est que l'échantillon doit produire des résultats « extrapolables » à la population d'intérêt. La formalisation de cette « idée » : définir la notion d'échantillon représentatif, est une quête que certains poursuivent ardemment et que d'autres ont abandonné. Nous proposons dans ce papier un regard sur cette notion qui débouche sur une définition statistique.

Mots clés : Echantillon représentatif, Sondage, Quotas, Probabilités d'inclusion.

1. Introduction

Qu'est ce qu'un échantillon représentatif, et pourquoi cette notion de représentativité est-elle un préalable à de nombreuses analyses de résultats de sondage ? Il est bien évident qu'une analyse par échantillonnage en vue de décrire, prédire ou extrapoler à la population entière est la base des statistiques, le problème étant de justifier les techniques de statistique mathématique associées. Avant même de parler de représentativité, on doit définir la notion de population, d'échantillonnage dans cette population, et enfin d'échantillon de cette population. Dans la suite, suivant les notations de Cochran (1977) ou Ardilly (2006), une population de taille finie N est classiquement définie comme un ensemble d'unités disjointes pouvant être indexées par les N premiers entiers et représentée par $P = \{u_i, i = 1, \dots, N\}$ sans distinction d'ordre. Une méthode d'échantillonnage est un algorithme permettant sans ambiguïté de créer un échantillon, c'est à dire de sélectionner sans distinction d'ordre une partie de P . Puisque les unités statistiques sont distinguables, l'échantillonnage revient à décrire comment sont sélectionnés les indices $\{i_1, \dots, i_n\}$ des n unités statistiques de la population. Ainsi, échantillonner dans P est équivalent à échantillonner dans $\{1, \dots, N\}$. Enfin, un échantillon $\{u_{i_1}, \dots, u_{i_n}\}$ est le résultat d'une méthode d'échantillonnage appliquée une fois sur la population P .

La représentativité d'un échantillon ne peut donc être envisagée (une fois définie) qu'en termes de qualité d'un échantillon provenant d'une méthode d'échantillonnage donnée. C'est donc la méthode d'échantillonnage qui donnera aux échantillons leur qualité de représentativité.

Dans cet article, nous proposons une définition d'un échantillon représentatif d'une population finie et démontrons quelques propriétés essentielles qui en découlent.

2. Notions de représentativité d'un échantillon dans la littérature.

La représentativité est souvent un argument avancé pour justifier l'usage des résultats d'une enquête par sondage. Il existe cependant de nombreuses interprétations de ce terme et nous avons relevé quelques argumentaires ou définitions de ce terme.

a. Aperçu de notions relevées dans la littérature.

Pour Yves Tillé (2001) « *Le concept de représentativité est aujourd'hui à ce point galvaudé qu'il est désormais porteur de nombreuses ambivalences. Cette notion, d'ordre essentiellement intuitif, est non seulement sommaire mais encore fautive et, à bien des égards, invalidée par la théorie.* ».

Pour Jean Vaillant (2005) « *La définition d'échantillon représentatif diffère selon que le plan d'échantillonnage est probabiliste ou non probabiliste :*

- *un plan probabiliste fournit un échantillon représentatif dès lors que chaque individu de la population a une probabilité connue et non nulle d'être inclus dans l'échantillon.*

- *un plan non probabiliste fournit un échantillon représentatif si la structure de l'échantillon pour certaines variables clés est similaire à celle de la population cible. Par exemple, on peut vouloir construire un échantillon pour lequel les proportions de catégories d'individus soient similaires dans l'échantillon à celles de la population cible (c'est le principe de la méthode dite des quotas).* ».

Pour Pascal Ardilly (2006) « *On dit qu'un plan de sondage est représentatif d'une expression donnée et numériquement connue construite à partir d'une variable auxiliaire (un total le plus souvent) lorsque, pour la méthode d'échantillonnage choisie, l'estimateur estime parfaitement bien (c'est-à-dire avec un biais nul et une variance nulle) l'expression donnée*

en question. Ce terme ne s'applique pas de façon satisfaisante à un échantillon. »

Pour Olivier Sautory (2010) « *Un échantillon n'est jamais représentatif "en soi", il est représentatif par rapport à certaines variables*».

