



## Mathématiques et sciences humaines

Mathematics and social sciences

167 | Automne 2004

Varia

---

# Probabilités, démographie et sciences sociales

*Probability, demography and social sciences*

Daniel Courgeau

---



### Édition électronique

URL : <http://journals.openedition.org/msh/2933>

DOI : 10.4000/msh.2933

ISSN : 1950-6821

### Éditeur

Centre d'analyse et de mathématique sociales de l'EHESS

### Édition imprimée

Date de publication : 1 septembre 2004

ISSN : 0987-6936

### Référence électronique

Daniel Courgeau, « Probabilités, démographie et sciences sociales », *Mathématiques et sciences humaines* [En ligne], 167 | Automne 2004, mis en ligne le 10 mars 2006, consulté le 01 mai 2019. URL : <http://journals.openedition.org/msh/2933> ; DOI : 10.4000/msh.2933

---

PROBABILITÉS, DEMOGRAPHIE ET SCIENCES SOCIALES<sup>1</sup>Daniel COURGEAU<sup>2</sup>

RÉSUMÉ — *Cet article replace les diverses notions de probabilité, apparues dès le XVII<sup>e</sup> siècle, dans le développement simultané des sciences sociales. Si l'approche objectiviste, pour laquelle la probabilité est la fréquence limite atteinte après un nombre infini d'épreuves, a longtemps prévalu, l'approche épistémique, pour laquelle la probabilité concerne l'évaluation des degrés auxquels il est raisonnable de croire en la vérité de propositions, semble mieux convenir pour les sciences sociales. Une présentation et une discussion de ces deux approches sont faites ici, avec de nombreux exemples d'application aux sciences sociales.*

MOTS-CLÉS — Démographie, Inférence statistique, Probabilité objectiviste, Probabilité épistémique, Sciences sociales

SUMMARY — Probability, demography and social sciences  
*This paper considers the different approaches to probability, that has been developed from the XVII<sup>th</sup> century, compared to the simultaneous development of social sciences. If the frequency-type approach, for which the probability is the mathematical limit to which the relative frequency of an outcome converges, the epistemological approach, for which the probability concerns the available evidence it is reasonable to have a degree of belief in the truth of a given proposition, seems to be more suitable for social sciences. A presentation and a discussion of these two approaches are made here, with numerous examples of application to social sciences.*

KEY WORDS — Demography, Frequency-type probability, Epistemological probability, Social sciences, Statistical inference.

Le concept moderne de probabilité émerge avec les réflexions de Fermat et de Pascal, publiées à la fin de la monographie de ce dernier sur le triangle arithmétique [Pascal, 1651], en vue de résoudre le problème des partis ; la démographie, l'épidémiologie et bien d'autres sciences sociales trouvent leur origine dans l'ouvrage de Graunt [1662], qui estime, entre autres, les probabilités de mortalité par cause dans la ville de Londres, en transformant des listes d'individus en des tableaux statistiques ; la même année paraît la *Logique* de Port Royal [Arnauld et Nicole, 1662], dont les derniers chapitres sont à l'origine d'une nouvelle forme d'inférence non-déductive, l'inférence probabiliste en contexte d'incertitude. La convergence de ces dates, montre la simultanéité d'apparition de ces concepts, et nous allons voir ici les liens étroits qui

<sup>1</sup> Article reçu le 15 mai 2004, révisé le 18 juillet 2004 ; accepté le 28 juillet 2004.

<sup>2</sup> Institut national d'études démographiques, 133 bd Davout Paris 75980 cedex 20, courgeau@ined.fr

se sont établis entre probabilités, démographie, sciences sociales et inférence probabiliste.

Nous avons développé dans un article précédent de cette même revue [Courgeau, 2002], l'évolution de la démographie au cours du temps. Nous n'essayerons pas ici de développer l'évolution des idées sur la probabilité au cours du temps, qui a maintenant fait l'objet de nombreux ouvrages (par exemple [Hacking, 1975, 1990] [Krüger *et al.*, 1986] [Porter, 1986] [Suppes, 2002]), mais plutôt de démêler les liens qui se sont établis entre les probabilités et les sciences sociales. Cependant, avant d'aborder le thème principal de cet article, il est nécessaire d'indiquer rapidement les divers points de vue pris sur les probabilités, pour pouvoir mieux voir ensuite comment la démographie et les sciences sociales se situent par rapport à eux et les problèmes posés par l'inférence statistique considérée comme une méthode d'induction.

## 1. PROBABILITES DIVERSES THEORIES

Si en mathématiques les problèmes des fondements sont résolus dès qu'une axiomatique satisfaisante a été élaborée à partir d'une logique déductive, il n'en est pas de même pour le calcul des probabilités. En effet, ce calcul est maintenant complètement axiomatisé<sup>3</sup> depuis plus d'un demi-siècle [Kolmogorov, 1933], et cependant des controverses se poursuivent, sans discontinuer, sur la nature des probabilités et les possibilités d'inférences qu'elles permettent, en particulier en sciences sociales<sup>4</sup>. Cela vient de ce que les chercheurs, qui utilisent le calcul des probabilités, tendent à le considérer non pas comme un simple objet de calcul, mais comme la formalisation d'un mode de connaissance qu'il importe de mettre en évidence. En effet, les propriétés mathématiques des probabilités s'appliquent à de nombreuses caractéristiques, qui peuvent n'avoir aucun lien entre elles, mais seulement un parallélisme formel avec ces notions – par exemple, la distribution de la masse totale du mobilier présent dans une pièce donnée a les mêmes propriétés qu'une distribution de probabilité [Savage, 1967] mais n'est pas revêtue de la même signification. Enfin l'axiomatisation de Kolmogorov porte sur des ensembles, alors que les probabilités peuvent porter sur des propositions – il n'est pas toujours possible de considérer une proposition se référant au monde réel comme une réunion de propositions élémentaires disjointes d'un ensemble qui aurait une signification dans le cas du problème étudié [Jaynes, 2003]. Il nous faut donc d'abord cerner la nature de ces problèmes et présenter les diverses théories sur la probabilité, avant de voir leur utilisation dans le domaine des sciences sociales.

### 1.1 UNE APPROCHE OBJECTIVISTE – CONCEPTION FRÉQUENTISTE OU PROPENSIONISTE

Le point de vue objectiviste s'appuie sur la loi des grands nombres établie par Jacques Bernoulli<sup>5</sup> [1713], qui nous indique que la valeur du rapport entre le nombre d'événements observés et le nombre d'individus soumis au risque, tend vers la probabilité de cet événement, lorsque la population soumise au risque, tend vers l'infini.

<sup>3</sup> On peut dire de façon synthétique que la probabilité est une fonction numérique non négative et additive, définie sur une tribu (collection de sous-ensembles), dont la masse totale est l'unité.

<sup>4</sup> Les sciences physiques et biologiques sont également concernées par ces controverses.

<sup>5</sup> Il est intéressant de voir que Jacques Bernoulli est considéré à la fois comme un objectiviste exemplaire [von Mises, 1928] et comme le père de la première conception subjectiviste de la probabilité [Hacking, 1975].

Il est donc possible de donner un statut objectif à la notion de probabilité, pour des événements susceptibles de se produire indéfiniment dans des conditions identiques. Les axiomes de Kolmogorov ont été conçus en fait pour servir de fondation mathématique à cette conception objectiviste des probabilités. Ils permettent cependant de définir, sans tenir compte de la façon dont les données sont recueillies, des probabilités comme des nombres associés à des ensembles et remplissant certaines règles [Shafer, 1990].

La conception fréquentiste, qui consiste à considérer la fréquence relative à laquelle un événement survient lorsque le nombre d'essais tend vers l'infini, existe depuis l'origine des probabilités. Ainsi Pascal en s'attaquant au «*problème des partis*» [1651] aborde une question que l'on peut considérer comme fréquentiste : comment distribuer le restant des mises lorsque la partie est soudainement interrompue. Les *Observations* de Graunt, qui sont consacrées à l'utilisation statistique des bulletins de mortalité, se placent parfaitement dans le point de vue fréquentiste. Cette approche a été reprise et développée par Venn [1866], pour prédominer pendant une grande partie du XX<sup>e</sup> siècle en particulier avec les travaux de Fisher [1925], de von Mises [1957], etc.

La conception propensioniste est apparue plus tard avec les travaux de Popper [1959, 1990]. Cette interprétation conserve l'idée que les probabilités sont des estimations portant sur des fréquences statistiques observées dans des suites longues (réelles ou virtuelles). Mais elle prend en considération le fait que ces suites sont définies par la façon dont leurs éléments sont engendrés : il en résulte que ces probabilités dépendent des conditions d'engendrement, et qu'elles peuvent changer lorsque ces conditions changent [Popper, 1990, p. 371-372]. Cette conception est particulièrement utile pour l'étude des processus stochastiques que l'on trouve dans la nature, tels que la radioactivité de certaines sources. Comme l'indique Popper, l'interprétation propensioniste est adéquate pour comprendre les théories physiques probabilistes, ainsi que les théories des jeux de hasard. Elle n'a guère été utilisée en sciences humaines.

Ce point de vue objectiviste va généralement refuser de considérer comme relevant du calcul des probabilités toute notion d'incertitude autre que celle qui joue sur l'arrivée de tels événements indéfiniment reproductibles et qui peut être vérifiée empiriquement. En particulier parler de la probabilité pour qu'une proposition donnée soit vraie n'a aucun sens pour un objectiviste, car on ne peut parler de la probabilité d'un événement par nature unique : la proposition est vraie ou fausse, en logique aristotélicienne. Popper [1990, p. 242] indique clairement que la probabilité d'une hypothèse en fonction des tests qu'elle a passés n'obéit pas aux règles du calcul des probabilités.

De ce point de vue, l'inférence statistique peut permettre de tester, à partir de l'observation d'un échantillon, la validité d'une hypothèse pour la population dont est extrait l'échantillon mais pas la probabilité qu'elle soit vérifiée. Cela est possible si l'on suppose que l'échantillon observé fait partie d'un ensemble beaucoup plus vaste d'échantillons tirés au hasard dans cette même population. C'est alors sur la probabilité d'obtenir l'échantillon observé, si l'hypothèse est vraie, que l'on va raisonner, et non sur la probabilité que l'hypothèse soit vraie, qui n'a aucun sens pour un objectiviste comme nous l'avons déjà indiqué.

