

La représentativité d'un échantillon et son test par le Khi-deux

Testing the representativeness of a sample

Louis Laurencelle

Université du Québec à Trois-Rivières

The value and credibility of a survey or opinion pool depend to a large extent on the observed sample's representativeness, which in turn depends on sample size. When dealing with an heterogeneous population, however, sample size may be insufficient to counterbalance possible biases due to differences in constituent subpopulations. If population's composition is known, the sample's representativeness can be verified by a Chi-square test: the test is illustrated and a special formula is proposed and validated for the case of finite populations.

La valeur d'une statistique issue d'un sondage, telle qu'une moyenne ou une proportion, dépend à la fois de la représentativité de l'échantillon d'éléments sur lequel elle est basée et de sa méthode de calcul. Dans les cas où l'enquête porte sur une population hétérogène et de composition connue, la représentativité de l'échantillon lui-même peut être vérifiée par un test statistique, le test Khi-deux d'ajustement. Nous examinons la réalisation de ce test pour les cas de population infinie ou finie et nous proposons une approximation originale pour ce dernier cas.

Article first published in Lettres Statistiques, 2005, vol. 12, p. 131-146.

Peut-on légitimement appliquer à la population entière les résultats et conclusions obtenus à partir d'un seul groupe d'éléments, d'un échantillon? C'est une question fondamentale que doit aborder la méthodologie des enquêtes et sondages, la réponse à cette question formant la trame des théories de l'échantillonnage. Au cœur de ces théories, trois concepts : la représentativité, le biais et la précision.

Imaginons une population quelconque, constituée d'un ensemble de N éléments pouvant chacun être mesuré et noté X_i , population dont la moyenne, ou espérance, est définie par $\mu_X = \sum_{i=1}^N X_i / N$, la taille N pouvant être infinie : cette quantité peut dénoter une grandeur (p. ex. le poids corporel, le revenu annuel, etc.), une réponse (p. ex. 1 = « Oui », 0 = « Non »), un attribut (p. ex. sexe, catégorie d'emploi, intention de vote).¹ Tirons un échantillon de n éléments de

cette population, soit $\{ X_1, X_2, \dots, X_n \}$, où $n < N$, et calculons $\bar{X} = \sum_{i=1}^n X_i / n$ à partir de ce sous-ensemble. Alors, le biais d'échantillonnage B est défini par :

$$B = E[\bar{X}] - \mu_X,$$

et la précision σ^2 , par :

$$\sigma^2 = E[\bar{X}^2] - (E[\bar{X}])^2.$$

La précision, exprimée ici par la variance d'erreur de l'estimateur \bar{X} , dépend bien sûr du plan d'échantillonnage, mais son principal facteur est la taille d'échantillon n : nous n'y reviendrons pas. Le biais, quant à lui, dépend

« Homme », 0 = « Femme ») ou multiple (p. ex. la catégorie d'emploi : « Sans emploi », « Col blanc », « Col bleu », « Entrepreneur indépendant », etc.), tout comme les quantités X_i et μ_X peuvent être scalaires ou vectorielles. Notre propos vaut également pour les deux types de quantités. Cependant, pour rester simple, nous nous en tiendrons aux scalaires dans la suite du texte.

¹ La donnée d'attribut peut être simple (p. ex. le sexe : 1 =

essentiellement du plan d'échantillonnage, c.-à-d. de la méthode employée pour mettre sur pied l'échantillon. Pour que le biais soit nul, c.-à-d. pour que la quantité \bar{X} obtenue par l'échantillon soit légitimement applicable à la population, la méthode de mise sur pied de l'échantillon doit en garantir la représentativité.

Le concept de représentativité n'est pas explicitement mentionné dans plusieurs grands traités de statistique sur l'échantillonnage (Cochran 1963 ; Kish 1965 ; Rossi, Wright et Anderson 1983 ; Thompson 1997) : on se contente plutôt de parler d'un échantillon ou d'un échantillonnage sans biais, confondant ainsi la méthode de sélection des éléments et leur valeur mesurée. Certains auteurs, comme Yates (1981) et S. K. Thompson (1992), identifient "échantillon représentatif" à "échantillon de représentants", les éléments étant choisis et spécifiquement désignés, voire mandatés, pour "représenter" la population. Le concept français d'échantillon représentatif n'a pas cette acception restrictive. Logiquement affilié à l'échantillon au hasard simple (Robert 1988), pour lequel chaque élément de la population a une chance égale d'être tiré, l'échantillon représentatif en est un dans lequel les différentes sous-populations ou strates de la population-mère ont des chances proportionnelles de se retrouver.

La population serait-elle homogène, c.-à-d. formée d'éléments tous interchangeables, à de menues variations près? Dans ce cas, - une utopie dans le monde humain mais un cas possible quand on s'intéresse à des populations du règne végétal ou du règne animal ou à des produits manufacturés, par exemple -, n'importe quel lot d'éléments peut être dit représentatif (puisque tous les lots sont équivalents) et sans biais. Pour une population hétérogène, d'autre part, le problème se pose avec acuité, et les concepts de biais et de représentativité y méritent alors une discussion nuancée.

Représentativité versus biais

Si population hétérogène il y a, c'est que la mesure X_i change systématiquement en fonction de certaines caractéristiques de ses éléments. Si la possibilité nous était donnée de mesurer d'avance toute la population ou une bonne fraction de celle-ci, nous pourrions établir quelles caractéristiques sont pertinentes en ce qu'elles influencent la valeur de la mesure X , ainsi que dresser un plan d'échantillonnage afin de les contrôler. Ce plan d'échantillonnage découperait la population selon différents niveaux des caractéristiques identifiées et permettrait de reconstituer à petite échelle, de taille n , la composition globale de la population, de taille N . Un tel mode d'échantillonnage serait, selon nous, à la fois *représentatif* et *sans biais*.