Pour l'AFNOR et l'ISO : Pas de définition d'échantillon représentatif ou de représentativité dans l'ISO 3534-1-Vocabulaire et symboles- mais une définition du terme échantillon : « *Une ou plusieurs unités d'échantillonnage prélevées dans une population et destinées à fournir des informations sur cette population. Note – Un échantillon peut servir de base à une décision concernant cette population ou le processus qui l'a produite* ».

Pour le Sénat dans la proposition de loi du 14 février 2010 sur les sondages : « *Un sondage est une enquête statistique visant à donner une indication quantitative, à une date déterminée, des opinions, souhaits, attitudes ou comportements d'une population par l'interrogation d'un échantillon représentatif de celle-ci, qu'il soit constitué selon la méthode des quotas ou selon la méthode aléatoire* »

Pour Sheldon M. Ross (1999) « *[...] Thus, although it may seem paradoxical, we are most likely to obtain a representative sample by choosing its members in a totally random fashion without any prior considerations of the elements that will be chosen. In other words, we need not attempt to deliberately choose the sample so that it contains, for instance, the same gender percentage and the same percentage of people in each profession as found in the general population.*».

Enfin, pour Philippe Dutarte (2005) « *Voilà une expression qui, si elle n'est pas précisée, peut signifier à peu près n'importe quoi. Un échantillon constitué selon la méthode des quotas est évidemment « représentatif » des critères correspondants aux quotas (sexe, âge, catégorie socioprofessionnelle, région, taille de la commune...) selon lesquels il a été fabriqué. Mais on n'a aucun moyen de savoir jusqu'à quel point il est « représentatif » de ce pour quoi il a été prélevé, c'est-à-dire le sujet du sondage, l'opinion, le pourcentage que l'on cherche à évaluer. L'expression « représentatif de la population française », que l'on lit souvent dans la presse, prête évidemment à confusion. On a l'impression que l'échantillon est « représentatif » de tout ce que l'on veut. En statistique, on désigne plutôt par « échantillon représentatif », un échantillon où le hasard permet d'éviter les biais inconnus et d'appliquer le calcul des probabilités. La méthode optimale pour obtenir un échantillon « représentatif » est celle du sondage aléatoire stratifié optimal. ».*

Cet éventail de définitions, propriétés ou remises en cause de la représentativité d'un échantillon montre l'apparente subtilité et complexité de cette notion. Pourtant, cette notion nous semble fondamentale. Nous pourrions éprouver, voir unifier les points de vue dès lors qu'une définition statistique unique s'impose. Dans cet article, nous proposons une définition d'un échantillon représentatif d'une population finie et démontrons plusieurs propriétés essentielles qui en découlent. Deux idées fondamentales ont guidé notre réflexion :

- La population est un échantillon représentatif
- Un échantillon simple au hasard (équiprobable) est un échantillon représentatif.

b. Résultats représentatifs ou échantillons représentatifs ?

Un échantillon représentatif n'est pas une fin en soi. Ce que nous souhaitons c'est que les résultats issus du traitement statistique de l'échantillon puissent être « extrapolés » à la population. Aussi, ce que l'on souhaite c'est que les résultats soient « représentatifs ».

Un résultat représentatif est une idée assez claire. En tout cas, nous pouvons nous accorder sur une définition statistique en ce qui concerne les résultats issus des méthodes d'échantillonnage probabilistes. Un résultat sur l'échantillon doit estimer une quantité dans la population. Ce que l'on souhaite clairement, d'un point de vue statistique, c'est que cette estimation soit sans biais. C'est sur cette idée qu'est construite la théorie des sondages (voir par exemple Ardilly (2006)).

Un résultat représentatif est donc, statistiquement, un résultat issu d'un estimateur sans biais. On peut donc dire que l'ensemble des résultats d'un sondage est représentatif si chaque résultat est représentatif. A ce stade, nous pourrions étudier la proposition de définition suivante :

Définition 0 : *Un échantillon est représentatif pour une étude (c'est à dire un ensemble de résultats) si et seulement si l'ensemble des résultats de l'étude est représentatif.*

Ici, un échantillon est représentatif si et seulement si tout résultat produit est un estimateur sans biais de la quantité recherchée dans la population. Ici, l'échantillon est représentatif au regard de la qualité de l'ensemble des résultats produits par l'étude.

