Mathématiques et sciences humaines
Mathematics and social sciences

176 | 2006
176, Varia

Contribution à l'histoire des tests laplaciens

Contribution to the history of laplacian tests

Michel Armatte

Édition électronique
URL : http://msh.revues.org/3674
DOI : 10.4000/msh.3674
ISSN : 1950-6821

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

Édition imprimée
Date de publication : 1 décembre 2006
Pagination : 117-133
ISSN : 0987-6936

Référence électronique
Michel Armatte, « Contribution à l'histoire des tests laplaciens », *Mathématiques et sciences humaines* [En ligne], 176 | Hiver 2006, mis en ligne le 28 juillet 2006, consulté le 12 octobre 2016. URL : http://msh.revues.org/3674 ; DOI : 10.4000/msh.3674

Ce document est un fac-similé de l'édition imprimée.

© École des hautes études en sciences sociales

Math. Sci. hum ~ Mathematics and Social Sciences (44e année, n° 176, *2006*(4), p. 117-133)

CONTRIBUTION À L'HISTOIRE DES TESTS LAPLACIENS

Michel ARMATTE[1]

RÉSUMÉ – *Au croisement des recherches historiques de Bernard Bru sur la statistique laplacienne, et des interrogations récurrentes depuis les années 1960 sur les usages peu satisfaisants des tests statistiques dans les sciences d'observation, nous revisitons la mise en place des tests statistiques dans l'œuvre de Laplace. On trouve déjà dans celle-ci un mélange entre plusieurs approches, et des audaces interprétatives, dignes de la version aujourd'hui pratiquée, hybride entre les apports de Bayes-Laplace, Fisher et Neyman.*

MOTS-CLÉS – Astronomie, Bayes, Cournot, Fisher, Laplace, Neyman, Sex-ratio, Tests d'hypothèses, Tests statistiques

SUMMARY – Contribution to the history of laplacian tests

Crossing the historical studies of Laplace statistics by Bernard Bru and the critical questioning of the modern uses of hypothesis that began in the 1960's, we are tracking the use of hypothesis testing in Laplace's texts. We can already find in these writings a combination of several approaches and interpretations, ressembling that of the current practice which is a hybrid of Bayes-Laplace's, Fisher's and Neyman's conceptions of statistical testing.

KEY-WORDS – Astronomy, Bayes, Cournot, Fisher, Hypothesis tests, Laplace, Neyman, Sex-ratio, Statistical tests

Depuis les alertes lancées par certains psychologues dès les années 1960, les textes séminaux de G. Gigerenzer [1987, 1993], et les travaux en France de Benjamin Matalon, Henri Rouanet, Bruno Lecoutre et Jacques Poitevineau, les vagues de publications se sont succédées pour dénoncer la version hybride de la théorie statistique des tests de signification. Celle-ci est accusée d'amalgame entre les approches de Fisher, Neyman, et des néobayésiens, et de promotion d'une vulgate hybride qui domine à la fois l'enseignement de la théorie des tests, la pratique quotidienne des chercheurs, et la politique de publication des éditeurs de journaux scientifique, dans de nombreux domaines des sciences expérimentales ou d'observation, au premier rang desquelles la psychologie et l'économétrie. Un article récent de Daniel Denis dans le *Journal de la Société Française de Statistique*, dénonçant l'écart des pratiques au dogme Fisherien en la matière – un écart que l'on constaterait d'ailleurs tout autant à l'approche de Neyman – a donné une fois de plus l'occasion de faire le point sur les multiples glissements de sens et de pratiques qu'ont subi ces méthodologies forgées dans les années trente dans le

[1] Département d'économie appliquée, Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny 75775 Paris cedex 16 et Centre A. Koyré, michel.armatte@dauphine.fr

cadre d'un paradigme de l'inférence inductive chez Fisher et dans celui d'une théorie de la Décision chez Neyman et Pearson.[2]