Hélas, de par la nature des choses, nous ne pouvons que

rarement mesurer toute une population, et nous ne sommes jamais à même d'établir avec confiance les caractéristiques démographiques pertinentes pour notre objet d'étude. Par contre, nous constatons assez facilement si et jusqu'à quel point notre mesure X varie d'un élément à l'autre, c.-à-d. si la population visée peut être réputée homogène ou hétérogène. Si hétérogène, l'échantillonnage au hasard simple, la méthode idéale en principe, devient en effet seulement *idéale* en ce qu'elle se heurte à de grandes difficultés d'opérationnalisation. Il reste donc l'échantillonnage stratifié, une méthode qui consiste à découper la population en sous-populations artificielles, ou strates, selon des critères ou caractéristiques externes dictés par un plan d'échantillonnage : on pense par exemple au sexe, au niveau socio-économique, au statut civique, etc.

Les critères, ou variables de stratification, appliqués pour un échantillonnage stratifié ne sont pas obtenus d'ordinaire par une étude faite sur l'objet d'enquête mais, plutôt, le chercheur suppose et argumente qu'il existe un lien de corrélation, quel qu'il soit, entre la mesure X et ces dits critères. Pour autant qu'un tel lien existe, l'échantillonnage stratifié va généralement produire des estimateurs plus précis que l'échantillonnage au hasard simple, à tailles d'échantillon égales (Cochran 1963 ; Kish, 1965 ; Laurencelle, 2001 ; Thompson 1997 ; Voyer, Valois et Rémillard 2000). De plus, en structurant la population-mère en îlots ou sous-ensembles bien définis d'éléments, ce type d'échantillonnage réduit les risques de dérapage électif qu'encourt l'échantillonnage censé au hasard simple et il contribue ainsi à mieux contrôler le biais.

L'échantillon stratifié proportionnel, qui reproduit dans ses éléments la composition connue de la population en fonction des variables de stratification convenues, peut être dit représentatif de la population. Le procédé de stratification, même si les critères utilisés sont sans relation aucune avec l'objet d'étude (la mesure X), ne produit pas de biais ; au contraire, en régissant la sélection des éléments et en divisant, pour ainsi dire, le risque global en un agglomérat de risques plus cernés, l'amplitude d'un biais possible s'en trouve amoindrie. Enfin, nonobstant la question centrale du biais et celle du lien entre les variables de stratification et l'objet d'étude, il est souvent important que la population, la population humaine, soit et se sente représentée dans une enquête qui la concerne, et que cette représentativité lui soit justifiable en termes de critères et de correspondances proportionnelles. La représentativité d'échantillon déborde ainsi et transcende la question du biais échantillonnel, de même que l'argument de représentativité fait son propre pont avec les concepts de légitimité et d'imputabilité d'une enquête.

Le test de représentativité par le Khi-deux

Que ce soit pour démontrer que l'échantillon constitue un reflet fidèle de la population dans sa diversité, en rassurant par exemple certaines sous-populations sensibles quant à leur présence proportionnelle dans les résultats d'enquête, ou pour contenir les risques d'un biais d'estimation, il peut être intéressant de vérifier la représentativité de l'échantillon *en démontrant que celui-ci ne s'écarte pas significativement de la répartition proportionnelle des sous-groupes trouvés dans la population*. Pour ce faire, le test d'ajustement du Khi-deux tombe à point nommé.

Notation et calcul du test d'ajustement du Khi-deux. Nous avons une population de taille N (N pouvant être infiniment grand), constituée de k sous-populations ou strates, de tailles respectives N_1, N_2, \dots, N_k . De cette population, par échantillonnage stratifié ou autrement, nous tirons un échantillon de n éléments, qui se répartissent en effectifs n_1, n_2, \dots, n_k parmi les k catégories. Sous une répartition proportionnelle exacte de l'échantillon, chaque case échantillonnale j compterait $n \times N_j/N$ éléments. Posons la fraction échantillonnale

$$f = n/N;$$

alors le test Khi-deux approprié à cette situation est :

$$\chi_{\text{classique}}^2 = \frac{(n_1 - f \times N_1)^2}{f \times N_1} + \frac{(n_2 - f \times N_2)^2}{f \times N_2} + \dots + \frac{(n_k - f \times N_k)^2}{f \times N_k} \quad (1)$$

la distribution de référence étant celle du χ^2 avec $k - 1$ degrés de liberté. Ainsi, l'obtention d'une valeur calculée telle que $\chi^2 \geq \chi_{k-1[1-\alpha]}^2$ c.-à-d. significative au seuil α , indiquerait que, dans ses effectifs, l'échantillon considéré trahit en quelque sorte la population dont on prétend l'avoir tiré.

La question lancinante de la valeur minimale des *fréquences théoriques* (représentées ici par les produits $f \times N_j$) ne se pose guère ici, comme elle se pose rarement par ailleurs. Kendall et Stuart (1979) rapportent des conclusions à l'effet que l'approximation χ^2 invoquée dans le test d'ajustement convient dès que la *moyenne* des fréquences théoriques atteint la valeur « 2 », en d'autres mots dès que l'échantillon contient au moins $2k$ éléments. Nos propres investigations, consistant à comparer la probabilité de la distribution χ^2 à celle issue de 10.000 échantillons Monte Carlo pour des tailles d'échantillon de plus en plus petites, vérifient ces conclusions.