Cependant, cette définition conditionne la représentativité de l'échantillon aux seuls résultats produits. En un sens, c'est suffisant car, « peut importe pour les résultats non produits ». Mais

ici, un échantillon ne peut être représentatif « en soi ». Or, nous avons admis, par exemple, qu'un échantillon simple au hasard doit être un échantillon représentatif, un échantillon représentatif « en soi » ou « par définition ».

Cette définition est néanmoins intéressante. En effet, si l'on démontre que tout résultat est issu d'un estimateur sans biais, alors l'échantillon est représentatif. On aurait donc une méthode de démonstration de la représentativité de l'échantillon au regard des traitements qui seront réalisés.

Malheureusement, si nous souhaitons estimer une quantité pour laquelle il n'existe pas d'estimateur sans biais quelque soit l'échantillon (par exemple, le maximum d'une variable quantitative n'est estimé sans biais que si $n=N$), nous devrions conclure qu'aucun échantillon représentatif n'existe pour cette quantité. Or nous voulons qu'un échantillon simple au hasard soit un échantillon représentatif. La définition 0 est donc sans suite.

3. Définition statistique d'un échantillon représentatif d'une population de taille finie.

La statistique est basée sur la notion d'information et on posera que la population contient toute l'information (sur cette population). La définition d'un échantillon représentatif ne peut pas dépendre des « maladresses » de méthodes d'estimation. On veut qu'un échantillon puisse être représentatif en soi.

Dans la suite, on reprend les notations de l'introduction en notant $S = \{i_1, \dots, i_n\}$ les indices des unités de l'échantillon et $E = \{u_i, i \in S\}$ l'échantillon lui même.

Définition 1 : *Une caractéristique d'une population de taille N est un vecteur de taille N qui consigne, pour cette population les valeurs prises par chaque unité de la population à un moment donné (ex : âge de chaque personne).*

Il est clair qu'une caractéristique C d'une population peut s'exprimer en termes d'une distribution empirique $F_N(C)$ simplement définie comme l'ensemble des fréquences dans la population des valeurs prises par la caractéristique dans la population.

Définition 2 : *L'ensemble des caractéristiques d'une population de taille N est une matrice de taille $N \times K$ qui consigne, pour cette population les valeurs prises par chaque individu pour*

l'ensemble des K caractéristiques de la population (ex : âge, taille, CSP, ...).

Il est clair que l'ensemble des caractéristiques d'une population peut s'exprimer en termes d'une distribution empirique de dimension K : $F_N(C_1, \dots, C_K)$. On note $C_{(i,k)}$ la valeur de la caractéristique C_k de l'individu u_i de la population P .

Définition 3 : Echantillon représentatif d'une caractéristique

*Un échantillon E composé de n unités $\{u_i\}_{i \in S}$ est **représentatif de la caractéristique C_k** d'une population de taille N s'il existe une méthode d'échantillonnage probabiliste dans E d'une unité $u_i \in E$ telle que la loi de probabilité de $C_{(i,k)}$ qui est la valeur de cette caractéristique pour l'individu $u_i \in E$, pris au hasard dans l'échantillon, est égale à la loi de distribution empirique $F_N(C_k)$ de cette caractéristique dans la population P .*

Remarque : Cette définition ne présuppose pas que la méthode d'échantillonnage initial de E dans P soit probabiliste. Elle impose simplement la possibilité de tirer au hasard dans E selon une loi dont le transport sur la caractéristique correspond à la loi empirique de la caractéristique dans P . Comme pour toute méthode probabiliste, en notant E_1 l'échantillon (de taille un) issu de cette méthode de tirage, on peut définir les probabilités d'inclusion conditionnelles :

$$P_i = P(u_i \in E_1 \mid u_i \in E) \text{ pour } i \in S.$$

NB : ces probabilités sont différentes des probabilités d'inclusion conditionnelles suivantes :

$$Q_i = P(u_i \in E_1 \mid u_{i_1} \in E, u_{i_2} \in E, \dots, u_{i_n} \in E), i \in S$$

qui sont les probabilités d'inclusion dans E_1 connaissant l'ensemble de l'échantillon E , et correspondent donc à la loi de tirage de E_1 dans E avec $\sum_{i \in S} Q_i = 1$.

Définition 4 : Echantillon représentatif d'une population de taille finie.

