JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

RODOLFO DE CRISTOFARO

L'influence du plan d'échantillonnage dans l'inférence statistique

Journal de la société statistique de Paris, tome 137, n° 4 (1996), p. 23-34

<http://www.numdam.org/item?id=JSFS_1996__137_4_23_0>

L'INFLUENCE DU PLAN D'ECHANTILLONNAGE DANS L'INFÉRENCE STATISTIQUE

Rodolfo de CRISTOFARO

Université de Florence[1]

Résumé

Selon un point de vue très répandu, le principe de vraisemblance est une conséquence directe du théorème de BAYES. Toutefois, cette déduction est loin d'être immédiate. En particulier, la connaissance du plan d'échantillonnage est une information qui est capable d'influencer la loi *a priori*. Ceci conduira à proposer une solution rationnelle au vieux problème des probabilités *a priori*. Selon nous, le principe de l'équiprobabilité des hypothèses ne doit pas être identifié avec l'idée d'ignorance, mais avec l'impartialité du plan d'échantillonnage par rapport à l'hypothèse h. On sera ainsi amené à confirmer la règle originaire proposée par JEFFREYS pour le choix de la loi *a priori*.

Abstract

LOGICAL PRIOR PROBABILITIES IMPLIED BY THE INFORMATION ON SAMPLING RULE

According to a widely held opinion, one should make an identical inference about a parameter from the data d irrespective of the design p which yields d. This thesis is not surprising for the supporters of the likelihood principle. Nevertheless, we question the fact that Bayesian inference is design-free. Namely, in principle, the piece of information about p is relevant for the purpose of inference. Only if the sampling rule is impartial with respect to the hypothesis h, the prior can be ignored. In other words, the equiprobability assumption is to be linked to the idea of the impartiality of sampling with respect to h. In line with such a conception, in this paper prior probabilities are determined in an objective way according to the performed sampling.

Key-words and phrases : Bayes' theorem, data translated likelihood, Jeffreys' rule, likelihood principle, prior probabilities, statistical inference.

---

1. Università di Firenze, Rodolfo de Cristofaro, Dipartimento di Statistica «*G. Parenti*», viale Morgagni, 59, I-50134 Florence (Italie), fax 00 39 55 422 35 60, # E-Mail : decrist @ stat.ds.unifi.it

Une première version de ce texte a été présentée aux XXVIII[es] Journées de la Statistique de l'Association pour la Statistique et ses Utilisations (ASU), du 27 au 30 mai 1996, à Québec, Canada.

1. A propos du principe de vraisemblance

Selon un point de vue très répandu, le principe de vraisemblance (p.v.) est une conséquence directe du théorème de BAYES (cf. BARNETT 1982, p. 226). En vérité, la déduction de ce principe n'est pas immédiate. BERGER & WOLPERT (1988, p. 19) donnent la version suivante du p.v. : *"All the information about θ obtainable from an experiment is contained in the likelihood function for θ given x. Two likelihood functions for θ (from the same or different experiments) contain the same information about θ if they are proportional to one another"*.

LINDLEY (1965, p. 59) indique aussi des conditions semblables auxquelles deux ensembles de données devraient se soumettre pour satisfaire le p.v. Il ajoute que *"the principle is immediate from* BAYES*'s theorem because the posterior distributions from the two sets will be equal"*. Toutefois, le point concerne justement la densité *a priori* des paramètres que LINDLEY suppose implicitement indépendants de la règle d'échantillonnage (ce qui serait en fait à démontrer). La question mérite d'être approfondie à partir de la nature conditionnelle de la probabilité.

Les affirmations de la théorie de la probabilité, comme dans toute la Mathématique, sont exprimées sous forme conditionnelle. Les présupposés de l'assertion conditionnelle sont normalement omis, et la conséquence seule est retenue, lorsque de tels présupposés semblent évidents du fait de leur contexte. Autrement, il serait opportun de toujours préciser l'antécédent conditionnel, comme il serait convenable de le faire pour toute application de la Mathématique. En particulier, bien que l'adjectif «conditionnel» soit normalement omis et que la probabilité d'un événement soit habituellement notée $\mathrm{p}(E)$, celle-ci peut être différente parce que l'espace des échantillons Ω est différent, ou, si l'on préfère, parce que l'expérience à laquelle on fait référence est différente. En conclusion, $\mathrm{p}(E)$ est une simplification. La notation de la probabilité dans son expression complète serait : $\mathrm{p}(E \,|\, \Omega)$.