## 1.2. UNE APPROCHE ÉPISTÉMIQUE<sup>6</sup> □ CONCEPTION SUBJECTIVISTE OU LOGICISTE

L'approche épistémique considère que l'on peut appliquer le calcul des probabilités pour évaluer les degrés auxquels il est raisonnable de croire en la vérité de propositions. Elle va donc prendre en compte le plus grand nombre possible de sentiments d'incertitude, aussi subjectifs soient-ils, mais surtout d'origine non-fréquentielle. Dans la mesure où les axiomes de Kolmogorov ne nécessitent pas le concept de répétition pour fonder mathématiquement les probabilités \_ car il n'intervient qu'à titre optionnel \_ ces axiomes peuvent aussi bien fonder les probabilités épistémiques.

La conception subjectiviste, d'une part, va chercher à montrer que si un individu est cohérent dans ses comportements face à l'incertain, alors les degrés de probabilité qu'il attribue aux hypothèses possibles, peuvent être représentés par des nombres réels qui vérifient les axiomes du calcul des probabilités. Les différents auteurs qui ont utilisé de tels concepts considèrent la probabilité comme « une mesure du degré de crédibilité d'un événement » [Granger, 1988], compte tenu de ce qu'ils savent des autres circonstances. La probabilité devient alors une mesure du degré d'intensité de l'attente d'un individu, variable de l'un à l'autre, et non plus la propriété d'un événement comme c'était le cas chez les fréquentistes. Mais, cependant, en introduisant des contraintes rationnelles de cohérence sur les comportements individuels, on peut retomber sur des plausibilités satisfaisant aux axiomes de Kolmogorov (voir [Allais, 1953] et [Matalon, 1967] pour une discussion de la validité de ces contraintes).

Cette notion de probabilité, comme les notions objectivistes, est présente dès le début du calcul des probabilités, par exemple dans le pari de Pascal, où il applique la théorie des jeux à la croyance en Dieu. Jacques Bernoulli, dans la partie IV de l'*Art de conjecturer* [1713] revendique explicitement une conception subjectiviste de la probabilité, en introduisant la notion de degré de certitude. Elle a été développée plus récemment par des mathématiciens et des statisticiens, tels que Ramsey [1931], de Finetti [1937], Savage [1954], etc. C'est à Savage que l'on doit la définition axiomatique la plus complète de ce point de vue.

Par rapport à cette conception subjectiviste, le point de vue logiciste ne se contente pas des contraintes de cohérence utilisées par les subjectivistes pour asseoir les probabilités. Ses représentants partagent bien la conception de la probabilité comme traduisant un degré de croyance, cependant ce degré ne représente pas un sentiment personnel, mais une relation logique valable pour tous. Ils proposent dès lors d'assigner des probabilités à l'aide d'une analyse logique de l'information incomplète rencontrée habituellement dans tous les problèmes d'analyse statistique. Cette nouvelle logique est différente d'une logique déductive □ un énoncé de probabilité logique n'est donc pas un énoncé formulé à partir d'un fait ou d'une observation, comme l'est un énoncé de probabilité objectiviste, mais un énoncé formulé à propos de la relation logique qu'entretiennent entre eux deux énoncés [Nadeau, 1999]. Elle permet également de justifier des formes de raisonnement plus faibles que le syllogisme, telles que : si P alors N □ mais N est vrai □ alors P devient plus plausible.

À nouveau cette notion de probabilité logique existe dès le début du calcul des probabilités □ Leibniz pensait déjà que le raisonnement probabiliste deviendrait une

---

<sup>6</sup> Le terme épistémique est pris ici dans le sens de « relatif à la connaissance ». Nous préférons parler ici d'approche épistémique plutôt que d'approche bayésienne, car cette dernière est souvent synonyme d'approche subjectiviste.

nouvelle forme de logique. Les travaux de Bayes [1763] et de Laplace sur la probabilité des causes tentent d'évaluer ce que l'on appelle aujourd'hui une probabilité inductive. Mais, après une période où l'approche objectiviste écrasa toute velléité subjectiviste ou logiciste [Hacking, 1990], il a fallu attendre les années 1920 pour que cette nouvelle logique prenne forme avec les travaux de Keynes [1921], de Jeffreys [1939], de Carnap [1950], et plus récemment de Jaynes [2003].

De ce point de vue, les énoncés de probabilité ne font plus référence au hasard ou à des variables aléatoires mais sont basés sur des principes logiques clairement définis et rendent la distinction entre théorie des probabilités et inférence statistique inutile. Ces principes logiques sont en effet des règles uniques et consistantes pour mener des inférences de tout type, lorsque toute l'information nécessaire pour mener un raisonnement n'est pas disponible. Bien entendu, lorsqu'elle l'est, le raisonnement déductif reprend sa place comme forme limite du raisonnement plausible.

Sous cette approche épistémique, l'inférence statistique, considérée comme la révision d'une probabilité à la lumière d'informations nouvelles, peut alors utiliser le théorème de Bayes [1763]. En effet, le problème abordé dans ce cas est celui de déterminer comment la probabilité d'une hypothèse *a priori* peut être modifiée par l'observation d'une expérience postérieure. On voit facilement pourquoi les fréquentistes sont amenés à rejeter cette méthode d'inférence statistique, car ils se refusent à parler de la probabilité d'une hypothèse qui est un événement unique. En revanche pour les épistémistes la notion de probabilité d'une hypothèse constitue la notion centrale du calcul des probabilités tel qu'ils le conçoivent et le théorème de Bayes leur permet d'interpréter l'inférence comme un processus de décision [Rouanet *et al.*, 1998].

## 2. DEMOGRAPHIE, SCIENCES SOCIALES ET POINT DE VUE PROBABILISTE

L'article précédemment publié dans cette revue [Courgeau, 2002] avait montré une évolution de la pensée démographique, qui se retrouvait, au moins en partie, dans d'autres sciences sociales, comme l'épidémiologie, la géographie humaine, les sciences de l'éducation, etc. [Courgeau, 2003]. Partant d'une approche du moment [Durkheim, 1895], passant par une approche longitudinale agrégée, toutes deux étant des holismes, puis par une approche biographique individualiste, la démographie aboutit à une approche contextuelle et multiniveau, qui permet de dépasser l'opposition entre holisme et individualisme méthodologique. Nous allons d'abord montrer comment le point de vue objectiviste a longtemps prévalu dans la plupart des sciences sociales, puis montrer ce qu'une approche bayésienne subjectiviste ou logiciste peut apporter dans ce domaine.

### 2.1 L'APPROCHE OBJECTIVISTE EN SCIENCES SOCIALES

Nous allons d'abord voir comment et pourquoi la démographie et les autres sciences sociales ont essentiellement utilisé une approche objectiviste tout au long de leur histoire. Nous verrons ensuite les critiques que l'on peut apporter à cette approche en les illustrant à l'aide d'un exemple.

#### 2.1.1 L'histoire de cette approche en démographie et en sciences sociales

Comme nous l'avons déjà indiqué, dès le départ Graunt [1662] se situe dans l'approche objectiviste lorsqu'il utilise les bulletins de mortalité pour en tirer, par exemple, des

fréquences de décès selon la cause. En fait les fondateurs des sciences sociales au XVII<sup>e</sup> et au XVIII<sup>e</sup> siècles recherchent des lois simples et universelles pour expliquer le monde des phénomènes aussi bien physiques qu'humains. Ainsi, le mémoire de Halley utilise les données de la ville de Breslau pour fournir une estimation de la mortalité du genre humain [1693] □ Kersseboom [1742] part de l'observation des décès survenus d'année en année dans les négoce pour construire une table de survie, dont il dit □ «J'ai tant de preuves de la précision de cette table que j'ose la présenter comme généralement valable □. Les conditions d'une approche fréquentiste sont parfaitement vérifiées, dès lors que l'on cherche à estimer la fréquence d'un phénomène démographique, qui est posé comme universel.

Cependant, très rapidement, l'observation de nombreuses populations montra que ces probabilités étaient loin d'être universelles et qu'elles vont dépendre de caractéristiques individuelles telles que le sexe et l'âge, de caractéristiques plus générales telles que la résidence en ville ou en campagne, dans des climats nordiques ou des pays plus méridionaux, etc., et qu'elles vont varier au cours du temps. Ainsi Laplace pouvait écrire en 1812 □ □ Tant de causes variables influent sur la mortalité que les tables qui la représentent doivent changer suivant les lieux et les temps □. Cependant cela ne change en rien l'approche objectiviste suivie □ il suffit de considérer maintenant des populations de définition plus restreinte que l'ensemble de l'humanité, pour y estimer des probabilités de même type que précédemment. En fait jusqu'à la fin du XVIII<sup>e</sup> siècle, l'essentiel des fréquences statistiques que l'on pouvait estimer concernait les phénomènes démographiques tels que la mortalité, la fécondité, la nuptialité.

Lors du XIX<sup>e</sup> siècle, la mise en place dans de nombreux pays de recensements non seulement de la population mais également de nombreux faits sociaux, tels que la profession, la maladie, le suicide, le crime, la folie, etc., vient renforcer l'approche objectiviste dans de nombreuses sciences sociales. Ainsi Durkheim [1895] indique que «La statistique nous fournit le moyen de les isoler. Ils sont, en effet, figurés, non sans exactitude, par le taux de natalité, de la nuptialité, des suicides, ... □. Il continue en indiquant que «Chacun de ces chiffres comprend tous les cas particuliers indistinctement, les circonstances individuelles qui peuvent avoir quelque part dans la production du phénomène s'y neutralisent mutuellement et, par suite, ne contribuent pas à le déterminer. Ce qu'il exprime, c'est un certain état de l'âme collective □.