*Un échantillon E composé de n unités $\{u_i\}_{i \in S}$ est **représentatif d'une population P** s'il existe une méthode d'échantillonnage probabiliste dans E d'une unité $u_i \in E$ telle que la loi de*

probabilité conjointe des caractéristiques $(C_{(i,1)}, \dots, C_{(i,K)})$ de l'individu $u_i \in E$ pris au hasard dans l'échantillon est égale à la loi de distribution empirique de l'ensemble des caractéristiques dans la population P , c'est à dire que $F_{E_i}(C_1, \dots, C_K) = F_N(C_1, \dots, C_K)$.

En substance, un échantillon est donc représentatif s'il est possible de tirer au hasard dans cet échantillon un individu dont la loi des caractéristiques est celle de la population. Le hasard de ce tirage n'est pas nécessairement équiprobable. On démontre qu'un échantillon est représentatif si et seulement si on démontre que $F_{E_i}(C_1, \dots, C_K) = F_N(C_1, \dots, C_K)$.

Propriété 1 : *La population P est un échantillon représentatif de la population P .*

Démonstration : Si on tire au hasard équiprobable (i.e. $\{Q_i = P_i = 1/N\}_{i=1..N}$) une unité u_i de la population, alors il est clair que $F_{E_i}(C_1, \dots, C_K) = F_N(C_1, \dots, C_K)$.

Propriété 2 : *Un échantillonnage aléatoire simple produit un échantillon représentatif de la population P .*

Démonstration : Si on tire au hasard équiprobable, n individus dans la population et que l'on tire un individu au hasard équiprobable dans cet échantillon, il est clair que cela revient à tirer un individu au hasard équiprobable directement dans la population, revenant ainsi au cas de la propriété 1 et donc $F_{E_i}(C_1, \dots, C_K) = F_N(C_1, \dots, C_K)$. Dans ce cas, $Q_i = P_i = 1/N$, pour tout i dans S .

Propriété 3 : *Si E est un échantillon de n individus d'une population P de taille N issu d'une méthode d'échantillonnage probabiliste tel qu'il existe une méthode d'échantillonnage probabiliste d'une unité $u_i \in E$ avec $P(u_i \in E_i) = 1/N$, pour tout $i = 1..N$, alors E est un échantillon représentatif de P :*

$$\text{Si } P(u_i \in E_i) = 1/N \text{ pour tout } i = 1..N \text{ alors } F_{E_i}(C_1, \dots, C_K) = F_N(C_1, \dots, C_K),$$

Démonstration : Cette propriété est une conséquence directe de la propriété 2.

Ainsi, en substance, un échantillon est représentatif si sa construction est « équivalente » à celle d'un échantillon simple au hasard.

Propriété 4 : Si E est un échantillon de n individus d'une population P de taille N est issu d'une méthode d'échantillonnage probabiliste avec des probabilités d'inclusion connues et supérieures ou égales à $1/N$, (i.e. $P(u_i) = P(u_i \in E) \geq 1/N$, pour $i = 1 \dots N$), alors E est un échantillon représentatif de P :

$$\text{Si } P(u_i) \geq 1/N \text{ pour tout } i=1 \dots N \text{ alors } F_{E_1}(C_1, \dots, C_K) = F_N(C_1, \dots, C_K)$$

Démonstration : si $P(u_i) \geq 1/N$, on pose $P_i = 1/(N \times P(u_i))$ pour $i = 1 \dots N$. On obtient alors un ensemble de probabilités qui peut être la base d'une méthode d'échantillonnage conditionnel d'un individu dans l'échantillon E . Alors, la probabilité d'inclusion de tout u_i dans E_1 est : $P(u_i \in E_1) = P(u_i \in E_1 | u_i \in E) \times P(u_i \in E) = P_i \times P(u_i) = 1/N$ pour tout $i=1 \dots N$. La propriété 3 nous dit que $F_{E_1}(C_1, \dots, C_K) = F_N(C_1, \dots, C_K)$.

Remarque : La condition portant sur des probabilités d'inclusion connues et supérieures ou égales à $1/N$ peut sembler surprenante. Pour autant cette condition est en réalité naturelle et essentielle. En effet, supposons que l'on tire un échantillon de taille deux sur une population de taille 10 avec une méthode d'échantillonnage telle que les probabilités d'inclusion soient $P(u_1) = \dots = P(u_9) = 0,22$ et $P(u_{10}) = 0,02$. Dans ce cas, $P(u_{10}) < 1/10$, l'échantillon n'est pas représentatif de la population P au sens de la définition 4. Cet exemple montre qu'un trop gros déséquilibre dans les probabilités d'inclusion peut conduire à un échantillon non représentatif.