Cette exigence a été soulignée par les défenseurs de la conception «logique» de la probabilité et notamment par R. CARNAP. Citons les paroles de CARNAP (1962, p. 31) : "The omission of any reference to evidence is often harmless". En accord avec cette nature conditionnelle, la probabilité d'une hypothèse h (appelée par CARNAP *degree of confirmation*) est conditionnée à l'expérience e et devrait toujours être notée $c(h, e)$. Du reste, cette nature conditionnelle est clairement reconnue et mise en évidence dans le traité bien connu sur la probabilité, écrit par KEYNES en 1921. Permettez-moi en outre de citer les mots de Sir Arthur EDDINGTON (écrits en 1935) : *"We can only speak of "probability in the light of certain given information", and the probability alters according to the extent of the information"*.

Une conséquence directe de la nature conditionnelle de la probabilité est d'utiliser toute l'information, implicite dans l'expérience, qui est capable d'influencer le résultat de l'inférence inductive. BARNETT (1982, p. 68) écrit à ce sujet (se référant à GOOD) : *"Again probability is an entirely conditional concept"*. Que signifie cela ? Cela signifie qu'elle est conditionnée à l'expérience e (tout ce qui est connu).

En effet, dans l'expérience, tout ce qui est connu doit être inclus. Si le plan d'échantillonnage p adopté dans l'enquête est connu, il est une partie de l'expérience disponible, qui ne devrait pas être ignorée. C'est-à-dire, en règle générale, que p est capable d'influencer la probabilité de l'hypothèse h.

Du reste, la probabilité des données est non définie, si, en plus de l'hypothèse h, on ne spécifie pas p. Ainsi, la probabilité de x succès en n épreuves est différente suivant le processus supposé de production des données (qui peut être constitué par un échantillonnage direct, inverse, avec ou sans remise, Polya, intentionnel, et ainsi de suite). En d'autres termes, p entre toujours dans l'inférence à cause de sa présence dans la détermination de la probabilité des données : $\mathrm{p}(d \mid h,\ p)$.

D'autre part, si $\mathrm{p}(d \mid h,\ p)$ est indéfini quand p n'est pas connu, la même chose est valable pour $\mathrm{p}(h \mid d,\ p)$, en vertu de la formule de Bayes :

$$\mathrm{p}(h \mid d,\ p) \propto \mathrm{p}(h \mid p)\ \mathrm{p}(d \mid h,\ p). \tag{1}$$

Selon certains, toutefois, l'information sur le plan d'échantillonnage, contenue dans la vraisemblance $\mathrm{p}(d \mid h,\ p)$, est habituellement absorbée dans la constante de proportionnalité. Par exemple, considérons des épreuves indépendantes avec pour probabilité de succès θ (constante pour chaque épreuve). L'observation de x succès sur n épreuves pourrait s'effectuer au moins de deux façons différentes : en préétablissant le nombre des épreuves, ou bien le nombre de succès. Et, comme l'on sait, dans les deux cas, les vraisemblances sont proportionnelles entre elles. On en déduit alors que l'inférence sur θ devrait être la même, aussi bien pour l'échantillonnage *direct* que pour l'échantillonnage *inverse*.

En réalité, les arguments rapportés ci-dessus négligent la présence de $\mathrm{p}(h \mid p)$ dans la relation (1), d'où il résulte que la probabilité de h dépend de p.[2] C'est-à-dire que l'*a posteriori* est le résultat d'un processus où p est toujours présent (si l'on préfère, afin de caractériser l'*a posteriori*, il n'est pas possible d'ignorer p).

Certains statisticiens trouvent surprenant que l'*a priori* puisse dépendre du plan d'échantillonnage. Ils supposent que leur «opinion» à propos d'une hypothèse restera la même quel que soit le plan d'échantillonnage. Nous pensons que l'erreur est conceptuelle. On croit que la probabilité *a priori* est un degré de croyance qui *précède* l'information expérimentale, mais la connaissance du plan d'échantillonnage est elle-même *une information* qui précède le recueil des données et elle est donc capable d'influencer l'*a priori*.

Suivant la théorie qui domine actuellement, la règle d'échantillonnage (ou *"stopping rule"*) peut être ignorée quand le recueil des observations ne dépend pas des informations disponibles sur θ. Mais, d'habitude, le plan

2. Il est juste que l'information sur le plan d'échantillonnage contenue dans la vraisemblance soit absorbée dans la constante de proportionnalité, car, différemment, on en tiendrait compte deux fois : d'abord dans la vraisemblance, puis dans l'*a priori*.

d'échantillonnage est décidé par les statisticiens d'enquêtes sur la base des informations disponibles sur θ. Alors, le plan d'échantillonnage est informatif dans le cas où il est décidé consciemment. Il n'est pas informatif si l'on préfère ignorer que le plan peut influencer l'inférence. Une conclusion assez incohérente.