On sait bien maintenant que l'approche de Fisher, forgée entre 1925 et 1935, définit une seule hypothèse dite « hypothèse nulle[3] », paramétrique ou non, suppose un échantillon aléatoire résultant éventuellement d'un plan d'expérience par « randomization » et extrait une seule fois d'une population « infinie ». Il juge que cette hypothèse doit être rejetée (au profit d'un effet dit significatif) si la probabilité *p* – « p-value » ou « probabilité critique » – d'obtenir un résultat (plus généralement une fonction pivotale de l'échantillon) au moins aussi extrême que celui qui est observé est suffisamment faible, c'est-à-dire inférieur à un « seuil de significativité » que Fisher a d'abord fixé conventionnellement à 5 % avant de préférer une indication du vrai niveau de signification *p*. Le test de signification est pour Fisher un élément d'une méthodologie plus générale de l'inférence inductive qui comprend aussi l'estimation ponctuelle et le calcul de probabilités fiducielles.

L'approche de Neyman et Pearson qui lui est opposée dès 1928 s'inscrit dans une autre démarche, celle de la décision. Elle définit *a priori* deux hypothèses H0 (nulle) et H1(alternative), et deux risques α et β correspondant aux probabilités respectives des erreurs de premier type et de second type, c'est-à-dire de choisir H1 sachant H0 vraie, et H0 sachant H1. Sur la base d'un échantillon tout aussi aléatoire extrait d'une population finie, et ayant vocation a être répété, une règle de décision fondée sur une certaine statistique d'échantillon S conduit à une bipartition de l'espace des échantillons possibles en une région d'acceptation de H0 et une région critique (rejet de H0 au profit de H1) et un couple de valeurs de α et β. Neyman propose de rompre la symétrie : de fixer d'abord α puis de choisir le test, c'est-à-dire la bipartition, qui minimise β (= qui augmente la « puissance » du test).

Or ce contre quoi les différents auteurs contemporains mettent en garde, c'est la diffusion massive d'une version hybride de ces deux approches qui confond falsification et décision, qui suppose 2 hypothèses mais n'en spécifie qu'une, qui ne se soucie plus de randomisation, qui prend $\alpha = 5$ % comme une norme, qui utilise indifféremment la p-value ou la valeur critique, et qui interprète *1 - p* comme un degré de corroboration de H1. Or cette interprétation associée à la règle implicite de ne publier que les effets significatifs provoque de nombreux biais de publication. Dans une discussion de l'article de Denis, nous avons déjà argumenté sur la diversité disciplinaire des méthodologies hybrides construites et surtout pratiquées – elles ne sont pas les mêmes et n'ont pas les mêmes fonctions en psychologie et en économie – et nous avons conclu que l'hybridisation, loin d'être une tare épistémologique, était sans doute une nécessité des pratiques de la recherche empirique. Nous avons aussi suggéré que ces hybrides pourraient bien avoir une très ancienne existence, du vivant même des premières formulations de tests statistiques.

Cependant peu nombreux sont les participants de la controverse qui ont fait le détour historique pour la préciser et l'enrichir, et moins nombreux encore sont ceux qui sont remontés aux XVII^{e} et XVIII^{e} siècle, et aux raisonnements en forme de syllogisme aléatoire qui se trouvent dans les discussions sur la Providence divine, sur l'efficacité d'une pratique médicale comme l'inoculation, sur des contrats aléatoires juridiques ou

[2] D.J. Denis, "The modern hypothesis testing hybrid R.A. Fisher's fading influence", avec réponses de M. Armatte, B. Bru, J. Gill, E.Kwan, B. et M.P. Lecoutre et J. Poitevineau, S. Stigler, *Journal de la dSFDS,* Vol 145, n° 4, 2004, p. 5-68.
[3] "To be nullified"

économiques, ou encore sur des conjectures scientifiques. La systématisation de ces raisonnements, armés du calcul analytique des probabilités triomphe avec les travaux de Laplace à la fin du XVIII[e] siècle. Si la théorie des tests ne commence pas avec Fisher, ni même un peu plus tôt avec Edgeworth comme l'affirme une tradition anglaise[4], il est intéressant de revisiter l'œuvre de Laplace qui a tant fait en la matière. Et je ne fais ici que suivre le pionnier en la matière et le meilleur des connaisseurs de cette œuvre, à savoir Bernard Bru, auquel ce texte, comme l'ensemble du recueil, veut rendre hommage. C'est dans le cadre des premières séances du Séminaire d'histoire du calcul des probabilités et des statistiques[5], dans les années 1983 et 1984, que j'ai entendu les premiers exposés de Bernard Bru sur l'œuvre de Laplace, exposés dont il ne reste guère que les notes et la postface de sa réédition de *l'Essai sur les probabilités*. De mes propres notes prises à une des séances, je retiens cette phrase[6] qui pose bien le problème :