La procédure statistique proposée, et l'approximation χ^2 qui la complète, supposent que le tableau d'effectifs $\{n_1, n_2, \dots, n_k\}$ correspond à un échantillon multinomial, donc à un échantillon tiré d'une population à sous-populations inépuisables ou tiré avec remise d'une population finie. Or, même si dans plusieurs cas le ratio n/N ($=f$) est quasi nul, suggérant une population pratiquement infinie, on

rencontre des cas où ce ratio empiète sérieusement sur l'unité. Pensons aux enquêtes de contrôle sur la qualité de service, où l'on interrogera jusqu'à une personne sur cinq ($f = 0,20$), ou aux relances sur les effets d'une nouvelle thérapie, qui rejoindront parfois une personne sur deux ($f = 0,50$).

Élaboration d'un test approprié aux populations finies. Le test d'ajustement du Khi-deux, tel qu'il est et que nous l'avons présenté ci-dessus, ne convient pas aux échantillons tirés d'une population finie et correspondant à une fraction d'échantillonnage substantielle. Pour combler cette lacune, nous avons exploré des variantes du test (1), basées sur la correction de variance pour population finie (Cochran 1963): la variance d'une moyenne basée sur n éléments tirés (sans remise) d'une population de taille N est donnée par :

$$\frac{\sigma^2}{n} \times \frac{N-n}{N-1}$$

et est approximativement égale à $\sigma^2 / n[1 - f]$. La première variante proposée incorpore cette correction de variance sur chaque composante du Khi-deux, soit: ²

$$\chi_{\text{corr.ind}}^2 = \sum_{j=1}^k \frac{(n_j - f \times N_j)^2}{f \times N_j} \times \frac{N_j - 1}{N_j - n_j} \quad (2)$$

la seconde variante, plus simple, impose une correction globale, proportionnelle à $1 - f$, soit :

$$\chi_{\text{corr.glo}}^2 = \sum_{j=1}^k \frac{(n_j - f \times N_j)^2}{f \times N_j} \times \frac{N-1}{N-n}. \quad (3)$$

Nous avons étudié sommairement la performance de ces deux variantes ainsi que celle du test classique (1) par simulation Monte Carlo. Les paramètres de l'étude ont été :

- la taille N des populations, soit 200, 1000, 5000 et 10000 ;
- le nombre k de sous-populations ou strates, soit 2, 4, 6, 8 et 10 ;
- la répartition des éléments parmi les strates, soit égalitaire ($N_j = N/k$), soit "aléatoire" ($N_i \sim N \times u_i$, où u_i est une variable aléatoire uniforme entre 0 et 1) ;
- la fraction échantillonnale, soit 1 % (ou $f = 0,01$), 5, 10, 20 et 40 %.

Dans chaque cas, nous générions à 5 reprises une population et, pour chaque population, nous prenions 100 échantillons aléatoires différents (pigés sans remise dans la population), pour lesquels les variantes (1), (2) et (3) étaient calculées. De plus, pour chaque échantillon produit, nous estimions la probabilité exacte du Khi-deux classique (1) en

² Rappelons au lecteur que le Khi-deux est essentiellement une somme d'écart-réduits carrés (z^2), chacun étant constitué sous le modèle $z = (X - \mu) / \sigma$. La variable X désignée, n_j , a pour espérance $\mu = f \times N_j$ et, étant considérée comme une variable de Poisson, son écart-type est $\sqrt{\mu}$. Ainsi, le dénominateur de la composante, $f \times N_j$, dénote la variance de n_j , et c'est à elle que la pondération corrective s'applique.