Un contre-exemple au p.v. est donné par les échantillonnages «*direct*» et «*inverse*» cités ci-dessus : l'échantillonnage inverse se termine toujours par un succès, qu'il serait juste de pénaliser dans l'inférence. Un autre contre-exemple est le suivant : Au cours d'un référendum pour l'abrogation d'une loi, le dépouillement des scrutins de tous les bureaux de vote se poursuit jusqu'à ce que l'on obtienne r votes consécutifs en faveur de la loi. Selon les défenseurs du p.v., à égalité de données, l'inférence devrait être la même par rapport à un échantillonnage aléatoire direct. Mais je pense que les promoteurs du référendum ne seraient pas d'accord avec cette théorie. En effet, toute argumentation sur l'insignifiance du plan d'échantillonnage servirait seulement à jeter le discrédit sur la Statistique, en tant que discipline.

Parmi les contre-exemples au p.v. fournis par la littérature statistique, nous comptons celui de STEIN (1962). Un autre exemple est le paradoxe de la *"stopping rule"* de SAVAGE (1962, pp. 75-76). D'autres contre-exemples ont été formulés par GODAMBE (1966), SCOTT (1977). Un exemple, proposé initialement par STONE contre l'acceptation d'un *a priori* uniforme, se présente en réalité comme un contre-exemple au p.v. (comme le notait FRASER dans la discussion.

Quand SAVAGE (Savage et coll., 1962, p. 76) eut connaissance des conséquences du p.v., il fut scandalisé de ce qu'un statisticien puisse soutenir une idée aussi manifestement fausse, mais il revint plus tard sur ses propos. Toutefois, c'est sa première impression qui apparaît maintenant correcte.

Lorsque l'échantillonnage est intentionnellement biaisé afin de favoriser une hypothèse fausse, disons h', il paraît raisonnable de poser $\mathrm{p}(h' \mid p)$ égal à zéro. De la même manière, il semble raisonnable de retenir que p est indépendant de h quand l'échantillonnage est impartial par rapport à h. Dans ce cas, $\mathrm{p}(h \mid p)$ peut être absorbé dans la constante de proportionnalité et

$$\mathrm{p}(h \mid d,\ p) \propto \mathrm{p}(d \mid h,\ p). \tag{2}$$

Comme il est mis en évidence par la relation (2), nous pouvons voir si la condition d'impartialité est satisfaite en regardant la courbe de vraisemblance en fonction de h (pour voir si p favorise quelques hypothèses). Mais nous reviendrons sur ce point par la suite.

Naturellement, p peut être impartial par rapport à un certain paramètre θ, mais non par rapport à un autre. Dans le prochain paragraphe, nous réexaminerons brièvement certaines propositions faites par plusieurs statisticiens à propos de la spécification de l'*a priori* dans le cas où la vraisemblance pour un paramètre θ est connue.

2. Les règles pour le choix de la loi *a priori*

Le problème le plus important de la solution bayésienne de l'inférence concerne le choix de la loi *a priori*. Que représentent les probabilités *a priori* et comment peuvent-elles être obtenues (cf. O'HAGAN 1994, p. 11) ? En réalité, il y a beaucoup d'insatisfaction à l'intérieur de la théorie bayésienne (cf. BARNETT 1982) quant à la manière dont il faudrait procéder si, *a priori*, nous ne savons rien de θ.

JEFFREYS (1961) proposa d'utiliser l'information $I(\theta)$ de FISHER pour exprimer cet état d'ignorance, afin de satisfaire la condition d'invariance par rapport aux transformations de variables. Dans le cas d'un seul paramètre, l'*a priori* de JEFFREYS pour θ (appelé *non informatif*) est proportionnel à la racine carrée de $I(\theta)$.

Un certain nombre d'objections ont été faites à la solution de JEFFREYS. En particulier, il a été observé que la règle de Jeffreys dépend de la manière dont les épreuves ont été réalisées et, dans le cas de deux différentes épreuves, dont chacune se réfère au même paramètre, le choix de l'*a priori* peut être différent. Par exemple, l'*a priori* pour θ dans l'échantillonnage direct (épreuves de Bernoulli) est différent de l'*a priori* pour θ dans l'échantillonnage inverse (épreuves de Pascal). Ceci transgresse le principe de vraisemblance. Toutefois, nous l'avons déjà vu, ceci n'est pas une objection, étant donné que l'*a priori* doit dépendre précisément de la façon dont les données ont été recueillies.