> *Les deux théories fisheriennes et bayésiennes sont toutes les deux de Laplace… mais contradictoires*

Bru rappelait alors que tous les tests laplaciens se ramènent à un seul principe : Quand la probabilité p(E/H0) d'une observation E conditionnelle à une hypothèse H0 est trop faible, on ne trouve pas de raison suffisante pour accepter l'hypothèse H0. Plus précisément, il ne s'agit pas de la probabilité de l'observation E mais de toute valeur de la statistique X aussi extrême, aussi éloignée de la valeur espérée sous H0, que celle qui résulte de l'observation faite une seule fois. C'est le principe de tous les tests. C'est la formule magique à deux négations qui pose tant de problèmes.

Vingt ans plus tard, Bernard Bru commentant l'article de Denis dans le JSFDS, prend la défense des chercheurs qui appliquent maladroitement les tests de signification :

> *certes ils se trompent et ce n'est pas bien, mais tant de savants et des plus grands se sont trompés sur ces questions qu'ils se trompent en bonne compagnie…Fisher lui-même y perd son latin…Mais le premier et le plus connu c'est « la faute de Laplace ».*

Et Bru évoque alors deux ou trois exemples de ces turpitudes laplaciennes. Trop rapidement. Ce texte se propose d'y revenir. Commençons d'abord par 3 exemples de la *Théorie analytique*.

1. PLANÈTES ET COMÈTES

Après le Livre I de la *Théorie analytique*[7] consacré aux fonctions génératrices, puis aux approximations « des formules qui sont fonctions de très grands nombres » et au modèle binomial obtenu par répétition indépendantes d'épreuves binaires à probabilités fixes, c'est au Livre II que Laplace développe une *Théorie générale des probabilités*. L'exposé des principes généraux de ce calcul explore l'approche classique de Huygens

[4] "The actual history of testing begin with Edgeworth", nous dit un member important de l'Ecole (économétrique) de Hendry à la London School of Economics dans son dernier manuel [Spanos, 1999, p. 682)].

[5] La liste des exposés à ce séminaire jusqu'en 1990 a été publiée dans *Mathématiques et Sciences humaines,* n° 113, 1991, p. 57-75.

[6] Des citations approximatives en l'absence d'enregistrement et de rédaction authentifiées.

[7] Nous travaillons sur l'édition numérique Gallica de la troisième édition de 1820.

et l'approche bayésienne de la probabilité des causes. Laplace se propose alors dans un long chapitre II d'appliquer les formules qu'il vient d'établir à « la probabilité des événements composés d'événements élémentaires ». Après les exemples d'urne et de dés (somme *s* des points de *i* dés à *n+1* faces), il en vient (p. 261) à une question de philosophie naturelle, à savoir :

> *la probabilité que la somme des inclinaisons à l'écliptique d'un nombre s d'orbites sera comprises dans des limites données, en supposant toutes les inclinaisons, depuis zéro jusqu'à l'angle droit, également possibles.*

La somme des inclinaisons des 10 planètes sur l'écliptique était de 91,4187 degrés[8] au commencement de 1801. Or la probabilité que la somme des orbites soit comprise entre 0 et cette valeur « si toutes les inclinaisons étaient également possibles » est de $(1/10\ !).(0{,}914187)^{10}$ qu'il faut combiner dit-il avec la probabilité $(1/2)^{10}$ d'avoir toutes les planètes tournant dans le même sens que la terre, soit $1{,}0972.10^{-10}$. Il en conclut que la probabilité de l'événement complémentaire :

> *s'approche tellement de la certitude que le résultat observé devient invraisemblable dans cette hypothèse ; ce résultat indique donc, avec une très grande probabilité l'existence d'une cause primitive qui a déterminé les mouvements des planètes à se rapprocher du plan de l'écliptique, ou plus naturellement du plan de l'équateur solaire et à se mouvoir dans le sens de la rotation du soleil.*

Laplace applique une méthode analogue aux 100 comètes identifiées jusqu'en 1811, dont 53 sont directes et 47 rétrogrades. L'inclinaison moyenne est de 51° 73677 soit 1° 73677 de plus qu'un demi angle droit.