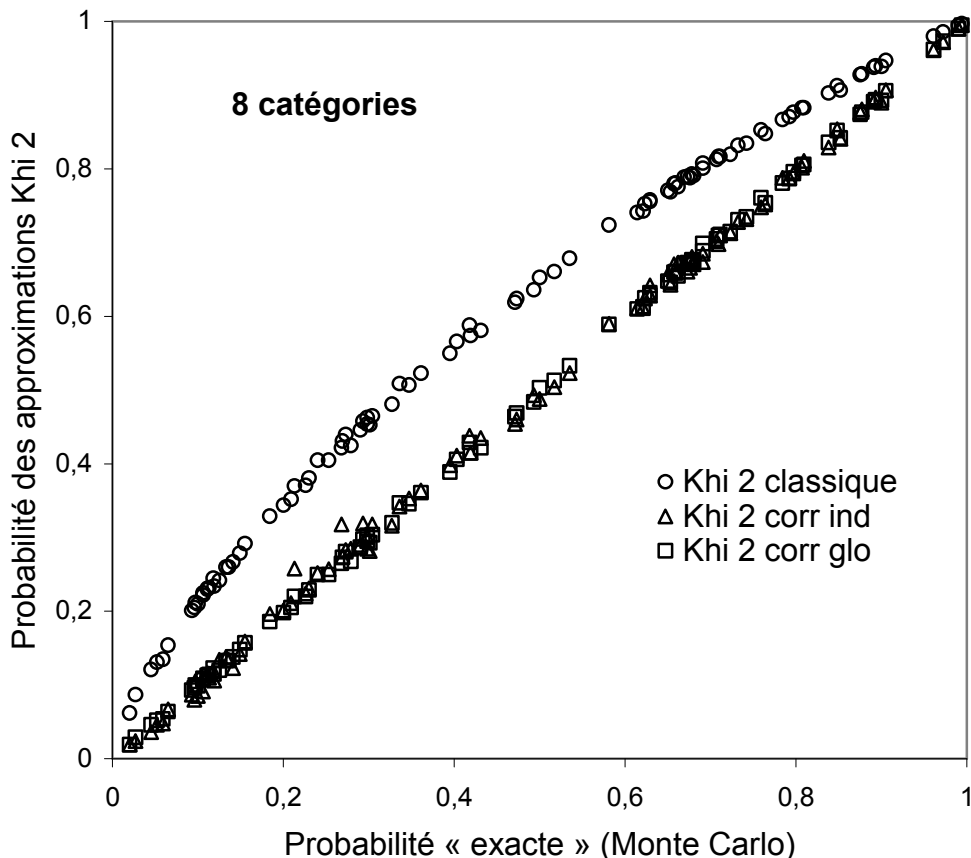


Figure 1. Comparaison des probabilités extrêmes correspondant aux formules (1), (2) et (3) du test Khi-deux pour 100 échantillons de $n = 1000$ éléments (fraction = 20 %) tirés sans remise d'une population de $N = 5000$ éléments répartis en 8 catégories ($N_j = 1090, 510, 285, 1064, 408, 678, 852$ et 113)

comparant sa valeur obtenue à celles de 10.000 échantillons Monte Carlo équivalents. C'est cette probabilité "exacte" (précise à au moins 0,01 grâce aux 10000 estimations) qui permet de juger la valeur des variantes comparées. À toutes fins utiles et parce qu'ils confirment les tendances générales en les illustrant au mieux, nous présentons seulement les résultats pour des populations de $N = 5000$ éléments, comportant $k = 2, 4, 6$ ou 8 strates à répartition aléatoire (et inégale), et des fractions d'échantillonnage de 5, 10, 20 et 40 %, correspondant à des tailles d'échantillons respectives de $n = 250, 500, 1000$ et 2000 éléments. Nous confirmons au lecteur que les données sur les strates ou catégories de population égales sont en tous points comparables à celles issues des catégories inégales. La figure 1 illustre l'allure des résultats d'une réalisation dans notre plan d'étude, soit l'un des cinq exemplaires obtenus pour $k = 8$ catégories de population et une fraction échantillonnale de 20 %, ou $n = 1000$.

La probabilité "exacte", P_{ref} , utilisée comme référence d'abscisse à la figure 1, est fournie, rappelons-le, par une

estimation Monte Carlo, précise à $\pm 0,01$: la diagonale principale constitue ainsi le critère de comparaison pour les trois quantités à juger, soit P_1 (Khi-deux classique), P_2 (Khi-deux corr ind) et P_3 (Khi-deux corr glo), selon les formules (1), (2) et (3) respectivement. La figure permet de constater que le test classique, selon P_1 , s'écarte systématiquement de la probabilité de référence, avec un biais positif qui culmine à $P_{ref} = 0,5$: la probabilité pour le rejet de l'hypothèse de conformité échantillonnale est trop forte. Quant à P_2 et P_3 , ces deux approximations suivent assez bien la diagonale, la formule (2) ayant un comportement tant soit peu plus erratique que la formule (3).

Nous avons comparé systématiquement les diverses formules à l'aide de trois indicateurs globaux, basés sur l'écart entre chaque probabilité approximative (\hat{P}) et la probabilité de référence (P_{ref}), soit :

l'erreur moyenne (EM) :
$$\sum_{i=1}^{100} (\hat{P} - P_{ref,i})/100 ;$$

l'erreur maximale (Emax) :
$$\max_{i=1..100} |\hat{P} - P_{ref,i}| ;$$

l'erreur absolue moyenne (EAM) :
$$\sum_{i=1}^{100} |\hat{P} - P_{ref,i}|/100 ;$$