Au contraire, il est possible d'affirmer que la règle de Jeffreys est une réponse raisonnable au problème de l'évaluation de l'information contenue dans le processus de production des données.

Par exemple, l'*a priori* de Jeffreys pour le paramètre θ de la binômiale est une distribution Bêta avec paramètres $\alpha = 1/2$ et $\beta = 1/2$, qui est en forme de U. Puisque la probabilité est une mesure du degré de certitude d'une prévision basée sur les informations disponibles sur θ, il est évident que la certitude est maximum quand $\theta = 0$ ou 1, et minimum quand $\theta = 1/2$. La forme en U de la Bêta confirme la validité de l'*a priori* de Jeffreys.

Dans l'échantillonnage inverse, l'*a priori* de Jeffreys est une distribution Bêta (impropre) avec pour paramètres $\alpha = 0$ et $\beta = 1/2$. Cette solution aussi apparaît raisonnable à partir du moment où l'échantillonnage se termine toujours par un succès, qu'il semble juste de pénaliser dans l'inférence.

On peut noter que, comme cela a été démontré par BERNARDO (1979), en utilisant l'information de Lindley, dans des conditions très générales, l'*a priori* de Jeffreys est excellent dans la mesure où il rend maximum l'information d'échantillonnage et minimum l'information *a priori*.

Une autre objection à la règle de Jeffreys concerne la possibilité d'obtenir un *a priori* impropre. Mais, dans un tel cas, l'*a priori* est admis par approximation. Par exemple, un *a priori* uniforme sur l'axe réel entier R doit être considéré comme un *a priori* «localement» uniforme, excluant l'intervalle dans lequel la vraisemblance est presque égale à 0. De toute façon, avec un système d'axiomes différent (de nature comparative ou qualitative), les distributions impropres de Jeffreys deviennent propres (cf. GOODMAN 1977).

D'autres objections ont été faites dans le cas de plusieurs paramètres. Mais nous reviendrons sur ce point par la suite.

Une autre formulation de la représentation de l'ignorance est basée sur la mesure de l'entropie. En utilisant la célèbre formule de Shannon, l'entropie de la densité $f(\theta)$ peut être pensée comme une mesure de la façon dont $f(\theta)$ est non informatif par rapport à θ. Maintenant, pour représenter l'absence d'information, on peut utiliser la densité *a priori* $f(\theta)$ qui rend l'entropie maximum. Ceci est précisément la solution proposée par JAYNES dans une série d'études faites de 1968 à 1983. Il a également utilisé le principe d'invariance de JEFFREYS de manière à éviter une incohérence dans le cas d'un changement de variable.

A leur tour, BOX & TIAO (1973) ont proposé de choisir l'*a priori* comme un standard de référence à utiliser lorsque l'on prend une attitude neutre dans l'inférence (appelé *jury principle*). En particulier, la distribution *a priori* de θ est rendue uniforme quand différents ensembles de données décrivent par translation la courbe de vraisemblance sur l'axe de θ, la laissant pour le reste inchangée. Dans le cas où cela ne se vérifie pas, il faut trouver une transformation de variable $\phi(\theta)$ telle que (au moins approximativement) divers ensembles de données entraînent une translation de la courbe de vraisemblance pour ϕ (interprété comme un paramètre de position). Le critère est aussi connu comme *"data translated likelihood"*.

Par exemple, dans le cas d'un échantillonnage aléatoire de n observations d'une population normale avec moyenne μ (quand la variance est connue), les données entrent dans la vraisemblance seulement au travers de la moyenne d'échantillonnage $\overline{y}$ (un résumé exhaustif de μ), et différentes valeurs de $\overline{y}$ donnent une translation de la courbe de vraisemblance sur l'axe μ ; sinon, elle reste inchangée. En conséquence, l'*a priori* est «localement» uniforme dans la métrique originaire. Au contraire, pour la variance d'une population normale (quand la moyenne est connue), l'*a priori* est «localement» uniforme par rapport à $\log \sigma$. En effet, les données entrent dans la vraisemblance seulement au travers de la variance d'échantillonnage s^2 et les courbes de vraisemblance prennent une forme différente pour différentes valeurs de s quand on se réfère à la métrique originaire. Afin que la vraisemblance soit *"data translated"*, il faut, justement, l'exprimer en fonction de $\log \sigma$.