Il propose dès lors une analyse, qui ne va pas considérer les circonstances sous lesquelles un individu a connu un fait donné mais va considérer «Les faits sociaux comme des choses □, et cherche ainsi à unifier l'analyse tant sociologique que démographique. Il précise □ que l'individu trouve les faits sociaux comme des choses «Toutes formées et il ne peut pas faire qu'elles ne soient pas ou qu'elles soient autrement qu'elles sont □. Ces faits sociaux, dans un type de société donné, ont également une grande permanence dans le temps, de même que les «Institutions sociales nous sont léguées toutes faites par les générations antérieures □. Ces faits sociaux, qui existent hors de la conscience individuelle et qui sont indépendants des diverses actions qu'ils déterminent, s'expriment dans les statistiques, qui fournissent le moyen de les isoler. La méthode des variations concomitantes, qu'il propose pour ce faire et qui n'est autre qu'une méthode de régression linéaire, utilise l'hypothèse d'une probabilité objective □ l'existence d'une telle probabilité, différente pour le suicide des protestants et celui des catholiques [Durkheim, 1897], lui permet de justifier l'utilisation de la méthode des variations concomitantes pour montrer que les protestants se suicident moins que les catholiques.

Plus tard, lorsque Henry [1959] recherche la probabilité d'arrivée d'un événement en fonction de l'âge d'un individu d'une cohorte donnée, il fait l'hypothèse d'une cohorte homogène, où les probabilités de connaître cet événement «sont les mêmes à chaque âge, pour tous les membres de la cohorte». C'est donc bien encore le concept de probabilité objective qu'il utilise. Même lorsqu'il considère ensuite des cohortes hétérogènes, il raisonne «comme si chaque cohorte hétérogène était formée par la réunion de cohortes homogènes d'effectif infini».

De façon semblable l'analyse des biographies en sciences sociales s'est essentiellement développée dans le cadre objectiviste. L'article fondateur de David Cox [1972] se place dans ce cadre et suppose que les observations sont disponibles sur une sous-population d'individus indépendants, étant sous-entendu que cette sous-population est issue d'une population plus large de mêmes caractéristiques. Des sociologues [Tuma et Hannan, 1984; Blossfeld *et al.*, 1986], des démographes [Courgeau et Lelièvre, 1989], des économistes [Lancaster, 1990], etc., envisagent tous cette analyse de façon objectiviste. Aucun de ces ouvrages ne comporte même le terme «Bayésien» dans son index.

Enfin l'analyse multiniveau s'est également développée au départ dans le cadre objectiviste. De ce point de vue, on peut considérer les différents niveaux comme un échantillon aléatoire de contextes dont les effets peuvent être considérés comme distribués de façon approximativement normale. Ainsi lorsque l'on dispose en sciences de l'éducation, par exemple, de notes d'élèves à deux âges, répartis dans différentes classes [Goldstein, 2003], on peut considérer les classes observées dans l'échantillon comme un échantillon aléatoire d'une population beaucoup plus vaste de classes, de la même façon que les élèves sont issus d'une population également plus vaste d'élèves. De la même façon que l'échantillon d'élèves peut fournir des estimations des moyennes et variances de la population qu'ils représentent, l'échantillon de classes pourra fournir des informations sur les caractéristiques de la population formée par les classes. Ainsi, il est possible de fournir des estimations de la variance et de la covariance entre écoles des paramètres estimés, lorsque l'on essaye par exemple d'expliquer les notes à un âge donné en fonction des notes antérieures et de diverses caractéristiques individuelles (modèle multiniveau), et de comparer des écoles avec différentes caractéristiques agrégées correspondant à chaque classe (modèle contextuel).

### 2.1.2 L'inférence statistique objectiviste

Cependant, le raisonnement par lequel on arrive à une inférence statistique fréquentiste, à partir de données observées, n'est pas aussi clair qu'il pourrait le paraître à première vue, et il nous faut examiner plus en détail son déroulement. Ainsi, l'énoncé, selon lequel l'intervalle de confiance à 95 % pour un paramètre inconnu,  $\theta$  (par exemple l'âge moyen à la naissance du premier enfant, dans la génération née en 1920 en France, estimé à partir d'un échantillon représentatif de cette génération), est compris entre deux valeurs  $\theta_1$  et  $\theta_2$ , semble indiquer que le paramètre se trouve dans cet intervalle avec une probabilité de 95 %, mais cela n'est pas vérifié, car cet intervalle ne peut être défini que par rapport à l'estimation de ce paramètre et non par rapport à ce paramètre lui-même, qui est inconnu. Dans ce cas on peut seulement affirmer que, si l'on effectue de nombreux tirages d'échantillons de même taille et si nous construisons un tel intervalle autour de la moyenne de chaque échantillon, alors on peut s'attendre à ce que 95 % des intervalles de confiance ainsi construits contiennent le paramètre inconnu. Il s'agit là d'une réponse à une question beaucoup plus complexe, que la première, qui paraissait beaucoup plus claire : quelle est la probabilité pour que le paramètre inconnu se trouve

dans un intervalle donné. La réponse à cette question n'existe pas, en fait, dans la théorie fréquentiste.

Allons plus loin, en essayant de voir plus précisément ce que l'inférence statistique objectiviste peut montrer. Supposons que l'on cherche à voir si un certain facteur affecte ou non le phénomène que l'on étudie – ainsi dans l'exemple que nous détaillerons plus loin, il s'agit de voir si le fait d'être agriculteur affecte la probabilité de migrer d'une certaine sous-population donnée. On se demande alors si les différences entre les probabilités estimées de migrer des agriculteurs et du reste de la population peuvent s'expliquer par le hasard seul ou si elles sont significativement différentes. Si ces probabilités se révèlent différentes à un seuil fixé d'avance, par exemple un «niveau de signification» à un pour cent, alors on pourra conclure qu'elles le sont, puisque le résultat observé n'a que peu de chances d'avoir été obtenu au hasard. On interprète donc ici ces probabilités en termes fréquentiels, en imaginant une population plus vaste que celle observée de laquelle on peut tirer au hasard un grand nombre de sous-populations, dont celle observée. Si la probabilité de l'échantillon observé est trop faible, on rejettera alors l'hypothèse testée. Cette constatation rejoint ce que nous avons dit précédemment – l'inférence statistique objectiviste permet de tester la probabilité d'obtenir l'échantillon observé, si l'hypothèse est vraie, mais non la probabilité de cette hypothèse elle-même, qui est soit vraie soit fausse [Matalon, 1967].

Cela rejoint également ce que dit Fisher [1956], qui exprime les idées d'un objectiviste<sup>7</sup> sur ces questions

*“This fundamental requirement [of no recognisable subset] for the applicability to individual cases of the concepts of classical probability shows clearly the role of subjective ignorance, as well as that of objective knowledge in a typical probability statement. It has been often recognised that any probability statement, being a rigorous statement involving uncertainty, has less factual content than an assertion of a certain fact would have, and at the same time has more factual content than a statement of complete ignorance. The knowledge required for such a statement refers to a well-defined aggregate, or population of possibilities within which the limiting frequency ratio must be exactly known. The necessary ignorance is specified by our inability to discriminate any of the different subaggregates having different limiting frequency ratios, such as must always exist”*  
(p. 33).

<sup>7</sup> Si dans une majorité de ses écrits Fisher peut être considéré comme un fréquentiste, il donne à la fin de sa vie une définition différente de la probabilité, en ce qu'il indique que «l'on ne peut reconnaître aucun sous-ensemble dans la fraction possédant la caractéristique, qui l'ait dans une proportion différente de l'ensemble» (“no sub-set may be recognizable having a fraction possessing the characteristic differing from the fraction  $P$  of the whole”, [Fisher, 1960]). Ce concept de probabilité est généralement considéré comme peu clair et n'a guère été utilisé par la suite [Savage, 1976].

<sup>8</sup> «Cette nécessité fondamentale [le fait qu'on ne peut pas identifier de sous-ensemble] pour pouvoir appliquer à des cas individuels le concept de probabilité classique montre clairement le rôle de l'ignorance subjective, aussi bien que celui de la connaissance objective dans un énoncé probabiliste typique. Il a souvent été reconnu que tout énoncé probabiliste, étant un énoncé rigoureux impliquant l'incertitude, a un contenu moins relié aux faits qu'un énoncé portant sur un fait certain, et en même temps a un contenu factuel plus certain qu'un énoncé de parfaite ignorance. La connaissance exigée par un tel énoncé se rapporte à une sous-population bien-définie, ou à une population de possibles à l'intérieur de laquelle la fréquence limite doit être parfaitement connue, lorsque le nombre d'essais tend vers l'infini. L'ignorance inéluctable est liée à notre incapacité de distinguer aucun des sous-groupes différents qui ont différentes fréquences limites, tels qu'ils devraient toujours exister».

Cette citation généralise la difficulté de définir un intervalle dans lequel se trouve un paramètre inconnu, en indiquant l'incapacité pour un fréquentiste de discerner avec une clarté suffisante les différentes sous-populations ayant différentes fréquences limites.

Ainsi, lorsque nous travaillons sur un échantillon issu d'une population plus vaste, les résultats d'une analyse statistique de ces données permettent, sous certaines hypothèses, une inférence sur le comportement étudié d'un membre de cette population, qui se trouve hors de l'échantillon, mais a certaines caractéristiques observées. Pour ce faire, il faut supposer que ce membre est semblable aux individus de l'échantillon ayant ces caractéristiques et qu'il fait partie du même sous-groupe, qui ne peut dès lors être divisé plus finement.

La situation est encore plus complexe lorsque l'on travaille non plus sur un échantillon mais sur l'ensemble de la population. On doit alors supposer que cette population observée est un échantillon issu d'une « super-population », dans laquelle on n'a pu faire qu'un seul tirage, sur lequel on doit donc raisonner. C'est bien ce qui est fait lorsque l'on cherche à projeter cette population vers l'avenir en faisant l'hypothèse que son comportement reste identique à celui que l'on avait observé. Les méthodes démographiques de projection de population ou de micro-simulation utilisent cette hypothèse [Van Imhoff et Post, 1997, 1998].