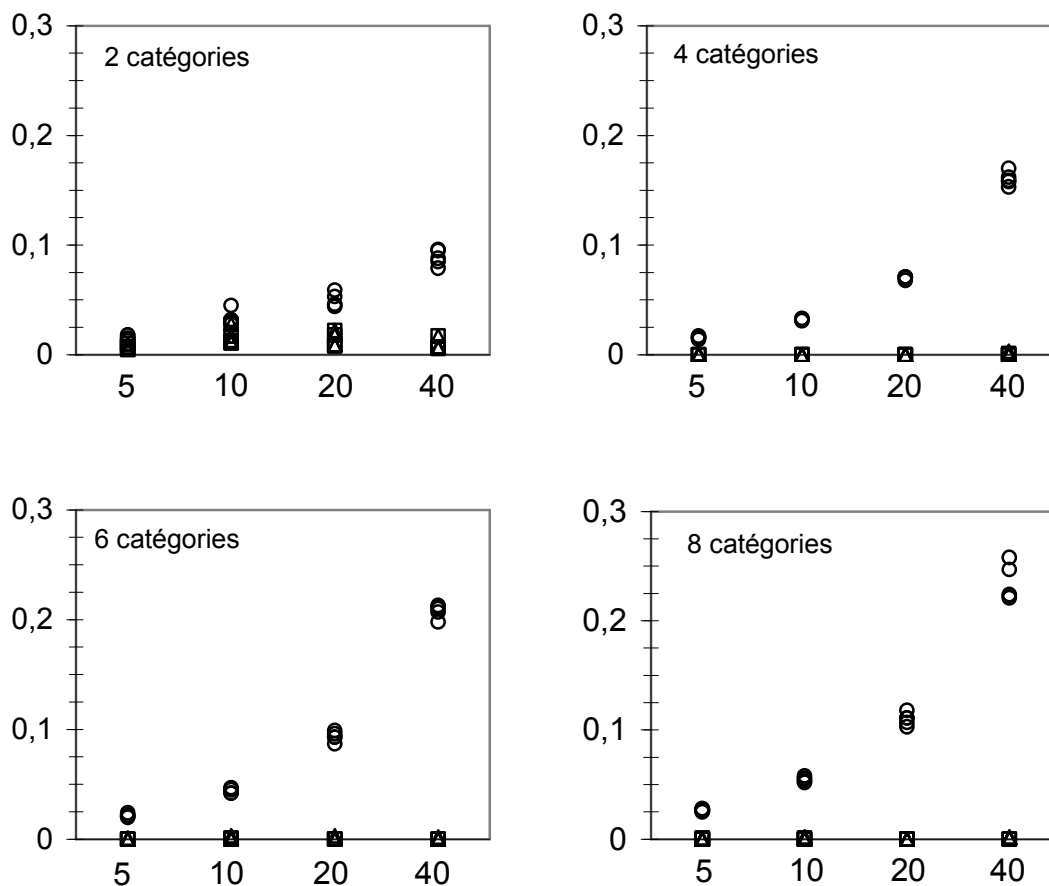


Figure 2. Erreur moyenne (EM) en probabilité pour le Khi-deux classique (O), le Khi-deux à correction individuelle (Δ) et le Khi-deux à correction globale (\square) en fonction de la fraction d'échantillonnage (en %, à l'abscisse), dans des populations réparties inégalement en 2, 4, 6 ou 8 catégories.

Les figures 2, 3 et 4 illustrent les résultats.

La figure 2 montre à l'évidence que l'erreur moyenne du Khi-deux classique augmente avec la fraction d'échantillonnage (f), cet effet étant plus marqué suivant le nombre de strates, ou catégories de population. Cette erreur est positive, la statistique Khi-deux elle-même étant trop faible, sa probabilité trop forte, par rapport à la référence Monte Carlo. Quant aux approximations (2) et (3), elles apparaissent équivalentes, avec une erreur moyenne quasi nulle.

L'erreur maximale est représentée sur des graphes semi-logarithmiques, à la figure 3. Le Khi-deux classique, on le voit, se disqualifie de lui-même. Par ailleurs, la figure fait apparaître une différence entre les approximations (2) et (3) : l'erreur maximale issue du Khi-deux à correction individuelle, ou formule (2), tend à croître avec f à mesure que le nombre de catégories augmente, tandis que, pour le Khi-deux à correction globale, l'erreur maximale estimée tend vers la valeur 0,01, soit à peu près la précision de calcul de notre référence Monte Carlo.

Les graphes logarithmiques de la figure 4 affichent l'erreur absolue moyenne, sans doute l'indice le plus informatif du comportement de nos trois approximations. La configuration des résultats reproduit à peu près celle obtenue pour l'erreur maximale, avec un comportement plutôt bon pour l'approximation (2) mais marqué d'une tendance à l'accroissement selon f , et avec une performance exemplaire de l'approximation (3) qui, très rapidement, se stabilise près de la valeur 0,003, encore une fois une valeur relative à la précision de nos estimations Monte Carlo.

Globalement, tel que prévu, la formule classique du Khi-deux, sans correction pour population finie, ne fonctionne pas dès $f = 0,05$ et elle devient rapidement catastrophique pour f croissant, l'erreur en probabilité augmentant aussi selon le nombre de catégories de population. La statistique Khi-deux obtenue par la formule (1) est trop petite, sa probabilité extrême trop élevée, de sorte qu'elle ne respecte pas le seuil de signification imposé et que, conséquence obligée, sa puissance en pâtit.