Dans le cas du paramètre θ de la binômiale, l'*a priori* ne peut être supposé uniforme par rapport au même θ, mais par rapport à

$$\arcsin \sqrt{\theta},$$

transformation qui est cohérente avec la règle de JEFFREYS.

Comme l'ont démontré BOX & TIAO (p. 45), la transformation de variable qui permet de prendre un *a priori* uniforme, est différente pour le paramètre θ qui se rapporte aux épreuves de Pascal (échantillonnage inverse), confirmant ainsi la dépendance de l'*a priori* de l'échantillonnage adopté. En particulier,

$$\phi(\theta) = \log \frac{1 - \sqrt{(1-\theta)}}{1 + \sqrt{(1-\theta)}}.$$

BOX & TIAO eux-mêmes ont démontré que, sous certaines conditions, leur critère coïncide essentiellement avec celui proposé par JEFFREYS, basé sur la mesure de l'information de FISHER. Ceci permet d'utiliser la règle de JEFFREYS dans le but de trouver la transformation de variable $\phi(\theta)$ adéquate au cas que l'on considère.

On peut remarquer que la condition de translation de la vraisemblance n'est pas suffisante pour la déduction d'une loi *a priori* uniforme. En effet, la condition en question est nécessaire, mais, pour être suffisante, l'*a priori* doit être déduit en accord avec la règle de Jeffreys. A ce propos, on peut comparer l'*a priori* pour le paramètre θ qui se rapporte aux épreuves de Bernoulli et de Pascal (les deux transformations sont *"data translated"*, aussi bien pour l'échantillonnage direct que pour l'inverse).

Dans le cas de plusieurs paramètres, étrangement, BOX & TIAO ne suivent pas la règle originaire de JEFFREYS, qui consiste à rendre l'*a priori* pour le vecteur Θ proportionnel à la racine carrée du déterminant de la matrice d'information $I(\Theta)$. Mais nous reviendrons sur ce point par la suite.

Nous renvoyons le lecteur à BERNARDO & SMITH (1994) pour une étude détaillée du problème de l'*a priori* «*non informatif*».

3. L'impartialité du plan d'échantillonnage

Selon JEFFREYS, l'*a priori* «*non informatif*» est une représentation d'ignorance totale. En réalité, nous n'avons pas affaire à une ignorance totale, à partir du moment où le plan d'échantillonnage est connu. En particulier, l'équiprobabilité ne doit pas être liée à l'idée d'ignorance, mais à l'impartialité du plan d'échantillonnage par rapport au paramètre étudié.

A ce sujet, le critère proposé par BOX & TIAO semble être celui qui est le plus apte à exprimer cette idée d'impartialité. Comme nous l'avons déjà dit auparavant, dans le but de prendre une densité *a priori* uniforme pour θ, le plan d'échantillonnage doit être impartial. C'est-à-dire que la valeur la plus vraisemblable de θ, disons $\widehat{\theta}$, doit être le point d'équilibre en comparaison des autres valeurs possibles de θ, dans le sens où l'échantillonnage ne doit pas non plus favoriser un certain écart particulier de θ par rapport à $\widehat{\theta}$. En bref, la vraisemblance d'écarts identiques par rapport à $\widehat{\theta}$ doit être la même, quel que soit $\widehat{\theta}$.

Puisque l'*a priori* est supposé uniforme, nous pouvons faire référence à l'estimateur du maximum de vraisemblance t. Et l'échantillonnage sera impartial par rapport à θ si la courbe de vraisemblance pour t est complètement déterminée *a priori* comme une courbe symétrique ne favorisant aucun écart possible de t de son maximum, quelle que soit l'observation des données. En d'autres termes, quand la vraisemblance est *data translated* pour θ, nous pouvons dire que l'échantillonnage est impartial par rapport au même θ.

Naturellement, si la courbe de vraisemblance n'est pas symétrique et invariante dans la forme, il sera nécessaire de recourir à une transformation de variable en

accord avec la règle de Jeffreys, comme le proposent BOX & TIAO. Les exemples et les graphiques rapportés par BOX & TIAO (1973, pp. 27-39) confirment cette propriété de la symétrie et de l'invariance de la forme de la vraisemblance, quel que soit l'échantillon observé, quand le plan d'échantillonnage est impartial par rapport à la transformation $\phi(\theta)$. Et, jusqu'à présent, il n'y a pas de contre-exemples.