> *On trouve ainsi 0,474 pour la probabilité que l'inclinaison des 100 orbites doit tomber dans les limites 50° ±1° 17377 ; la probabilité que l'inclinaison moyenne [soit] inférieure à l'inclinaison observée est donc 0,737.*

Cette probabilité n'est pas assez grande pour que le résultat observé fasse rejeter l'hypothèse d'une égale facilité des inclinaisons des orbites et pour indiquer l'existence d'une cause primitive qui a influé sur ces inclinaisons. La même chose a lieu pour le sens du mouvement puisqu'il montre que la probabilité que 47 comètes au plus soient rétrogrades est de 0,3046 :

> *cette probabilité est beaucoup trop grande pour indiquer une cause qui ait favorisé dans l'origine les mouvements directs.*

Ainsi peut-il conclure que :

> *la cause qui a déterminé le sens des mouvements de révolution de rotation des planètes et des satellites ne parait pas avoir influé sur le mouvement des comètes.*

Le principe du test Laplacien est bien établi à travers ces 2 exemples aux conclusions opposées. Sous l'hypothèse d'un hasard qui consisterait à jeter planètes et comètes au hasard (mais quel hasard ?… Ici c'est la loi uniforme sur les inclinaisons) la probabilité d'obtenir au moins l'écart qui est observé est si faible que *le résultat observé devient invraisemblable dans cette hypothèse* et justifie son rejet, et l'acceptation d'une autre hypothèse – générale et non spécifiée – qu'une cause particulière a joué.

[8] Le degré de Laplace est la centième partie du quart de cercle.

Près de vingt-cinq ans plus tard, Cournot reprenant la question dans son *Exposition de la théorie des chances* (Chap. 12, n° 145, p. 173) fait un autre calcul, un calcul fondé sur la règle de Bayes, qui donne la probabilité *a posteriori* que la chance du mouvement direct des planètes surpasse ½ : soit $(2^{12} - 1)/2^{12} = 4095/4096$.

> *Il y a donc 4095 contre 1 à parier que qu'une cause quelconque a rendu les mouvements directs plus aptes à se produire que les mouvements rétrogrades*

écrit il, semblant donner raison à Laplace. Mais trouvant un résultat analogue pour la concentration des nœuds ascendants dans la moitié de l'écliptique où les longitudes sont moindres que 180°, Cournot juge cette accumulation totalement fortuite. Un « pur effet du hasard » que de très petits déplacements d'orbite pourraient faire disparaître. Projetons une sphère 11 fois sur le sol et relevons les points de contact. Il n'y a rien d'extraordinaire à ce qu'ils soient tous du même côté d'un méridien passant par l'un d'entre eux, rien qui motive l'intervention d'une cause spéciale :

> *et tous les calculs de probabilités sur lesquels on fonderait l'existence d'une telle cause seraient illusoires.*

Imaginer par ailleurs pour les inclinaisons des planètes que le hasard se traduise par une équiprobabilité des directions n'est pas plus pertinent qu'une équiprobabilité de leurs sinus. L'arbitraire dans la construction du hasard est flagrant (comme dans le problème de Borel d'une corde « prise au hasard » sur un cercle).