Quant aux approximations (2) et (3) avec correction pour

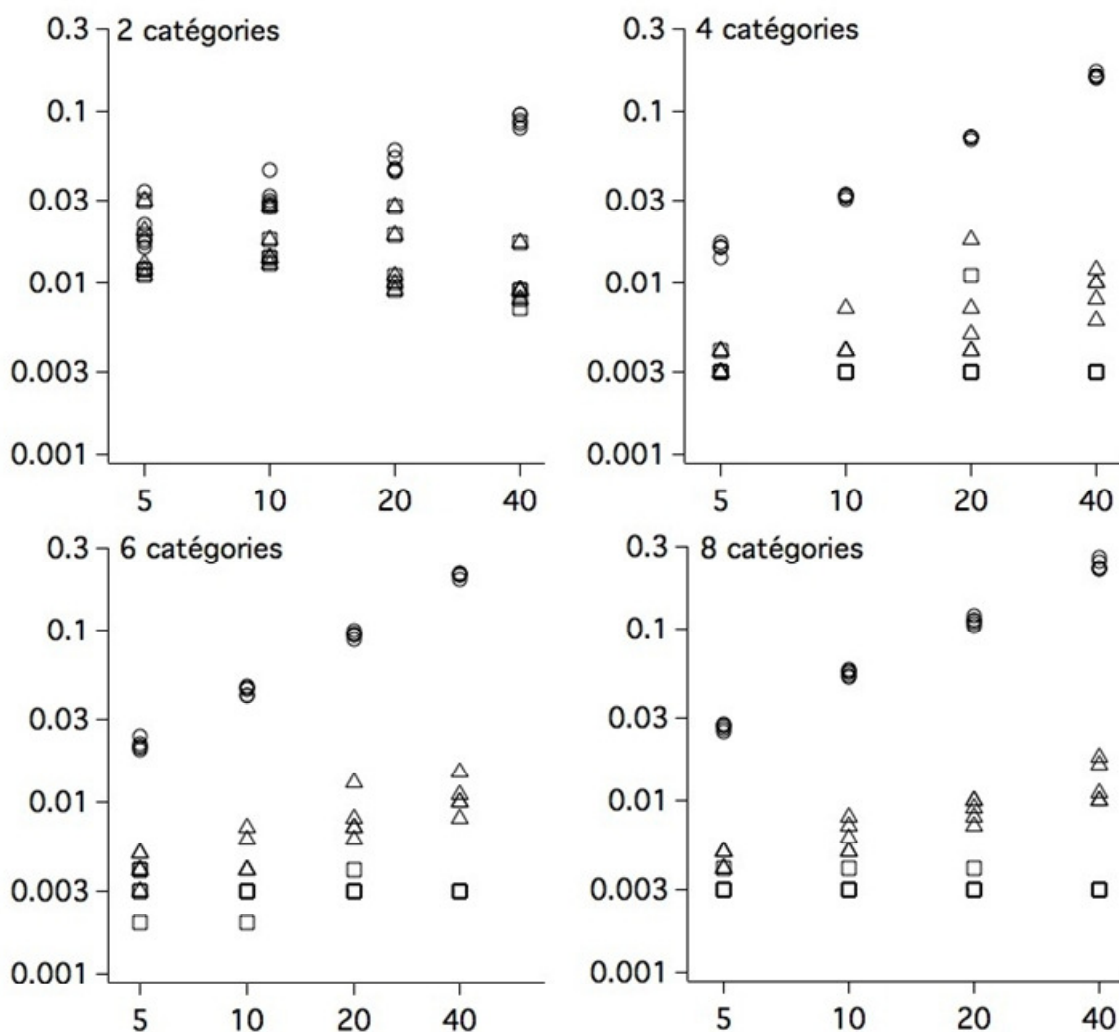


Figure 3. Erreur maximale (E_{max}) en probabilité pour le Khi-deux classique (○), le Khi-deux à correction individuelle (△) et le Khi-deux à correction globale (□) en fonction de la fraction d'échantillonnage (en %, à l'abscisse), dans des populations réparties inégalement en 2, 4, 6 ou 8 catégories

population finie, si elles sont équivalentes et excellentes en termes de probabilité moyenne, la formule (2), avec correction individuelle, se distingue par une instabilité qui non seulement la rend moins fiable mais surtout sensible aux effets de la fraction d'échantillonnage et du nombre de catégories. La formule (3), à correction globale, n'est pas sensible à ces effets et conserve (sauf peut-être pour le cas de

2 catégories) une performance quasi parfaite en termes d'erreur maximale et d'erreur absolue moyenne.

Dans un second temps, nous avons voulu apprécier l'exactitude des trois tests, d'une part en vérifiant s'ils respectent ou non le taux d'erreur attendu et d'autre part en estimant la corrélation de chacun avec la décision "exacte" dérivée des 10.000 échantillons Monte Carlo. Pour ce faire,

Tableau 1. Pourcentages d'échantillons trouvés significatifs (sur 10.000) par le test Khi-deux de représentativité selon une estimation de la probabilité exacte et trois approximations, aux seuils de 5% et 1%

n	5%				1%			
	Exact	(1)	(2)	(3)	Exact	(1)	(2)	(3)
250	4,91	3,89	4,92	4,91	0,99	0,74	1,15	1,01
500	5,04	2,89	5,24	5,00	1,10	0,49	1,24	1,14
1000	5,38	1,52	5,45	5,40	1,15	0,23	1,18	1,14
2000	5,03	0,11	5,24	4,99	0,83	0,00	0,98	0,85