Cette limite n'est pas si surprenante : si l'on s'intéresse à l'estimation d'une proportion d'une caractéristique D par l'estimateur de Horvitz-Thomson, on peut trouver la formule par exemple dans Ardilly (2006 page 133) :

$$\ll \hat{P}_D = \frac{1}{N} \sum_{\substack{i \in S \\ i \in D}} \frac{1}{P_i} \gg \text{ qui est, avec nos notations, égale à } \frac{1}{N} \sum_{\substack{i \in E \\ i \in D}} \frac{1}{P(u_i)} .$$

Or, si $\exists P(u_i) < 1/N$, l'estimation de la proportion est supérieure à 1. Dans notre exemple, si l'individu n°10 est le seul à posséder le caractère D et si l'individu n°10 appartient à l'échantillon E , on estimera la proportion de D dans la population à : $\frac{1}{10} \times \frac{1}{0,02} = 5$, soit

500%.

NB : La propriété d'estimation sans biais n'est clairement pas suffisante pour garantir la représentativité de l'échantillon. En effet, dans l'exemple précédant, l'estimation de la proportion de D dans la population est sans biais puisque :

- u_{10} appartient à E avec probabilité 0,02 et l'estimation vaut alors 500%,
- u_{10} n'appartient pas à E avec probabilité 0,92 et l'estimation vaut alors 0%.

Donc l'espérance est bien de $0,02 \cdot 500\% = 10\%$ (comme dans la population). Mais il est clair qu'aucun statisticien ne cautionnerait une estimation de 500%. Donc on devra limiter l'estimation à 100%. Mais dans ce cas, l'espérance de l'estimateur sera de $0,02 \cdot 100\% = 2\%$. L'estimateur est alors biaisé.

4. Représentativité et méthode des quotas.

La méthode des quotas est la plus fréquemment utilisée pour tenter de construire un échantillon « représentatif ». Pour Jérôme Fourquet en 2011 « *C'est en respectant cette méthode des quotas qu'on arrive à avoir un échantillon représentatif.* ». Pour L'IPSOS (Page Web) « *L'inconvénient majeur de la méthode des quotas est de ne pas permettre de calculer scientifiquement la marge d'erreur du sondage. Les lois statistiques qui permettent de la déterminer ne valent théoriquement que pour les sondages aléatoires. En pratique, on considère cependant que la marge d'erreur des sondages par quotas est égale ou inférieure à celle des sondages aléatoires.* ».

Le principe des quotas est simple : connaissant dans la population la distribution de certaines caractéristiques de la population, on sélectionnera des individus de la population séquentiellement jusqu'à satisfaction (en proportions exactes) de ces distributions. Cette seule contrainte ne définit pas les probabilités d'inclusion de chaque individu de la population. Pour autant, cette méthode ne signifie pas que les probabilités d'inclusion soient nécessairement inconnues.

Exemple : supposons que l'on dispose d'une base de sondage exhaustive et que l'on sélectionne un échantillon de n individus séquentiellement selon la méthode des quotas de la manière suivante : le premier individu de l'échantillon est sélectionné au hasard simple ($1/N$) dans la population. Le second au hasard simple ($1/(N-1)$) et ainsi de suite. Si, à un moment, l'individu ne satisfait pas aux quotas, on le retire de la base et on poursuit au hasard simple. Une petite analyse de l'échantillonnage montre que chaque individu de la population a la

même probabilité d'inclusion dans l'échantillon. On est donc ici dans le cas d'un échantillonnage stratifié, représentatif selon la définition 4 et dont « *la marge d'erreur [...] est égale ou inférieure à celle des sondages aléatoires* » (IPSOS).

La méthode des Quotas n'est donc pas « en soi » une méthode empirique ou non probabiliste. Elle ne le devient que si les probabilités d'inclusion sont inconnues.