Enfin Cournot conteste vivement la conclusion de Laplace sur les Comètes : « par une méprise bien singulière, Laplace a conclu que les comètes… paraissent avoir une tendance contraire aux Planètes » dit Cournot. Choisissant un autre système de coordonnées que latitudes et longitudes qui n'offrent pas de symétrie, Cournot prouve par une même démarche probabiliste le contraire de ce qu'a affirmé le grand Laplace ! Il évalue pour un certain angle à 9999 contre 1, la probabilité :

> *que des causes constantes ont tendu à rapprocher du plan de l'écliptique les plans des orbites cométaires.*

Cournot pousse l'étude jusqu'à identifier la cause d'une dominance de certains types d'orbites dans les conditions mêmes de l'observation. Enfin il déclare que :

> *contrairement à un préjugé accrédité, il n'est pas toujours nécessaire dans ce type de discussion de disposer d'un très grand nombre d'observations.*

On est d'ailleurs frappé par l'accent mis non pas sur la précision (celle des approximations autant que celle des résultats annoncés par Laplace avec 8 chiffres significatifs) mais sur les conventions mêmes de la construction du hasard, ou encore sur ce que Fisher aurait appelé un « plan d'expérience ».

2. PRESSION ATMOSPHÉRIQUE

Après un chapitre III consacré aux « lois de la probabilité qui résultent de la multiplication indéfinie des événements », le chapitre IV de la *Théorie analytique* traite complètement « de la probabilité des erreurs des résultats moyens d'un grand nombre d'observations », c'est-à-dire du rôle de la loi en $\exp(-x^2)$ et de la méthode des moindres carrés en théorie des erreurs, et le chapitre V (paragraphe 25, p. 355) traite de « l'application du calcul des probabilités à la recherche des phénomènes et de leurs causes ». Le principe général y est ainsi exposé :

> *Les phénomènes de la nature se présentent le plus souvent accompagnés de tant de circonstances étrangères, un si grand nombre de causes perturbatrices y mêlent leur influence qu'il est très difficile lorsqu'ils sont très petits, de les reconnaître. On ne peut alors y parvenir qu'en multipliant les observations afin que les effets étrangers venant à se détruire, le résultat moyen des observations ne laisse plus apercevoir que ces phénomènes. On conçoit par ce qui précède que cela n'a lieu rigoureusement que dans le cas d'un nombre infini d'observations. Dans tout autre cas, les phénomènes sont indiqués par le résultat moyen que d'une manière probable, mais qui l'est d'autant plus que les observations sont en plus grand nombre.*

Laplace prend alors l'exemple de la variation diurne du baromètre, c'est-à-dire des écarts à peine perceptibles de la pression atmosphérique entre jour et nuit. Il se propose de prouver par le calcul des probabilités « l'existence du phénomène et son étendue », dans une approche qui a fait écrire à Sheynin, dans son article sur Newton, que « Laplace y anticipe la méthode de l'hypothèse nulle ». D'où notre intérêt pour cet exemple. Supposant observée la hauteur du baromètre en un même lieu vers 9h du matin et 4h du soir pendant s = 400 jours – c'est une expérience fictive – Laplace s'intéresse à la différence q entre somme des mesures du matin et somme des mesures du soir :

> *Cette différence indiquera une cause constante qui tend à élever le baromètre vers neuf heures et à l'abaisser vers quatre heures du soir. Pour déterminer avec quelle probabilité cette cause est indiquée, concevons que cette cause n'existe point et que la différence observée q résulte des causes perturbatrices accidentelles et des erreurs des observations.*

Sous cette hypothèse d'absence de cause (que Fisher baptise un siècle plus tard « hypothèse nulle »), donc d'une certaine forme de hasard qui reste à préciser, les écarts de températures se compenseraient en moyenne et la valeur attendue de q serait 0. Mais bien sûr la valeur observée est bien plus grande. Pour une étendue de la variation diurne de l'ordre du millimètre – « c'est à peu près ce que M. Ramond a trouvé dans nos climats » – on pourrait trouver (l'expérience est fictive) une valeur de q observée de 400 × 1 = 400mm. Laplace cherche donc à calculer la probabilité que la différence q soit inférieure à cette valeur, sous cette même hypothèse « nulle ». Sous cette hypothèse, Laplace ne manque pas d'identifier la loi de q à « la loi des erreurs[9] », et son calcul passe donc par l'intégrale de cette loi et ses approximations développées dans les chapitres précédents. Laplace conclut :