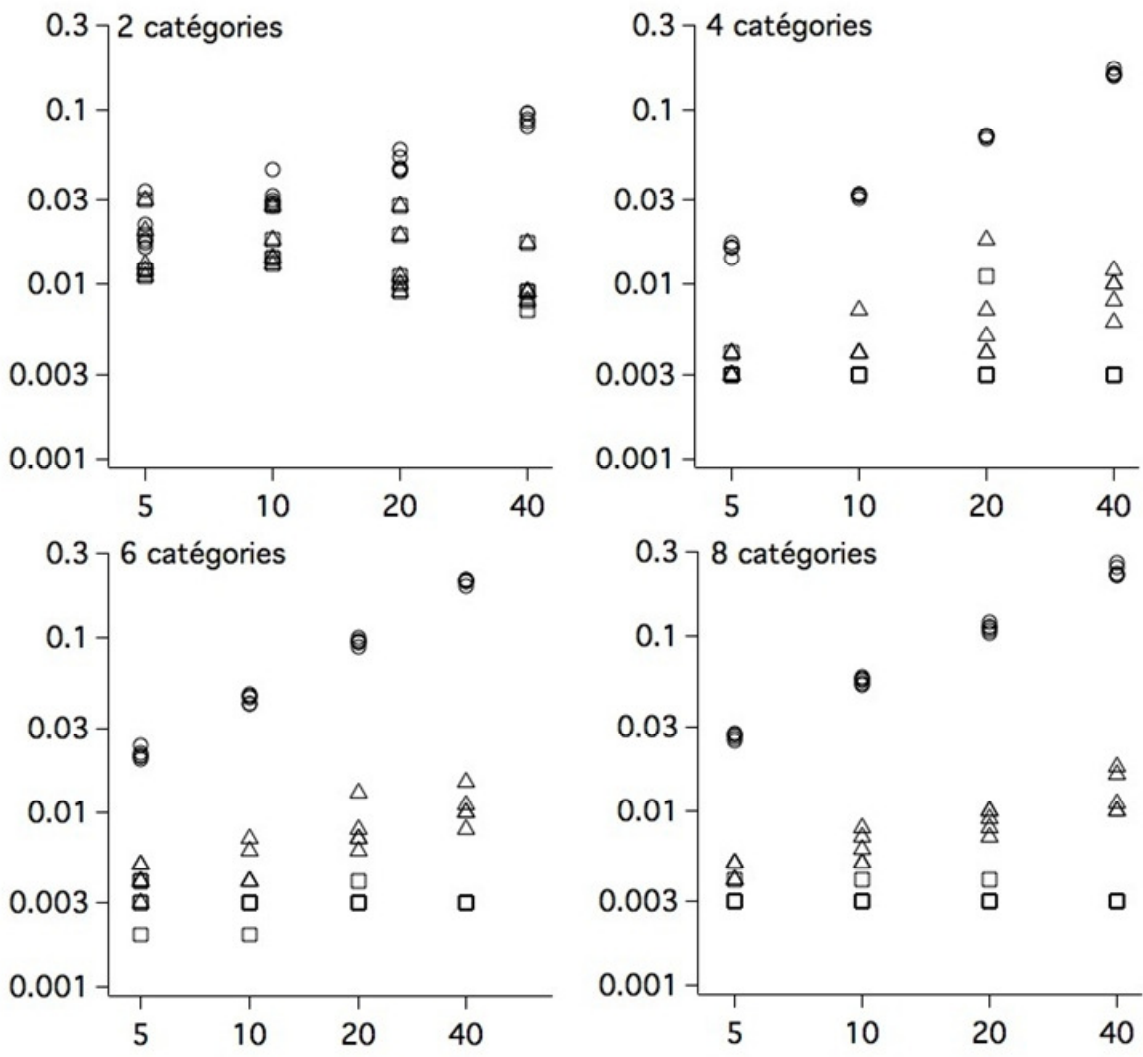


Figure 4. Erreur absolue moyenne (EAM) en probabilité pour le Khi-deux classique (O), le Khi-deux à correction individuelle (Δ) et le Khi-deux à correction globale (□) en fonction de la fraction d'échantillonnage (en %, à l'abscisse), dans des populations réparties inégalement en 2, 4, 6 ou 8 catégories.

nous avons retenu le contexte d'une population de $N = 5000$ éléments catégorisés en $k = 8$ strates inégales (et aléatoires), ce pour des échantillons sans remise de $n = 250$ (5%), 500 (10%), 1000 (20%) et 2000 (40%) éléments : les quantités rapportées sont basées aussi sur 10.000 répliques.

Les niveaux de signification indiqués "Exact" au tableau 1 correspondent au nombre d'échantillons, parmi 10.000 échantillons tirés au hasard, qui ont été trouvés significatifs à 5% ou 1% : les pourcentages rapportés sont très plausibles, s'écartant à peine des valeurs attendues, et ce sont les échantillons correspondants que nous considérons comme étant réellement significatifs.

Le test d'ajustement classique, selon la formule (1), perd de plus en plus d'efficacité à mesure que la fraction échantillonnale ($f = n / 5000$) augmente, jusqu'à devenir pratiquement insensible sous une fraction de 40%. Les deux

approximations pondérées s'en tirent bien mieux, l'approximation (3) gardant une belle proximité avec le test exact, ce quelle que soit la fraction considérée.

Les deux tableaux suivants complètent l'analyse des résultats portant sur l'exactitude des approximations (1), (2) et (3) en termes de significatif / non-significatif. Au tableau 2, le test classique, qui corrèle d'abord assez bien avec le test exact sous une petite fraction échantillonnale ($\varphi = 0,885$ à 5% et 0,872 à 1%), perd rapidement de son mérite quand cette fraction croît, les approximations (2) et (3) étant épargnées de cette dérive. Le tableau 3 affiche quant à lui un taux d'erreurs de décision, soit $(\epsilon^+ + \epsilon^-) / \alpha_R$, où ϵ^+ dénote le taux de faux positifs (échantillons pour lesquels l'approximation rend un verdict significatif alors que le test exact ne le fait pas), ϵ^- le taux de faux négatifs (au contraire du premier), et α_R est le taux de signification réel (tel que retrouvé sous la

Tableau 2. Degré d'association (par corrélation φ) entre chacun des tests Khi-deux d'ajustement selon les approximations (1), (2) et (3) et le test exact (10000 échantillons)

n	$\alpha = 5\%$			$\alpha = 1\%$		
	(1)	(2)	(3)	(1)	(2)	(3)
250	0,885	0,967	0,983	0,872	0,898	0,959
500	0,749	0,951	0,975	0,662	0,898	0,955
1000	0,521	0,948	0,986	0,449	0,944	0,969
2000	0,114	0,927	0,975	–	0,864	0,952

Tableau 3. Taux d'erreurs de décision des tests Khi-deux d'ajustement selon les approximations (1), (2) et (3), soit $(\varepsilon^+ + \varepsilon^-) / \alpha_R$ (voir texte)

n	$\alpha = 5\%$			$\alpha = 1\%$		
	(1)	(2)	(3)	(1)	(2)	(3)
250	0,208	0,063	0,032	0,237	0,227	0,082
500	0,427	0,095	0,048	0,555	0,218	0,091
1000	0,717	0,099	0,026	0,800	0,113	0,061
2000	0,984	0,141	0,048	1,000	0,301	0,096

colonne "Exact" au tableau 1).