Le choix de l'estimateur du maximum de vraisemblance se justifie aussi pour un autre motif. On sait en effet que, s'il existe, l'estimateur de vraisemblance t est un *résumé exhaustif* de θ. Nous pouvons dire alors que, si l'échantillonnage est impartial par rapport à θ, seul t est important pour obtenir des inférences sur θ, indépendamment du plan d'échantillonnage adopté[3].

On note un cas où le plan d'échantillonnage n'est pas impartial dans la métrique originaire : celui de la binômiale. En effet, la binômiale n'est symétrique que pour $\theta = 1/2$. Toutefois, avec l'augmentation du nombre des succès x et des faillites $n-x$, la binômiale s'approche de la normale, qui est symétrique, en confirmant et en précisant ainsi (abstraction faite d'autres informations) la validité du choix d'une loi *a priori* uniforme dans le cas de grands échantillons (disons : x et $n-x > 20$).

La loi *a priori* est ainsi obtenue par la connaissance de l'information de FISHER et donc par la connaissance de la courbe de vraisemblance dans sa complète expression (y compris la partie habituellement absorbée dans la constante de proportionnalité). Dans un certain sens, le principe de vraisemblance est confirmé dans son sens original à partir du moment où toute l'information expérimentale (y compris celle sur le plan d'échantillonnage) est contenue dans la vraisemblance.

O'HAGAN (1994, p. 12) écrit : *"The Bayesian method requires prior probabilities to be given explicit values. A logical prior probability must be the unique value logically implied by the available prior information. Unfortunately, for almost every kind of prior information, this value cannot be found; the necessary theory simply does not exist"*. Au contraire, nous avons vu qu'il est possible de fournir des valeurs explicites pour les probabilités *a priori* basées sur l'information contenue dans le plan d'échantillonnage.

Même les objections faites à l'utilisation de la formule de BAYES semblent disparaître. RÉNYI (1966, p. 77) écrit à ce propos : «Le théorème de Bayes est parfaitement démontré, personne ne met en doute sa justesse ; c'est seulement de ses applications pratiques qu'on dispute... Mais les $P(B_k)$ [les probabilités *a priori*] sont souvent inconnues et on leur attribue généralement des valeurs arbitraires ; c'est ce procédé qui est véritablement contestable». En effet, un tel arbitrage n'a pas de raison d'être, tout au moins quand le plan d'échantillonnage est connu.

A ce propos, on peut noter que FISHER n'a jamais nié l'importance de la formule de BAYES au cas où les probabilités *a priori* seraient obtenues de

3. On peut remarquer que la notion d'*exhaustivité* n'est plus la même si l'échantillonnage n'est pas impartial par rapport à θ.

façon objective et c'est pourquoi il devrait maintenant accepter l'emploi de cette dernière dans l'induction. On peut dire la même chose à propos d'autres statisticiens, comme NEYMAN et WALD, qui ont proposé des méthodes différentes d'inférence en raison de leur insatisfaction en ce qui concerne l'attribution des probabilités *a priori*.

4. La règle générale de Jeffreys

Comme nous l'avons déjà dit auparavant, dans le cas multiparamétrique, la règle de JEFFREYS consiste à rendre l'*a priori* pour le vecteur Θ proportionnel à la racine carrée du déterminant de la matrice d'information $I(\Theta)$. Cette règle a été modifiée par JEFFREYS lui-même, car : *"though consistent, leads to results that appear to differ too far from current practice"*. La pratique à laquelle on se réfère concerne la réduction du nombre des degrés de liberté, qui s'obtient quand l'estimation d'un paramètre (comme par exemple la variance) requiert la connaissance d'un autre paramètre (comme par exemple la moyenne).

La règle modifiée consiste à obtenir, d'abord, la distribution *a priori* de chaque paramètre et de supposer, ensuite, qu'ils sont indépendants entre eux. Par exemple, dans le cas de la distribution normale, avec moyenne et variance non connues, le déterminant de la matrice d'information de Fisher est égal à $2\sigma^{-4}$ et donc

$$p(\mu,\ \sigma) \propto \sigma^{-2}. \tag{3}$$

Avec l'utilisation de la règle modifiée,

$$p(\mu,\ \sigma) = p(\mu \mid \sigma)\ p(\sigma \mid \mu) \propto \sigma^{-1}. \tag{4}$$

La différence dans ce cas est peu évidente, mais la règle originaire semble préférable pour différentes raisons. Tout d'abord pour des raisons de cohérence. En effet, l'invariance par rapport aux transformations de variables, sur laquelle se base la proposition de JEFFREYS, est assurée par la règle générale et non par celle qui est modifiée. D'autre part, la réduction du nombre des degrés de liberté se justifie dans la théorie «non-bayésienne» de l'inférence par le fait que la distribution d'échantillonnage est conditionnée par la moyenne $\overline{y}$ (d'où l'expression *degrés de liberté*) ; mais pour une correcte formulation du problème de l'induction un tel lien n'existe pas et n'a pas de raison d'être.