Seul un postulat d'ignorance permet dans ces cas une inférence statistique, en disant que certaines choses sont inconnues et que la validité de l'argument implique qu'elles n'aient pas à l'être [Fisher, 1958]. Voyons plus en détail sur un exemple la difficulté de définir une sous-population dans laquelle une probabilité reste constante.

### 2.1.3 *Un exemple en démographie*

Prenons pour ce faire l'exemple développé dans l'article précédemment publié dans cette revue [Courgeau, 2002]. Nous y étudions l'effet d'être agriculteur sur la probabilité d'émigrer des diverses régions du pays à un âge donné, d'un individu né en Norvège en 1948<sup>9</sup>. Voyons, selon le type d'analyse faite, comment nous pouvons définir les sous-groupes ayant la même fréquence limite de migrer.

Supposons d'abord que la probabilité de migrer des agriculteurs soit identique pour tous les individus de la population totale, et qu'il en soit de même pour les non-agriculteurs. Sous cette hypothèse, on a deux possibilités d'estimer ces probabilités. En premier lieu on peut utiliser la méthode proposée par Durkheim, ce qui amène à travailler sur les proportions de migrants et d'agriculteurs dans chaque région norvégienne, et qui conduit à une probabilité de migrer égale à 0,597 pour les agriculteurs et à 0,119 pour les non-agriculteurs. En second lieu on peut travailler sur l'ensemble de la population norvégienne, qui conduit à une probabilité de migrer égale à 0,094 pour les agriculteurs et à 0,150 pour les autres. Le fait que ces résultats soient contradictoires nous indique que l'hypothèse initiale n'est pas vérifiée. Si la première possibilité d'estimation implique forcément cette hypothèse et doit donc être rejetée, la seconde peut toujours être vérifiée si l'on estime la valeur moyenne d'une probabilité de migrer, qui peut cette fois-ci varier selon les individus considérés. Comment dès lors

---

<sup>9</sup> Nous remercions ici les services statistiques norvégiens qui nous ont permis d'avoir accès à des fichiers créés à partir des données du Registre de Population et des recensements de 1970 et 1980 par Kjetil Sørliet et Øjsten Kravdal.

estimer cette hétérogénéité non observée avec les données dont nous disposons. Deux nouvelles possibilités se présentent à nous.

En utilisant d'abord la proportion d'agriculteurs présents dans chaque région, on peut estimer un modèle contextuel faisant intervenir non seulement le fait d'être agriculteur, mais également cette proportion et une interaction entre le fait d'être agriculteur et cette proportion. Le modèle ainsi estimé nous montre que, si pour les agriculteurs leur taux d'émigration ne dépend pas de cette proportion de façon significative, pour les non-agriculteurs la probabilité de migrer, toujours plus forte que pour les agriculteurs, augmente fortement avec cette même proportion. Ce modèle permet dès lors d'harmoniser les contradictions précédentes : une forte densité d'agriculteurs dans une zone va augmenter les chances de migrer des autres professions, sans affecter celle des agriculteurs. Les sous-populations, dont la probabilité de migrer sera identique pour tous leurs membres, sont dans ce cas celles des agriculteurs ou des non-agriculteurs, présents dans une ou plusieurs régions dont le pourcentage d'agriculteurs est constant.

Une autre approche consiste à travailler en multiniveau, avec en plus du niveau individuel, un niveau régional. Un tel modèle, sans faire intervenir en premier lieu, les pourcentages de migrants dans chaque région, nous permet d'estimer des probabilités de migrer qui dépendent à la fois du fait d'être agriculteur et du fait d'habiter une région donnée. Un tel modèle donne un aléa tout à fait significatif pour les non-agriculteurs. Mais lorsque l'on fait intervenir dans ce même modèle les pourcentages d'agriculteurs, au niveau des paramètres fixes, cet aléa se réduit de moitié et devient beaucoup moins significatif. En revanche l'aléa pour les agriculteurs reste au même niveau que dans le modèle précédent, et n'est pas significativement différent de zéro (voir le modèle objectiviste présenté dans le Tableau 1, p. 43). Cela montre à nouveau clairement que l'introduction des pourcentages d'agriculteurs vient expliquer une part importante de la variance au niveau régional correspondant aux non-agriculteurs. Les sous-populations, dont la probabilité de migrer peut être considérée comme identique, seront dans ce cas celles des agriculteurs ou des non-agriculteurs présents dans chaque région.

Comme on peut le voir, on teste chaque fois la probabilité d'obtenir l'échantillon ici observé si une hypothèse donnée est vraie. Cette hypothèse est d'abord celle d'une probabilité de migrer des agriculteurs identique dans toutes les régions de même que celle des non-agriculteurs. L'échantillon observé ne vient pas soutenir cette hypothèse. Notons ici que cet échantillon bien que constitué par l'ensemble des individus observés (données exhaustives du registre de population), est considéré ici comme un des tirages possibles d'une population hypothétique beaucoup plus vaste, que les objectivistes appellent souvent une « super population ». L'hypothèse devient ensuite celle d'une probabilité de migrer des agriculteurs et des non-agriculteurs dépendant de façon différente du pourcentage d'agriculteurs présents dans chaque région. L'échantillon observé soutient alors cette hypothèse.

## 2.2 L'APPROCHE ÉPISTÉMIQUE EN SCIENCES SOCIALES

Si l'approche épistémique est apparue simultanément à l'approche objectiviste dans l'histoire des probabilités, elle ne s'est développée que beaucoup plus récemment, en particulier en sciences humaines. Nous présenterons d'abord l'histoire de cette approche et les raisons qui ont retardé son application aux sciences sociales, en particulier à la démographie. Nous verrons ensuite qu'elle conduit à une inférence statistique dont la

signification nous paraît plus claire que lorsqu'on utilise une approche objectiviste des probabilités.

### 2.2.1 L'histoire de cette approche en sciences humaines et sociales

Cette approche épistémique concerne l'évaluation des degrés auxquels il est raisonnable de croire en la vérité de propositions non assurées. Ces propositions sont de type beaucoup plus général que celles portant sur l'occurrence d'événements susceptibles de se répéter dans des conditions identiques, comme pour les probabilités objectivistes.

Comme nous l'avons déjà indiqué, cette approche s'est développée, simultanément à l'approche objectiviste, tout au long du XVII<sup>e</sup> et XVIII<sup>e</sup> siècles. Le pari de Pascal, qui se trouve dans un passage intitulé *Infini-rien* des *Pensées* [1670], montre comment des considérations sur l'aléatoire peuvent conduire à une prise de décision dans le domaine *théologique*. Il n'est pas question ici de donner une vue détaillée des raisonnements suivis par Pascal [Hacking, 1975], mais d'en donner les principes. Sa démarche est la suivante. Considérons un individu qui hésite entre la foi et l'incrédulité, mais ne veut pas se fier aux témoignages des croyants, aux docteurs de l'Église ni aux miracles. En l'absence de données expérimentales, il faut parier sur l'existence de Dieu. Prenons le gain et la perte, en prenant croix que Dieu est. Estimons ces deux cas : si vous gagnez, vous gagnez tout ; si vous perdez, vous ne perdez rien. Gagez donc qu'il est, sans hésiter. C'est donc bien sur une hypothèse, celle de l'existence de Dieu, que porte le raisonnement.

Leibniz utilise également cette approche lorsqu'il s'intéresse aux degrés de preuve dans le *domaine juridique*. C'est également sur l'hypothèse de l'existence d'un fait extraordinaire, comme un miracle, que Hume [1748] s'interroge, faisant intervenir d'une part la fiabilité d'un témoin oculaire, d'autre part la crédibilité du phénomène indépendamment de ce témoignage.

Il nous faut présenter rapidement ici l'article de Bayes [1763], publié après sa mort, qui pose de façon plus précise le problème des probabilités épistémiques. Dans les problèmes de probabilité classiques, on spécifie à l'avance la nature d'une expérience ou d'un jeu, pour poser ensuite une question relative à un ou plusieurs résultats attendus de cette expérience ou de ce jeu. Ici, on va se poser la question inverse : on dispose des résultats d'une expérience ou d'un jeu sans savoir leur nature, et l'on se demande quelle est la probabilité que cette expérience ou que ce jeu soit d'un type donné parmi de nombreux autres [Molina, 1931]. Bayes pose alors le problème de la façon suivante : ayant observé le nombre de fois où un événement inconnu se produit ou non, quelle est la probabilité (chance) que sa probabilité d'occurrence dans une expérience unique se trouve entre deux valeurs données ? En s'exprimant avec les termes et les notations actuels, Bayes montre, en prenant d'abord l'exemple d'une table de billard, que l'on peut calculer la probabilité pour qu'elle se situe entre deux valeurs  $a$  et  $b$  :

$$P(a < \theta < b \mid X = p) = \frac{\int_a^b \binom{n}{p} \theta^p (1-\theta)^{n-p} d\theta}{\int_0^1 \binom{n}{p} \theta^p (1-\theta)^{n-p} d\theta} = \frac{\int_a^b \theta^p (1-\theta)^{n-p} d\theta}{\int_0^1 \theta^p (1-\theta)^{n-p} d\theta}$$

où  $p$  est la proportion du nombre de fois où l'événement  $X$  se produit rapporté au nombre d'observations antérieures et  $\theta$  la probabilité inconnue recherchée. C'est donc bien la probabilité d'une hypothèse qu'il cherche à calculer.

Il généralise ensuite cet exemple, dans un *Scholium*, au cas où  $X$  serait un événement dont on ignore tout, avant l'expérience, qui doit donc être uniformément distribué et dont la probabilité inconnue est posée égale à  $\theta$ , pour aboutir à une formule identique à la précédente. Diverses interprétations ont été données de cette distribution initiale uniforme. Molina [1931] a montré qu'une probabilité initiale (prior) uniforme sur toutes les valeurs possibles de  $\theta$  est une condition nécessaire pour une probabilité uniforme sur les résultats observés. Stigler [1982] amoindrit l'argument utilisé par la suite, selon lequel la distribution uniforme s'applique à la distribution inobservable de  $\theta$ , pour dire qu'elle s'applique en fait à la distribution observable de  $X$ . Dans ce cas ce n'est que lorsque la distribution de  $X$ ,  $p(x)$ , permet de déterminer la distribution de  $\theta$ ,  $p(\theta)$ , que l'inférence bayésienne est possible. Ainsi dans le cas binomial où l'on a une séquence échangeable<sup>10</sup> de 0 et de 1, alors, si  $p(x_1, \dots, x_n)$  est spécifié pour tout  $n$ , le théorème de de Finetti [1974] montre que cela est une condition suffisante pour que  $p(\cdot | \theta)$  et  $p(\theta)$  soient déterminés.