En plus de confirmer l'inaptitude de l'approximation (1), les données du tableau 3 font voir à nouveau la prédominance de l'approximation (3) sur les deux autres. Nonobstant la sévérité du critère d'erreur en raison du dénominateur α_R , l'approximation (3) se distingue à la fois par les taux d'erreurs les plus bas et par la constance de ceux-ci à travers les fractions échantillonnelles ($\sim n$) exploitées.

Le test de représentativité Khi-deux avec correction globale, selon la formule (3), se recommande donc de lui-même.

Un exemple, avec calculs

Une équipe, dont faisait partie mon collègue Martin Descarreaux, D.C., Ph. D., a mené une enquête auprès de la population des chiropraticiens exerçant alors au Québec. La population recensée comptait $N = 1113$ professionnels homologués, répartis dans les 14 régions du territoire, lesquelles ont été regroupées en 9 régions. L'échantillon comptait $n = 407$ membres, pris « au hasard » dans chaque région. Le tableau 4 sur la page suivante présente, par numéros de région administrative, les répartitions de la population et de l'échantillon, respectivement.

Ce cas, peut-être un peu exceptionnel, illustre très bien la situation d'un échantillonnage sur une population finie, la fraction d'échantillonnage étant ici de $f = 407 / 1113 \approx 0,3657$.

Le Khi-deux classique, d'après la formule (1), est la somme des composantes $(n_j - f \times N_j)^2 / f \times N_j$. Par exemple, pour les régions 01+11, nous avons $f \times N_j \approx 10,605$ et $(17-10,605)^2 / 10,605 \approx 3,856$. En sommant les 9 composantes, nous obtenons $\chi^2_{\text{classique}} \approx 16,49$; avec $k-1 = 8$ degrés de liberté, la probabilité de rejet est égale à 0,0358 : significative au seuil de 0,05 mais pas au seuil de 0,01. Ce test, on l'a vu, n'est pas crédible dans le cas d'un échantillonnage à fraction f conséquente, comme c'est le cas présent.

Avec la formule (2), chaque composante du Khi-deux est corrigée par son facteur de variance propre. Ici, pour la première composante relative aux régions 01+11, le facteur de variance est $(29-17) / (29-1) \approx 0,4286$, qui sert à diviser la composante brute, d'où la valeur corrigée devient $3,856 / 0,4286 \approx 8,997$. La somme des composantes donne 30,64 ; toujours avec 8 degrés de liberté, la probabilité extrême serait d'environ 0,0002.

La formule (3) consiste simplement à multiplier la valeur du Khi-deux classique par un facteur de population finie global. Ici, ce facteur (inversé) est $(N-1) / (N-n) = (1113-1) / (1113-407) \approx 1,5751$, d'où $\chi^2_{\text{corr glo}} \approx 16,49 \times 1,5751 \approx 25,97$; la probabilité extrême associée est d'à peu près 0,0011.

Bien entendu, nous avons procédé aussi à l'estimation de la probabilité exacte par échantillonnage Monte Carlo sans remise, la base d'estimation étant cette fois de 1.000.000 échantillons. En tirant au hasard, sans remise, dans la population des 1113 éléments décrite ci-dessus, nous

Table 4: Exemple de données

Régions	01+11	02	03+12	04+17	05	06+13+14+15+16	07	08	09+10	Total
Population	29	32	201	84	45	648	47	17	10	1113
Échantillon	17	15	79	27	26	218	18	7	0	407

obtenons de 850 à 910 échantillons dont le χ^2 est plus fort que 16,49, établissant sa probabilité extrême à environ 0,0009, proche de celle obtenue par la formule (3).

Force nous est d'admettre que les différentes régions du Québec n'étaient pas adéquatement représentées dans cet échantillon.

Références

- Cochran, W. C. (1963). *Sampling techniques* (2^e édition). New York, Wiley.
- Kendall, M., Stuart, A. (1979). *The advanced theory of statistics. Volume 2 : Inference and relationship* (4^e édition). New York : Macmillan.
- Kish, L. (1965). *Survey sampling*. New York : Wiley
- Laurencelle, L. (2001). *Hasard, nombres aléatoires et méthode Monte Carlo*. Québec : Presses de l'Université du Québec.
- Robert, M. (1988). Validité, variables et contrôle, dans M Robert (dir.) : *Fondements et étapes de la recherche scientifique en psychologie* (3^e édition) (p. 79-118). Saint-Hyacinthe (Qc) : Edisem.
- Rossi, P. H., Wright, J. D., Anderson, A. B. (1983). *Handbook of survey research*. New York : Academic Press.
- Thompson, M. E. (1997). *Theory of sample surveys*. Londres : Chapman & Hall.
- Thompson, S. K. (1992). *Sampling*. New York : Wiley.
- Voyer, J. B., Valois, P., Rémillard, B. (2000). La sélection des participants, dans RJ Vallerand et U Hess : *Méthodes de recherche en psychologie* (p. 91-129). Montréal : Gaëtan Morin.
- Yates, F. (1960). *Sampling methods for censuses and surveys* (3^e édition). Londres : Charles Griffin.