JEFFREYS, lui-même, introduisit un autre argument afin de justifier la règle modifiée, repris par la suite par d'autres statisticiens. Il s'agit de l'indépendance entre les croyances *a priori* concernant la moyenne et celles concernant la variance. En réalité, cette indépendance n'existe pas, puisque l'inférence sur la moyenne est influencée par la connaissance de la variance. En particulier, plus la variance est grande et plus d'observations sont nécessaires pour l'estimation de la moyenne : à la limite, avec une variance égale à zéro, il suffit d'une seule observation pour déterminer la moyenne de façon exacte.

Encore une fois, l'erreur est conceptuelle. C'est-à-dire que l'on retient que l'*a priori* dépend des connaissances qui précèdent l'enquête, ignorant le plan d'échantillonnage, qui, en revanche, joue un rôle dans la détermination de l'*a priori*.

Le même critère de la translation nous amène à confirmer la règle originaire de Jeffreys. En particulier, l'expression de la vraisemblance pour la distribution normale est de la forme (BOX & TIAO 1973, p. 50) :

$$(5) \qquad l(\mu, \sigma | \mathrm{x}) \propto F\left(\frac{\mu - \widehat{\mu}}{\widehat{\sigma}}, \log\sigma - \log\widehat{\sigma}\right) G(\log\sigma - \log\widehat{\sigma}).$$

où l'on voit que la forme logarithmique apparaît aussi bien en F qu'en G. En d'autres termes, l'absence de connaissance de l'écart type intervient deux fois : dans la représentation de la moyenne et dans celle de l'écart type lui-même.

On remarque en outre que le *paradoxe de marginalisation* intervient lorsqu'on utilise la règle originaire et non la règle modifiée, comme le confirme une lecture soignée de l'article de DAWID, STONE & ZIDEK (1973). Il y a enfin des cas (comme pour le modèle aléatoire d'analyse de la variance) où il faut utiliser la règle originaire pour certains paramètres et la règle modifiée pour d'autres (cf. TIAO & TAN, 1965). Un exemple de cohérence qui étonne vraiment.

5. Le cas d'un échantillonnage de populations finies

Selon la plupart des statisticiens, la règle d'échantillonnage peut être ignorée quand elle ne dépend pas des informations disponibles sur θ. Au contraire, le plan d'échantillonnage, pour être ignoré, doit être réalisé de façon à ce qu'il soit impartial par rapport à θ.

Toutefois, le raisonnement habituellement suivi pour les populations finies est tout à fait particulier. Par exemple, admettons un échantillonnage aléatoire d'une population finie et considérons comme paramètre le vecteur

$$\mathbb{Y} = (y_1, y_2, \ldots, y_N)$$

des observations d'une variable Y associée aux N individus de la population dans un ordre déterminé. En comparaison à toute détermination paramétrique de $\mathbb{Y}$ dans l'échantillon, la règle d'échantillonnage semblerait impartiale : tous les échantillons ordonnés sont équiprobables, quel que soit $\mathbb{Y}$.

Supposons maintenant que la variable Y prenne comme valeur 1 ou 0, selon que les individus possèdent ou non un certain caractère (disons : boule blanche ou non blanche) et indiquons par θ la fraction de boules blanches. Abstraction faite des étiquettes, la probabilité des échantillons possibles dépend de θ. En particulier, la distribution est binômiale si l'échantillonnage est avec remise. Maintenant, la distribution binômiale est symétrique pour $\theta = 1/2$ et asymétrique dans tous les autres cas. Cela signifie que l'échantillonnage n'est pas neutre par rapport à θ. Il le serait par rapport aux étiquettes. Mais l'inférence concerne la variable associée aux individus et non les étiquettes !