Le problème épistémique est repris, en particulier par Condorcet dans la cinquième partie de son *Mémoire sur le calcul des probabilités* [1786, 1994], intitulée *Sur la probabilité des faits extraordinaires*. Si  $u$  et  $e$  représentent les probabilité de la vérité et de la fausseté du fait et si  $u'$  et  $e'$  représentent la probabilité qu'un témoignage sera ou non conforme à la vérité alors que le témoin aura assuré la vérité de cet événement, alors sa probabilité conditionnelle au fait que le témoin le confirme est  $\square$

$$\frac{uu'}{uu' + ee'}$$

Si l'on suppose, comme Condorcet, que  $e = \frac{999999}{1000000}$ ,  $u = \frac{1}{1000000}$ ,  $u' = \frac{999}{1000}$ ,  $e' = \frac{1}{1000}$  nous aurons alors  $\frac{uu'}{uu' + ee'} = \frac{999}{1000998}$ . On voit alors que le fait que le témoin confirme un événement très extraordinaire dont la probabilité est égale à un millionième, donne toujours une très faible probabilité conditionnelle de cet événement au fait que le témoin le confirme, moindre que le millième. Ce raisonnement est conforme à la règle de Bayes-Laplace pour l'évidence d'un témoignage [Sobel, 1987].

Enfin, comme bien d'autres auteurs de son temps, Laplace [1812] développe les applications du calcul des probabilités aux sciences morales, pour l'étude de la probabilité des témoignages, des choix et des décisions des assemblées, des élections, des jugements des tribunaux, de l'effet de l'inoculation de la vaccine sur la mortalité par petite vérole, etc.

Ces raisonnements très fréquents tout au long du XVIII<sup>e</sup> siècle vont disparaître au XIX<sup>e</sup> et au début du XX<sup>e</sup> siècles, alors qu'en sciences humaines les recensements de population prennent de l'ampleur et que l'approche fréquentiste devient prépondérante. Ce n'est que vers le début des années vingt que diverses sciences sociales vont reprendre cette approche épistémique.

<sup>10</sup> Voir le paragraphe suivant pour la définition plus précise du terme «échangeable».

L'économiste Keynes [1921] introduit les probabilités dans une conception logiciste. Il y indique clairement : «In the sense important to logic, probability is not subjective. A proposition is not probable because we think it so. When once the facts are given which determine our knowledge, what is probable or improbable in those circumstances has been fixed objectively, and is independent of our opinion»<sup>11</sup>. Il est intéressant de voir que, si sa position logiciste a pu être développée de façon plus précise au cours de la fin du XX<sup>e</sup> siècle [Jeffreys, 1939□Carnap, 1950□Jaynes, 2003], la position subjectiviste est apparue très peu de temps après. En effet en 1926 Ramsey publie un article, repris dans l'ouvrage publié après sa mort en 1931, qui critiquait la position de Keynes et proposait une théorie basée sur les comportements rationnels face à l'incertitude. Les probabilités subjectives vont mesurer des degrés de certitude pour définir les probabilités. Cette approche a été prolongée par les travaux non seulement d'économistes mais de statisticiens [de Finetti, 1937□Savage, 1954□etc.].

L'approche épistémique va par la suite se développer dans diverses sciences humaines. Mais plutôt que de les prendre séparément, ce sont des nouvelles analyses, applicables à de nombreuses sciences, qui vont utiliser ces méthodes. Ainsi l'analyse des biographies, utilisée en épidémiologie, en sciences actuarielles, en biologie, en médecine, en économie, en démographie, etc., va développer une approche bayésienne [Ibrahim *et al.*, 2001□Gustafson *et al.*, 2003]. Les difficultés de l'analyse biographique bayésienne, à la fois pour les concepts et les calculs, viennent des difficultés de modéliser les quotients cumulés avec des distributions *a priori* non-paramétriques, par exemple, des processus beta ou gamma, dont la théorie et le calcul sont très souvent redoutables pour les chercheurs. Pour éviter ces difficultés des méthodes pleinement bayésiennes, on tend de plus en plus à utiliser des méthodes de simulation (*bootstrap* bayésien, par exemple), que la puissance des ordinateurs actuels permet de traiter en des temps non-prohibitifs [Kim et Lee, 2003]. On peut maintenant dire que cette approche présente de nombreux avantages par rapport à une approche plus objectiviste□ la signification des intervalles de confiance y est plus claire que dans le cas objectiviste, il y a la possibilité d'introduire simplement des informations antérieures sur les estimations à réaliser, etc.

De même, l'analyse multiniveau conduit à des modèles hiérarchiques bayésiens [Draper, 1995] plus parfaitement adaptés à ces travaux. À nouveau les difficultés de calcul ont pu être surmontées par l'utilisation de simulations telles que les méthodes de Monte Carlo par chaînes de Markov (*MCMC*) ou par la méthode du *bootstrap*. On peut en particulier voir le chapitre sur les *Modèles hiérarchiques* dans l'ouvrage de Lee [1989], le chapitre sur les *Modèles multiniveau, l'analyse multivariée et les modèles longitudinaux* de l'ouvrage de Congdon [2001] et les chapitres sur l'estimation bayésienne de divers types de modèles multiniveau dans la dernière version de l'ouvrage de Goldstein [2003]. Cette analyse se trouve en fait au cœur des modèles bayésiens lorsque l'on doit faire des hypothèses sur la structure du phénomène étudié, qui ne dépend pas seulement des caractéristiques individuelles mais également d'autres caractéristiques contextuelles telles que la présence d'un professeur pour les élèves d'une classe, la qualité des soins pour les patients traités dans un hôpital, etc. Voyons un peu plus en détail les cas où l'approche objectiviste n'est plus utilisable en multiniveau, et où une approche épistémique permet de résoudre ce problème.

---

<sup>11</sup> «□Dans toute l'acception logique du terme, la probabilité n'est pas subjective. Une proposition n'est pas probable parce que nous pensons qu'elle l'est. Une fois que les faits qui déterminent nos connaissances sont établis, ce qui est probable ou improbable dans ces conditions a été fixé objectivement, et est indépendant de notre opinion□.

Tant que les niveaux d'agrégation introduits (classes, écoles, communes, etc.), peuvent être considérés comme tirés d'une population sinon infinie du moins suffisamment vaste, l'approche objectiviste peut être appliquée dans de tels modèles multiniveau. En revanche supposons que l'on travaille sur un découpage exhaustif du territoire national, en régions, comme en géographie humaine, et que l'on dispose d'un échantillon de la population tiré dans chaque région pour déterminer, par exemple, si un individu est fumeur habituel ou non, en fonction de caractéristiques individuelles et régionales [Jones, 1993]. Dans ce cas, il est difficile de considérer que ces régions sont issues d'un échantillon plus vaste de régions puisqu'elles recouvrent tout le territoire. On peut bien sûr ajouter des régions d'autres pays pour élargir l'échantillon, mais la condition de normalité risque fort de ne plus être vérifiée et l'on a de fortes chances d'avoir une distribution qui aura, par exemple, deux ou plusieurs maxima locaux. Il est dès lors utile de voir si un point de vue subjectiviste ne permet pas de résoudre cette difficulté.

On montre dans ce cas que l'utilisation d'un estimateur de Bayes empirique (empirical Bayes estimator), par exemple, est possible sans avoir à faire intervenir un nombre infini de régions. En partant d'une estimation *a priori* (prior estimation) faite sur l'ensemble du pays et en appliquant le théorème de Bayes aux observations régionales, on obtient une telle estimation *a posteriori* du fait que l'individu soit un fumeur habituel. On montre que cet estimateur est bien plus précis qu'un estimateur conventionnel [Greenland, 2000]. Cependant l'application de cette méthode n'est valable que si une hypothèse d'échangeabilité des estimateurs *a priori* peut être considérée comme applicable dans ce cas. Il nous faut donc voir plus en détail ce que cette hypothèse implique, pour pouvoir ensuite justifier cette approche.

### 2.2.2 L'inférence statistique épistémique

Ce concept d'échangeabilité, introduit par de Finetti [1974], va jouer un rôle comparable à celui de l'appartenance à une sous-population, que nous avons présenté plus haut et qui a été introduit par Fisher [1974].

La notion d'échangeabilité, prise ici dans son sens subjectif, implique des jugements de similarité des comportements entre individus, qui permettent de faire des prévisions basées sur ce qui a été observé. Ainsi on peut regarder deux quantités inconnues comme échangeables si, pour tout jugement impliquant l'une ou les deux simultanément, les affectations de probabilités (probability assignments) ne changent pas lorsque les deux quantités sont échangées. Ainsi, dans l'exemple traité plus haut, on va supposer qu'avant d'observer les données régionales, la seule observation des données nationales ne nous donne aucune possibilité de dire si les données régionales sont ou non-différentes entre elles. Il est donc alors légitime de considérer ces données régionales comme échangeables entre elles. En revanche les estimateurs *a posteriori* peuvent, bien entendu, ne plus être échangeables, si des différences importantes apparaissent entre les régions. On peut montrer que ces estimateurs auront de toute façon une bien meilleure précision que les estimateurs conventionnels, même si les hypothèses d'échangeabilité des estimateurs *a priori* sont très incorrectes [Greenland, 2000].

Cette notion peut se généraliser à celle d'échangeabilité conditionnelle, si la distribution jointe de  $n$  variables aléatoires,  $X$ , conditionnée par la valeur d'une autre variable aléatoire,  $Y$ , qui peut être multidimensionnelle, est invariante quelle que soit la permutation des unités [Lindley and Novik, 1981 □ Draper *et al.*, 1993; Greenland,

1998a]. Ainsi dans l'exemple précédent, si l'on observe deux individus, dont on ne sait pas s'ils sont fumeurs habituels ou non, mais dont les caractéristiques considérées comme jouant sur ce fait sont identiques, on peut considérer que leur statut de fumeur habituel sera échangeable.