> *que la probabilité cherchée (de q < 400) approche tellement de l'unité ou de la certitude qu'il est extrêmement probable que s'il n'existait point de cause constante de l'excès observé* (…) *cet excès serait plus petit que 400 mm. Il indique donc avec une extrême vraisemblance l'existence d'une cause constante qui l'a produit.*

C'est bien un raisonnement « par l'absurde » ou par double négation qui démontre l'existence du phénomène : sous une hypothèse (nulle) de pur hasard il est quasiment impossible d'atteindre (et dépasser) un tel écart observable. Notons dès maintenant que le raisonnement ainsi outillé porte avec lui un certain nombre d'assimilations implicites qui pourraient bien poser problème :

1. Concevoir que l'effet n'existe point n'est pas si facile. C'est avoir un modèle du hasard. Ici l'auteur ne conçoit qu'une forme de hasard, celle qu'il identifie à la loi

[9] Dite plus tard « loi de Laplace-Gauss », et bien plus tard encore avec Galton « loi normale ».

des erreurs de Laplace-Gauss. Or cette hypothèse très féconde, placée au centre de la théorie laplacienne, est loin d'être acceptée dans tous les cas. Elle donnera lieu à de nombreuses controverses tout au long du siècle.

2. Le critère de décision est la probabilité critique, c'est-à-dire non pas la probabilité de l'observation, mais la probabilité d'obtenir un résultat au moins aussi excessif que celui qui est observé. La voie est ouverte à une pratique de data mining, celle même qui a été tant de fois critiquée de Cournot aux économètres des années 1980 : répétons l'épreuve et on finira bien par obtenir un résultat significatif, c'est-à-dire suffisamment extrême pour justifier d'une cause (cf. *infra*).
3. La troisième assimilation opère un glissement d'interprétation : Si la probabilité critique de l'observation E (observer 400 et plus pour q) sous H0 est faible, c'est que la probabilité de H0 sachant E est faible. Autrement dit, on glisse par simple abus de langage d'un jugement sur une probabilité conditionnelle $p_{H0}(E)$ à un jugement équivalent sur une autre probabilité conditionnelle $p_E(H0)$. La faiblesse de la première probabilité serait « indicatrice » du degré de réfutation de l'hypothèse nulle. Or rien n'autorise une telle inférence dans le calcul classique des probabilités conditionnelles. Si la première probabilité est calculable la seconde ne l'est pas, sauf à postuler une distribution de probabilité *a priori* sur les hypothèses. Et c'est bien dans le seul cas où ce postulat est celui de l'équiprobabilité, que le calcul bayésien conduit à la proportionnalité des deux jeux de probabilités conditionnelles. Il est impossible de tirer du test une probabilisation des hypothèses, donc une quantification de leur corroboration, sans adhérer au *principe de raison insuffisante* identifiant ignorance et loi uniforme sur les hypothèses. Un principe que Laplace accepte mais que Fisher rejettera avec véhémence.
4. La quatrième assimilation consiste à confondre hasard et absence de cause constante, donc à transformer le rejet du hasard (H0) en une acceptation d'un effet de l'heure sur la pression atmosphérique, ce qui devient implicitement la seule hypothèse alternative plausible. Une sorte d'hypothèse alternative H1 qui n'ose dire son nom… avant les travaux de Neyman.
5. La question de la taille d'échantillon qui permet de faire un tel test n'est pas anodine. Commentant, sans en préciser l'origine, ces quelques résultats, Cournot (§ 162, p. 191) est obligé de reconnaître qu'il faut, ici , « embrasser un grand nombre d'observations » pour faire disparaître les anomalies et surgir les causes constantes, encore qu'on peut faire baisser ce nombre en prenant non pas l'écart moyen de deux séries indépendantes relatives au matin et au soir, mais la moyenne des écarts entre des observations consécutives, ce qui élimine l'influence de causes irrégulières journalières. La question soulevée par Cournot est à la fois celle de la validité de la démarche laplacienne pour des petits échantillons alors même que les approximations ne fonctionnent plus (les travaux de Student en 1908 y apporteront une réponse) et celle du plan d'expérience qui donnera le plus grand pouvoir de discrimination au test (ce que Neyman appellera sa puissance).