5. Représentativité et méthodes de redressement a posteriori.

Un redressement a pour but de retrouver dans l'échantillon redressé certaines distributions empiriques existant dans la population. Au final, nous obtenons pour les caractéristiques redressées, des distributions égales à celles de la population. Qu'en est-il précisément pour les autres caractéristiques (en particulier celles d'intérêt) ?

Pour répondre à cette question, nous devons d'abord répondre à la question : pourquoi avons-nous redressé a posteriori l'échantillon ? Il est clair que si l'échantillon est représentatif, nous n'avons aucune raison d'effectuer un redressement a posteriori. Donc, notre échantillon n'est, pour une raison ou pour une autre, pas représentatif, nous le savons et nous voudrions le rendre représentatif. La définition 4 nous dit que tout cela n'est que vaine peine. En effet, nous ne pourrions en aucun cas démontrer la représentativité de cet échantillon redressé.

Remarque 1 : Si un échantillon n'est pas représentatif, un redressement même mûrement construit ne le rendra représentatif, au mieux, que pour les caractéristiques de redressement au sens de la définition 3 (pour un calage sur marge par exemple).

Remarque 2 : Un échantillon par quotas est représentatif des distributions marginales des caractéristiques prises en compte pas les quotas. De ce fait, cette méthode est identique à un échantillonnage aléatoire (non représentatif dû à un défaut de couverture ou des non-réponses) avec redressement par calage sur les mêmes caractéristiques.

6. Conclusion

Ce papier propose une définition d'un échantillon représentatif d'une population finie. Cette définition n'est pas basée sur la sélection au hasard d'individus de la population mais la sélection au hasard d'un individu de l'échantillon. De ce fait, il n'y a aucun a priori sur le mode de sélection de l'échantillon. Les propositions qui en découlent sont riches d'enseignements. Dans un premier temps, on démontre qu'un échantillon aléatoire simple est

bien représentatif. De plus, on montre que si la méthode d'échantillonnage a des probabilités d'inclusion trop disparates, la représentativité n'est pas évidente. Enfin, on voit que la représentativité de la méthode des quotas n'a rien à envier à celle des échantillons aléatoires avec redressement pour défaut de couverture ou non-réponse.

Les différentes citations de la partie (2.a.) peuvent être commentées au regard de cette définition de la représentativité d'un échantillon.

Pour Yves Tillé et Pascal Ardilly, la représentativité est une notion « inadaptée ». Nous espérons que la définition proposée dans cet article les fera changer d'avis. Pour Olivier Sautory, « *Un échantillon n'est jamais représentatif "en soi"* ». Notre définition est basée sur le contraire. Pour Jean Vaillant, il suffit que les probabilités d'inclusion soit connues et non nulles. Nous avons montré le problème que pose l'utilisation de probabilités d'inclusion inférieures à $1/N$. Notre définition est en revanche, totalement en accord avec la citation de Sheldon M. Ross et pour l'essentiel, en accord avec Philippe Dutarte.

Il est clair pour tous que la représentativité de l'échantillon est un idéal. En effet, un taux de couverture de 100%, un taux de non-réponse de 0%,... font partie des hypothèses du théoricien mais rarement de la réalité du statisticien. On peut alors se poser la question de la construction d'un indice de représentativité qui aurait pour but de relativiser le plus objectivement possible (et de la manière la plus consensuelle) le défaut de représentativité d'un échantillon.

Bibliographie

- [1] Ardilly P. (2006), *Les techniques de sondage*, Edition TECHNIP.
- [2] Cochran W.G. (1977), *Sampling techniques*, 3rd edition, Wiley & Sons, NY.
- [3] Dutarte P. (2005), *L'induction statistique au lycée* (ed : Didier).
- [4] Fourquet J. (2011), *Emission C dans l'air du 17 Février 2011*.
- [5] Norme ISO 3534-1 (1996) *Vocabulaire et symboles*.
- [6] Ross S.M. (1999), *Introduction to Probability and Statistics*, Elsevier
- [7] Sautory O. (2010), *Journée d'études sur la représentativité*, ENS Paris.
- [8] Senat (2010), *Proposition de loi du 14 février 2010 relative à la publication et à la diffusion de certains sondages d'opinion*.
- [9] Tillé Y. (2001), *Théorie des sondages*, Edition DUNOD.
- [10] Vaillant J. (2005), *Initiation à la théorie de l'échantillonnage*, Web.