Cela permet également une possibilité d'inférence statistique à partir de cet échantillon. Ainsi, lorsque la variable aléatoire  $X$  dépend d'un certain nombre de caractéristiques  $Y$ , pour  $n$  individus, un nouvel individu ayant toutes ses caractéristiques  $Y$  identiques à celles des précédents, pourra être également considéré comme échangeable en  $X$ . Par exemple, pour un individu dont on ignore s'il est fumeur habituel, mais dont toutes les caractéristiques, connues pour jouer sur ce fait (âge, sexe, profession, statut matrimonial, etc., par exemple) sont identiques à celles de  $n$  individus, jugés échangeables quant au fait qu'ils fument ou non, ce nouvel individu aura également son statut de fumeur échangeable avec celui des précédents.

On peut voir que cette notion d'échangeabilité conditionnelle d'individus peut être rapprochée de celle de leur appartenance à une sous-population, due à Fisher, où la discrimination indiquée par ce dernier se rapporte à la condition qu'ils fassent partie de la sous-population dont les caractéristiques  $Y$  sont identiques. Cependant il y a deux différences importantes entre ces concepts [Lindley and Novik, 1981]. En premier lieu, l'échangeabilité d'individus se réfère explicitement à une variable aléatoire, alors qu'une sous-population ne peut pas s'y référer. Ainsi les individus peuvent être échangeables en  $Y$  étant donné  $X$ , mais pas en  $X$  : si nous tirons un nouvel individu vivant dans un autre pays on peut, par exemple, penser que la propension à fumer diffère de façon importante selon les pays, mais que les caractéristiques considérées jouent de la même façon sur cette propension, c'est-à-dire que  $P(Y|X)$  et  $P(Y|\bar{X})$  sont les mêmes dans les deux pays considérés. Mais en revanche la propension à être un fumeur ayant les caractéristiques  $Y$  ne pourra plus être estimée, car le théorème de

Bayes nous donne :  $P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$ , et dans ce cas, sans jugement

d'échangeabilité en  $X$ , il n'est plus possible d'obtenir la probabilité d'être un fumeur dans le nouveau pays considéré si l'on a certaines caractéristiques  $Y$ . En second lieu, nous avons indiqué la difficulté, sinon l'impossibilité, de placer un individu réel dans une sous-population bien définie du point de vue objectiviste. En revanche, l'échangeabilité étant un jugement sur une caractéristique individuelle, il permet de dire si l'ajout d'un individu à un ensemble d'individus échangeables entre eux, maintient cette échangeabilité ou non. L'inférence statistique est maintenant vue "«As a passage from data to a unit and not, except as an intermediary, to a parameter»"<sup>12</sup> [Lindley and Novik, 1981].

Bien entendu, il n'est pas possible de donner ici toute la richesse de cette approche bayésienne multiniveau, qui peut utiliser un nombre beaucoup plus élevé de niveaux et mettre en place des méthodes d'analyse très diverses et complexes [Lindley and Smith, 1972 ; Gelman *et al.*, 1995 ; Lee, 1997]. Ces progrès techniques, permis par l'approche subjective, sortent du cadre de cet article et ne seront pas détaillés ici. Nous allons plutôt essayer de montrer en quoi l'approche «Bayésienne» qui, comme nous

<sup>12</sup> «Comme un passage des données vers un individu et non pas vers un paramètre, sauf s'il s'agit d'un intermédiaire vers l'individu».

l'avons déjà indiqué, formalise le degré de croyance personnel de l'approche subjective en une relation logique valable pour tous, permet de dépasser celle-ci.

L'approche «logicienne» reprend, bien entendu, tous les apports de l'approche bayésienne, mais elle permet d'aller plus loin, en particulier lors d'une phase exploratoire que tous les chercheurs en sciences humaines rencontrent. En effet, elle propose un moyen précis d'assigner une distribution de probabilité *a priori*, alors que les hypothèses fréquentistes sont invalides et les méthodes bayésiennes ne sont pas encore disponibles et développe l'assignation précise de cette distribution dans les cas où l'on ne dispose que d'une information parcellaire. Elle utilise pour ce faire le principe de l'entropie maximum (principle of maximum entropy), qui permet d'assigner, sans aucune hypothèse, autre que ce qui est contenu dans les données, un modèle unique qui utilise toute l'information disponible, mais évite de supposer que des informations que l'on n'a pas sont cependant utilisables. Ce modèle est optimal sous de très nombreux critères. On peut dire, par exemple, que ce modèle est le plus simple qui capture toute l'information contenue dans les contraintes; c'est le seul modèle pour lequel ces contraintes "would have been sufficient statistics"; etc. [Jaynes, 2003]. Il permet de fournir une distribution *a priori* (prior) satisfaisante dans un modèle multiniveau.

Plus généralement on voit que cette approche permet de raisonner sur une information incomplète, en la codifiant logiquement et numériquement sous la forme de probabilités. Celles-ci ne décrivent plus un état des choses, comme dans l'approche fréquentiste, mais un état de connaissance. Elles ne sont plus «mesurées», mais assignées selon un processus basé sur l'évidence qui conduit à la même assignation de probabilité, quelle que soit la personne qui effectue cette assignation, contrairement à l'approche subjective.

Cette approche, surtout développée pour le moment dans les sciences dures, pourrait apporter aux sciences humaines une meilleure cohérence. Les premiers essais de Jaynes [1991] d'appliquer le principe de l'entropie maximum à l'économie, en distinguant macro-économie et micro-économie, ne permettent pas encore de dire ce que cette utilisation peut amener dans ce domaine, car il s'agit plus d'idées lancées en vue de stimuler une discussion avec des économistes que d'une théorie bien avancée. Seuls des travaux à venir montreront si ce point de vue logiciste peut permettre de nouvelles avancées en analyse multiniveau.

### 2.2.3 Poursuite de l'exemple démographique

Pour estimer des modèles multiniveau épistémiques, les méthodes classiques bayésiennes n'étaient pas applicables car elles conduisaient à des intégrations d'une trop grande complexité pour permettre l'estimation de leurs paramètres. Seule l'utilisation de simulations permettent cette estimation par des méthodes de Monte Carlo par chaînes de Markov (MCMC) ou par la méthode du *bootstrap*, que la puissance des ordinateurs actuels permet d'utiliser. La présentation de ces méthodes n'étant pas l'objet de cet article nous renvoyons à la thèse de Browne [1998], pour les méthodes MCMC, et aux articles de Carpenter *et al.*, [1999], et de Hutchinson *et al.* [2003].

Reprenons l'exemple précédent des migrations norvégiennes des agriculteurs et des non-agriculteurs. Le tableau 1 donne les paramètres estimés avec une approche objectiviste dans l'article antérieur ([Courgeau, 2002], Tableau 3, Modèle multiniveau contextuel) et ceux estimés sous une approche épistémique par des méthodes de Monte Carlo par chaînes de Markov (MCMC). Ces méthodes permettent en effet, à partir d'une

distribution *a priori* uniforme ou de type Gamma, d'estimer par itérations successives les paramètres du modèle et leur écart-type.

Paramètres	Modèle objectiviste	Modèles épistémiques <i>MCMC</i> avec distributions <i>a priori</i> de type:	
		Gamma	Uniforme
Fixes			
$\alpha_0$ (non-agriculteur)	-2,067 (0,119)	-2,073 (0,129)	-2,090 (0,120)
$\alpha_1$ (agriculteur)	-2,017 (0,340)	-2,023 (0,306)	-2,118 (0,364)
$\alpha_2$ (part d'agriculteurs)	5,420 (1,209)	5,580 (1,690)	5,597 (1,400)
$\alpha_3$ (agr. X part d'agr.)	-8,691 (3,238)	-8,556 (2,988)	-7,984 (3,660)
Aléatoires			
$\sigma_{u_0}^2$ (non-agriculteur)	0,047 (0,024)	0,060 (0,024)	0,084 (0,041)
$\sigma_{u_{01}}$ (covariance)	0,085 (0,042)	0,105 (0,048)	0,142 (0,089)
$\sigma_{u_1}^2$ (agriculteur)	0,181 (0,119)	0,237 (0,121)	0,431 (0,281)

Tableau 1. Paramètres estimés avec deux aléas (écart-types entre parenthèses)

Les paramètres fixes estimés avec un modèle épistémique sont peu différents de ceux estimés avec un modèle objectiviste. Cependant l'intervalle de confiance à 95 % par exemple, que l'on peut tracer autour de ces estimations correspond maintenant bien à ce que l'on attend il s'agit de l'intervalle dans lequel on a une probabilité égale à 0,95 de trouver le paramètre inconnu. Également l'hypothèse de normalité de ces estimateurs n'est plus nécessaire et l'on peut disposer de leur distribution dans tout son détail.

Pour les aléatoires cependant, les différences sont beaucoup plus importantes l'estimateur de la variance pour les agriculteurs est plus du double dans l'estimation *MCMC*, avec distribution *a priori* uniforme, que dans l'estimation objectiviste. Ce fait a déjà été observé [Browne, 1998] pour des modèles de régression linéaire, où les variances au niveau agrégé sont beaucoup plus faibles qu'au niveau individuel. Mais malheureusement pour les modèles logistiques il est difficile de comparer les variances des niveaux agrégés et individuels. On peut cependant penser que cela est dû au fait que la variance estimée n'est pas toujours significative. Qui plus est, si l'on cherche à estimer un modèle épistémique par la méthode du *bootstrap* le modèle ne converge pas et ne permet plus aucune estimation. On voit donc que les modèles épistémiques sont beaucoup plus sensibles, et ne fournissent plus d'estimateur stable lorsqu'une caractéristique n'est pas vraiment significative.