Tout ceci pourrait sembler évident. Au contraire, l'échantillonnage de populations finies suit une tendance conditionnée par une formulation discutable du principe de vraisemblance. A ce sujet, faisons référence à CHAUDHURY & VOS (1988, pp. 65-75), où ils montrent qu'il faudrait effectuer une inférence identique sur l'état de nature inconnue Y à partir des données d en faisant abstraction du plan de sondage qui a généré d.

Les conséquences de cette formulation du problème sont encore plus surprenantes. Ainsi l'échantillonnage aléatoire semblerait inutile. Un paradoxe appelé par GODAMBE (1969) : *"the problem of randomization"*. D'autres statisticiens soutiennent que l'opération en vue de donner un caractère aléatoire à l'échantillon est, en revanche, très importante (à ce propos, on peut consulter GELMAN, CARLIN, STERN & RUBIN, 1995). De toute façon, le sujet mérite d'être reconsidéré, en vérifiant toujours l'impartialité du plan d'échantillonnage par rapport au paramètre considéré.

Nous pouvons remarquer que le résultat de cette étude représente un apport révolutionnaire pour les fondements de l'échantillonnage et de l'inférence statistique en comparaison de la théorie qui domine actuellement la discipline.

Références

BARNETT, V. (1982) *Comparative Statistical Inference*, J. Wiley, New York.

BERGER, J.O. & WOLPERT, R.L. (1988) *The Likelihood Principle*, Inst. of Math. Statistics, Hayward.

BERNARDO, J.M. (1979) "Reference Posterior Distributions for Bayesian Inference", *J. Royal Statistical Society,* B 41, 113-147.

BERNARDO, J.M. & SMITH, A.F.M. (1994) *Bayesian Theory*, Chichester : Wiley Series in Probability and Mathematical Statistics.

BOX, G.E.P. & TIAO, G.C. (1973) *Bayesian Inference in Statistical Analysis*, Addison Wesley, Californie.

CARNAP, R. (1962) *Logical Foundations of Probability*, The University of Chicago Press.

CHAUDHURI, A. & VOS, J.W.E. (1988) *Unified Theory and Strategies of Survey Sampling*, North-Hollet Series in Statistics and Probability : Vol. 4. North-Holland.

DAWID, A.P., STONE, N. & ZIDEK, J.V. (1973) "Marginalization Paradoxes in Bayesian and Structural Inference", *J. Royal Statistical Society,* B 35, 189-223.

DE CRISTOFARO, R. (1996) "A propos du principe de vraisemblance", *Recueil des résumés des communications des XXVIII*^{es} *Journées de la Statistique,* ASU, Université Laval, Québec.

GELMAN, A., CARLIN, J.B., STERN, H.S. & RUBIN, D.B. (1995) *Bayesian Data Analysis*, Londres : Chapman & Hall.

GODAMBE, V.P. (1966) "A New Approach to Sampling from Finite Populations. I : Sufficiency and Linear Estimation", *J. Royal Statistical Society,* B 28, 310-319.

GODAMBE, V.P. (1969) "A Fiducial Argument with Application to Survey Sampling", *J. Royal Statistical Society,* B 31, 246-260.

GOODMAN, T.N.T. (1977) "Qualitative Probability and Improper Distributions", *J. Royal Statistical Society,* B 39, 387-393.

JEFFREYS, H. (1961) *Theory of Probabilities*, Clarendon Press, Oxford.

LINDLEY, D.V. (1965) *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Vol. 2, Cambridge University Press.

O'HAGAN, A. (1994) *Kendall's Advanced Theory of Statistics,* vol. 2B : *Bayesian Inference*, Londres : Arnold.

RÉNYI, A. (1966) *Calcul des probabilités*, avec un appendice sur la théorie de l'information. Traduit par C. Bloch, Paris : Dunod.

SAVAGE, L.J. (1962) *The Foundations of Statistical Inference*, Methuen, Londres.

SAVAGE, L.J. ET COLL. (1962) *The Foundations of Statistics : A Discussion*, Methuen, Londres.

SCOTT, A.J. (1977) "On the Problem of Randomization in Survey Sampling", *Sankhya,* C, 1-9.

STEIN, C. (1962) "A Remark on the Likelihood Principle", *J. Royal Statistical Society,* A 125, 565-568.

STONE, M. (1976) "Strong Inconsistency from Uniform Priors", *J. Amer. Statist. Ass.,* 71, 114-123.

TIAO, G.C. & TAN, W.Y. (1965) "Bayesian Analysis of Random-Effect Models in the Analysis of Variance. I.- Posterior Distribution of Variance-Components", *Biometrika,* 52, 37-52.