Il nous faut donc maintenant considérer uniquement les effets significatifs supprimer l'aléa régional correspondant aux agriculteurs, remplacer les termes contextuels correspondant aux précédents paramètres  $\alpha_2$  et  $\alpha_3$ , pour ne considérer qu'un seul terme contextuel correspondant à la part d'agriculteurs présents dans la

région, où se trouvent les non-agriculteurs. Le tableau 2 porte ces résultats pour un modèle multiniveau objectiviste et des modèles épistémistes utilisant des estimations du type *MCMC* et *bootstrap*.

Paramètres	Modèle objectiviste	Modèles épistémistes		
		<i>MCMC</i> (Gamma)	<i>MCMC</i> (Uniforme)	<i>Bootstrap</i>
Fixes				
$\alpha_0$ (non-agriculteur)	-2,062 (0,113)	-2,072 (0,114)	-2,052 (0,118)	-2,050 (0,110)
$\alpha_1$ (agriculteur)	-2.260 (0,084)	-2,262 (0,084)	-2,264 (0,084)	-2,263 (0,089)
$\alpha_2$ (non ag. x part ag.)	5,329 (1,475)	5,475 (1,503)	5,379 (1,561)	5,266 (1,507)
Aléatoires				
$\sigma_{u_0}^2$ (non-agriculteur)	0,054 (0,020)	0,061 (0,026)	0,071 (0,032)	0,053 (0,020)

Tableau 2. Paramètres estimés avec un seul aléa (écart-types entre parenthèse)

Maintenant les aléas estimés avec les divers modèles sont beaucoup plus proches et l'estimation *bootstrap* converge sans problème.

### 3. CONCLUSION

La plupart des sciences sociales suivent très souvent une approche objectiviste des phénomènes humains et, rares sont encore les chercheurs dans ce domaine qui prennent une approche pleinement épistémique. Il nous semble cependant que cette situation soit difficile à justifier d'un point de vue logique lorsque l'on examine plus en détail les bases de ces approches. Reprenons et complétons dans cette conclusion les principaux résultats auxquels cet article nous a conduits.

En premier lieu l'utilisation de données de recensements ou de registres de population en démographie, conduit à une observation exhaustive mais qui ne peut être répétée, car les conditions dans lesquelles vivent ces populations vont changer en permanence au cours du temps. Il en est de même en économie, où à nouveau les expériences ne peuvent être répétées. Si l'on désire utiliser des probabilités objectives, il est alors nécessaire d'inventer un ensemble d'univers imaginaires pour pouvoir définir un espace d'échantillonnage. L'existence d'une telle «*super population*» est une hypothèse superflue par rapport à l'objectif du chercheur en sciences sociales, qui est de tester si une relation peut être considérée comme vérifiée ou non dans la population observée. Il paraît dès lors préférable d'utiliser une approche épistémique qui évite de poser de telles conditions et qui permet une estimation utilisant seulement les données observées.

Supposons maintenant que l'on fasse une enquête, ce qui revient à dire que l'on tire un échantillon d'individus, par exemple d'une même génération, et que l'on cherche à estimer à partir des réponses à cette enquête la probabilité pour qu'un individu de cette

génération connaisse un événement donné, à partir de la fréquence observée dans cet échantillon.

Si, dans ce cas, on dispose bien d'un échantillon que l'on peut tirer autant de fois que possible de l'ensemble de la population, il est important de définir et de situer avec précision un intervalle de confiance dans lequel on puisse assurer que se situe une telle probabilité. Sa définition devrait être du type suivant□il fournit l'intervalle dans lequel le statisticien peut être certain de trouver la probabilité recherchée avec une probabilité de 95 % par exemple. L'interprétation objectiviste conduit à considérer l'échantillon sur lequel on travaille comme issu d'une population plus vaste, d'où l'on peut tirer un grand nombre d'échantillons de même taille. On ne peut alors définir qu'un intervalle, dans lequel se trouvera la probabilité que l'on cherche à estimer, cela dans 95 % des tirages. Or ce n'est pas cela que l'on désire estimer, comme nous l'avons indiqué plus haut.

Seule une approche épistémique permet en fait de répondre à cette question. La probabilité ne peut plus se définir en termes absolus, mais en termes de degré d'évidence d'une hypothèse donnée. On peut alors dire que l'intervalle de confiance représente bien un intervalle dans lequel le statisticien peut penser à juste titre qu'il y a une probabilité de 95 % d'y trouver la probabilité pour qu'un individu connaisse l'événement étudié, qui est une hypothèse que l'on peut maintenant tester [Lee, 1989]. Cette approche justifie l'utilisation du théorème de Bayes qui parle de la probabilité d'une hypothèse, notion n'ayant pas de sens pour un objectiviste, alors qu'elle est fondamentale dans l'approche épistémique.

La notion de sous-population, introduite par Fisher pour définir l'inférence statistique dans l'approche objectiviste, pose également de nombreux problèmes en sciences sociales. Le chercheur se trouve dans l'incapacité de discerner et de définir avec clarté les différentes sous-populations à considérer. En revanche, la notion d'échangeabilité, introduite par de Finetti dans l'approche épistémique, se réfère explicitement à des variables aléatoires précises, tandis qu'une sous-population ne le peut□ainsi des individus peuvent être échangeables par rapport à une variable aléatoire, mais non par rapport à une autre ni aux deux simultanément. De même Fisher ne donne aucun moyen de reconnaître si un individu fait partie ou non d'une sous-population, alors que l'échangeabilité, étant un jugement sur des individus, le permet□un individu ne peut changer une fréquence limite, tandis qu'il peut changer l'échangeabilité d'un groupe d'individus auquel on veut le rajouter [Lindley et Novick, 1981].

Enfin, parmi les théories épistémiques, la conception logiciste nous semble la plus prometteuse. Elle permet à partir de principes généraux de logique, ne faisant plus référence au hasard ni aux variables aléatoires, de définir avec précision les règles du calcul des probabilités. Ces règles sont les seules consistantes pour mener une inférence probabiliste [Jaynes, 2003]. Ces principes font disparaître la distinction entre *théorie des probabilités* et *inférence statistique*. Ils permettent de replacer les probabilités objectivistes, de même que les probabilités subjectivistes ou bayésiennes, dans une théorie plus générale des probabilités vue comme une nouvelle logique. Il reste à voir comment on peut les appliquer aux sciences sociales.

## BIBLIOGRAPHIE

- ALLAIS M., «Le comportement de l'homme rationnel devant le risque – critique des postulats et axiomes de l'école américaine», *Econometrica* 21, 4, 1953, p. 503-546.
- ARNAULD A., NICOLE P., *La Logique ou l'Art de penser*, Paris, 1662.
- BAYES T. R., “An essay towards solving a problem in the doctrine of chances”, *Philosophical Transactions of the Royal Society of London* 53, 1763, p.370-418.
- BERNOULLI J. I, *Ars conjectandi*, Bâle, Impensis Thurnisiorum fratrum, 1713.
- BLOSSFELD H.-P., HAMERLE A., MAYER K.U., *Ereignisanalyse*, Frankfurt/New York, Campus Verlag, 1986.
- BRESLOW, N., “Biostatistics and Bayes” (with comments), *Statistical Science* 5, 3, 1990, p. 269-298.
- BROWNE W. J., *Applying MCMC methods to multi-level models*, PhD Thesis, University of Bath, 1998, [accessible via le site internet : <http://www.maths.nott.ac.uk/personal/pmzwbj/bill.html>].
- CARDANO J., *Opera Omnia*, vol. I *De Ludo Aleae*, Amsterdam, 1663.
- CARNAP R., *The logical foundations of probability*, Chicago, University of Chicago Press, 1950.
- CARPENTER J., GOLDSTEIN H., RASBASH J., “A non-parametric bootstrap for multilevel models”, *Multilevel Modelling Newsletter* 11, 1, 1999, p. 2-5.
- CONDORCET, « Mémoire sur le calcul des probabilités. Cinquième partie. Sur la probabilité des faits extraordinaires », *Mémoire pour l'Académie Royale des Sciences pour 1783, 1786*, p. 553-559. [Repris dans *Arithmétique politique. Textes rares ou inédits (1767-1789)*, Paris, Institut National d'études démographiques, 1994, p. 431-436].
- CONGDON P., *Bayesian statistical modelling*, Chichester, John Wiley & Sons, Ltd., 2001.
- COURGEAU D., « Évolution ou révolutions dans la pensée démographique ? », *Mathématiques et Sciences humaines* 160, 2002, p. 49-76.
- COURGEAU D. (éd.), *Methodology and epistemology of multilevel analysis. Approaches from different social sciences*, Methodos Series, vol. 2, Dordrecht/Boston/London, Kluwer Academic Publishers, 2003.
- COURGEAU D., LELIÈVRE E., *Analyse démographique des biographies*, Paris, Ined, 1989. (Traduction anglaise, *Event history analysis in demography*, Oxford: Clarendon Press, 1992. Traduction espagnole, *Análisis demográfico de las biografías*, México, El Colegio de México, 2001).
- COX D. R., “Regression models and life-tables” (with discussion), *Journal of the Royal Statistical Society* 34, 2, 1972, p. 187-220.
- COX R., “Probability, frequency, and reasonable expectation”, *American Journal of Physics* 14, 1946, p. 1-13.
- DRAPER D., “Inference and hierarchical modelling in the social sciences” (with discussion), *Journal of Educational and Behavioural Statistics* 20, 1995, p. 115-147, 233-239.

- DRAPER D., HODGES J. S., MALLOWS C. L., PREGIBON D., "Exchangeability and data analysis", *Journal of the Royal Statistical Society A*, 156, 1993, p.9-37.
- DURKHEIM E., *Les règles de la méthode sociologique*, Paris, Alcan, 1895.
- DURKHEIM E., *Le suicide*, Paris, Alcan, 1897.
- FINETTI (de) B., « La prévision □ ses lois logiques, ses sources subjectives », *Annales de l'Institut Henri Poincaré* 7, Paris, 1937, p. 1-68.
- FINETTI (de) B., *Theory of probability*, 2 vols., London/New York, Wiley & Sons, 1974.
- FISHER R. A., *Statistical Methods for Research Workers*, Edimburgh, Olivier and Boyd, 1925.
- FISHER R. A., *Statistical Methods and Scientific Inference*, Edimburgh, Oliver and Boyd, 1956.
- FISHER R. A., "The nature of probability", *Centennial Review* 2, 1958, p. 261-274.
- FISHER R. A., "Scientific thought and the refinement of human reasoning", *Journal of the Operations Research Society of Japan* 3, 1960, p. 1-10.
- FRANCK R. (ed.), *The explanatory power of models. Bridging the gap between empirical and theoretical research in the social sciences*, Boston/Dordrecht/London, Kluwer Academic Publishers, 2002.
- FRIEDMAN M., SAVAGE L. J., "The utility analysis of choices involving risk", *The Journal of Political Economy* LVI, 4, 1948, p. 279-304.
- GELMAN A., KARLIN J. B., STERN H. S., RUBIN D. B., *Bayesian data analysis*, New York, Chapman and Hall, 1995.
- GOLDSTEIN H., *Multilevel statistical models*, London, Edward Arnold, 2003.
- GRANGER G.-G., *Essai d'une philosophie du style*, Paris, éditions Odile Jacob, 1988.
- GRAUNT J., *Natural and political observations mentioned in a following index, and made upon the bills of mortality*, Londres, 1662.
- GREENLAND S., "Probability logic and probabilistic induction", *Epidemiology*, 1998a, p. 322-332.
- GREENLAND S., "Induction versus Popper: substance versus semantics", *International Journal of Epidemiology* 27, 1998b, p. 543-548.
- GREENLAND S., "Principles of multilevel modelling", *International Journal of Epidemiology* 29, 2000, p. 158-167.
- GUSTAFSON P., AESCHLIMAN D., LEVY A. R., "A simple approach for fitting Bayesian survival models", *Lifetime data analysis*, 9, 2003, p. 5-19.
- HACKING I., *The emergence of probability*, Cambridge, Cambridge University Press, 1975.
- HACKING I., *The taming of science*, Cambridge, Cambridge University Press, 1990.
- HACKING I., *An introduction to probability and inductive logic*, Cambridge, Cambridge University Press, 2001.
- HALLEY E., "An estimate of the degrees of the Mortality of Mankind, drawn from curious Tables of the Births and Funeral's at the City of Breslau; with an Attempt to ascertain the price of the Annuities upon Lives", *Philosophical Transactions for the Month of January* 196, 1693, p. 596-610.

HENRY L., « D'un problème fondamental de l'analyse démographique », *Population*, 13, 1959, p. 9-32.

HOLLAND P., "Statistics and Causal inference" (with Comments), *Journal of the American Statistical Association* 81, 1986, p. 945-970.

HULTING F. L., HARVILLE D. A., "Some bayesian and non bayesian procedures for the analysis of comparative experiments and for small area estimation: Computational aspects, frequentist properties, and relationships", *Journal of the American Statistical Association* 86, n° 415, 1991, p. 557-568.

HUME D., *Philosophical essays concerning human understanding*, 1748.

HUTCHINSON D., MORRISON J., FELGATE R., "Bootstrapping the effects of measurement errors", *Multilevel Modelling Newsletter* 15, 2, 2003, p. 2-10.

IBRAHIM J. G., CHEN M.-H., SINHA D., *Bayesian survival analysis*, New York/Berlin/Heidelberg, Sptinger-Verlag, 2001.

JAYNES E. T., *How should we use entropy in economics?*, 1991, [article non publié accessible via le site internet: <http://www.leibniz.imag.fr/LAPLACE/Jaynes/prob.html>].

JAYNES E. T., *Probability theory: the logic of science*, Cambridge, Cambridge University Press, 2003.

JEFFREYS H., *Theory of probability*, New York, Clarendon Press, 1939.

JENNINSON C., TURNBULL B. W., "Statistical approaches to interim monitoring of medical trials: a review and commentary", *Statistical Science* 5, 3, 1990, p. 299-317.

JONES K., *Everywhere is nowhere: Multilevel perspectives on the importance of place*, The University of Portsmouth Inaugural Lectures, 1993.

KEYNES J. M., *A treatise on probability*, London, Macmillan, 1921.

KIM Y., LEE J., "Bayesian bootstrap for proportionnal hazards models", *The Annals of Statistics* 31, 6, 2003, p. 1905-1922.

KOLMOGOROV A., "Grundbegriffe der wahrscheinlichkeitsrechnung", *Ergebnisse der mathematik*, vol.2, Berlin, 1933.

KRÜGER L., MORGAN M., GIGERENZER, *The probabilistic revolution*, Cambridge, Cambridge University Press, 1986.

KERSSEBOOM W., « Troisième traité sur la grandeur probable de la population de Hollande et de Frise occidentale », *Essais d'Arithmétique politique contenant trois traités sur la population de la province de Hollande et de Frise occidentale*, Paris, éditions de l'Ined, 1970, (1<sup>e</sup> édition, 1942).

LANCASTER T., *The econometric analysis of transition data*, Cambridge, Cambridge University Press, 1990.

LAPLACE P. S., *Théorie analytique des Probabilités*, 2 vols., Paris, Coursier imprimeur, 1812.

LEE P. M., *Bayesian statistics*, first ed. London, Arnold, 1989.

LINDLEY D. V., NOVICK M. R., "The role of exchangeability in inference", *The Annals of Statistics* 9, 1981, p. 45-58.

LINDLEY D. V., SMITH A. F. M., "Bayes estimates for the linear model", *Journal of the Royal Statistical Society, Series B (Methodological)* 34, 1972, p.1-41.

- MATALON B., «Épistémologie des probabilités», Piaget J. (éd.), *Logique et connaissance scientifique*, Paris, Gallimard, 1967, p. 526-553.
- MISES von R., *Wahrscheinlichkeit, statistik und wahrheit*, Wien, Springer, 1928.
- MISES von, R., *Probability, Statistics and truth*, London, Allen and Unwin, 1957.
- MOLINA E. C., “Bayes’ theorem: an expository presentation”, *The Annals of Mathematical Statistics* 2, 1, 1931, p. 23-37.
- NADEAU R., *Vocabulaire technique et analytique de l’épistémologie*, Paris, Presses Universitaires de France, 1999.
- PASCAL, B., *Traité du triangle arithmétique*, Paris, 1651.
- PASCAL B., *Pensées*, Paris, 1670.
- POLYA G., *Mathematics and plausible reasoning*, 2 vols, Princeton, University Press, 1954.
- POPPER K., “The propensity interpretation of probability”, *Philosophy of Science* 10, 1959, p. 25-42.
- POPPER K., *Le réalisme de la science*, Paris, Herman, 1990.
- PORTER T. M., *The rise of statistical thinking 1820-1900*, Princeton, University Press, 1986.
- RAMSEY F. P., *Mr Keynes and Probability*, Cambridge Magazine, 1922.
- RAMSEY F. P., “Truth and probability”, in Ramsey, F.P. (1931), *The foundations of mathematics and other logical essays*, R.B. Braithwaite ed., Ch. VII., Londres, Kegan, Tribner & Co., New York, Harcourt, Brace and Company, 1926, p. 156-198.
- ROBINS J. M., SCHEINES R., SPIRITES R., WASSERMAN L., “Uniform consistency in causal inference”, *Biometrika* 90, 2003, p. 491-515.
- ROSENBAUM P. R., “The consequence of adjustment for a concomitant variable that has been affected by the treatment”, *Journal of the Royal Statistical Society A* 147, 1984a, p. 656-666.
- ROSENBAUM P. R., “From association to causation in observational studies: the role of tests of strongly ignorable treatment assignment”, *Journal of the American Statistical Association* 79, 1984b, p. 41-48.
- ROSENBAUM P. R., RUBIN D. B., “Reducing bias in observational studies using subclassification on the propensity score”, *Journal of the American Statistical Association* 79, 1984, 516-524.
- ROUANET H., BERNARD J.-M., BERT M.-C., LECOUTRE B., LECOUTRE M.-P., LE ROUX B., *New ways in statistical methodology. From significance tests to Bayesian inference*, Bern, Peter Lang, 1998.
- RUBIN D. B., “Bayesian inference for causal effects: the role of randomization”, *The Annals of Statistics* 6, 1978, 34-58.
- SAVAGE L. J., *The foundations of statistics*, New York, Wiley, 1954.
- SAVAGE L. J., “Implications of personal probability for induction”, *The Journal of Philosophy* 64, 19, 1967, p. 593-607.
- SAVAGE L. J., “On reading R.A. Fisher”, *The Annals of Statistics* 4, 3, 1976, p. 441-500.

SHAFER G., “The unity and diversity of probability” (with comments), *Statistical Science* 5, 4, 1990, p. 435-462.

SOBEL J. H., “On the evidence of testimony for miracles: A Bayesian interpretation of David Hume analysis”, *The Philosophical Quarterly* 37, 147, 1987, p. 166-186.

STIGLER S. M., “Thomas Bayes Bayesian inference”, *Journal of the Royal Statistical Society, Series A* 145, 1982, p. 250-258.

SUPPES P., “Representation of probability”, Suppes P., *Representation and invariance of scientific structures*, Stanford, CSLI Publication, 2002, p. 129-264.

TRUSSELL J., “Introduction”, Trussell J., Hankinson R., Tilton J. (eds.), *Demographic applications of event history analysis*, Oxford, Clarendon Press, 1992, p. 1-7.

TUMA N. B., HANNAN M., *Social dynamics*, London, Academic Press, 1984.

VAN IMHOFF E., POST W., «Méthodes de micro-simulation pour des projections de population», D. Courgeau éd., *Population* 52 ,4, 1997, p. 889-932. “Microsimulation methods for population projections”, D. Courgeau éd., *Population. An English Selection*, 10, 1, 1998, p. 97-138.

VENN J., *The logic of chance*, London, Macmillan, 1866.

WUNSCH G., (1994), «L’analyse causale en démographie», Franck R. (éd.), *Faut-il chercher aux causes une raison? L’explication causale dans les sciences humaines*, Paris, Librairie Philosophique J. Vrin, 1994, p. 24-40.