Où l'on voit donc que Laplace développe ici une méthodologie des tests de signification qui est de même nature que celle de Fisher un siècle plus tard. Mais l'argumentation emprunte déjà certaines de ses interprétations soit à un raisonnement bayésien soit à l'introduction d'une hypothèse alternative qui ne sera explicite que chez Neyman. L'hybridation n'est pas une création des psychologues du XX^{e} siècle et de nos manuels contemporains. Elle est là, depuis le début.

3. SEX RATIO

Le Chapitre VI de la même *Théorie analytique* est consacré à « la probabilité des causes et des événements futurs, tirée des événements observés ». Laplace y combine la théorie de la probabilité *a posteriori* que Laplace a adoptée dès 1774 avant même d'avoir lu Bayes, et qu'il a développé au chapitre I du même Livre II. Le principe en est l'application de ce que l'on appelle aujourd'hui la formule de Bayes, couplé avec le lemme d'une équiprobabilité *a priori* des valeurs possibles de la probabilité cherchée.

Soit x la vraie possibilité à connaître de l'événement E et y la probabilité du résultat observé en fonction de x. Si l'on considère les différentes valeurs de x comme autant de causes de ce résultat, la probabilité que x soit comprise entre θ et θ' sera donnée par :

$$p(\theta < x < \theta') = \frac{\int_{\theta}^{\theta'} y dx}{\int_{0}^{1} y dx}$$

Qui se lit comme la probabilité *a posteriori*, une fois le résultat de probabilité y observé, que la possibilité *a priori* x soit comprise dans un certain intervalle. Laplace justifie le lemme d'équiprobabilité non pas comme il l'est parfois malheureusement par un principe d'ignorance – si je ne sais rien alors cela équivaut à choisir une répartition uniforme des valeurs possibles de x – mais par le fait que l'on peut toujours s'y ramener même si l'on a une autre hypothèse de distribution *a priori* [10].

Le second exemple développé par Laplace à cet endroit est celui du sex ratio à la naissance. C'est un sujet qui a déjà une longue tradition dans l'histoire du calcul des probabilités depuis Arbuthnot, s'Gravesande, et Süssmilch[11]. Rappelons en particulier que pour le médecin écossais Arbuthnot (1710) la supériorité des naissances masculines observée à Londres pendant 82 années successives aurait, sous l'hypothèse d'une équiprobabilité des sexes, une probabilité si faible – $(1/2)^{82} = 4{,}3\ 10^{-24}$ – que cette hypothèse serait inadmisible. Arbuthnot en concluait que :

> *cette inégalité des mâles et des femelles n'est pas l'effet du Hasard mais celui de la Providence Divine.*

Dans une correspondance avec s'Gravesande et Montmort, Nicolas Bernoulli (neveu de Jacques) réinterprète l'argument d'Arbuthnot :

> *ce serait un miracle que les nombres des enfants de chaque sexe nés à Londres ne se sont pas plus éloigné les uns des autres pendant 82 ans de suite*

et en prend le contre-pied. Il prétend :

> *qu'il n'y a aucun sujet de s'étonner et qu'il y a une grande probabilité pour que le nombre des mâles et des femelles tombe entre des limites encore plus petites que celles que l'on a observées*

---

[10] « Quand les valeurs de x, considérées indépendamment du résultat observé, ne sont pas également possibles, en nommant z la fonction qui exprime leur probabilité, il est facile de voir, par ce qui a été dit dans le chapitre 1er de ce Livre, qu'en changeant dans la formule y en yz, on aura la probabilté que x est dans les limites x = θ et x = θ'. Cela revient à supposer toutes les valeurs de x également possibles *a priori*, et à considérer le résultat observé comme étant formé de deux résultats indépendants dont les probabilités sont y et z».

[11] Cf. le dossier établi par Norbert Meusnier (CAMS, n° 162).

observant que le taux de masculinité sur les 82 années est de l'ordre de 18/17, Bernoulli évalue sur la base de cette hypothèse – « imaginons 14 000 dés à 35 faces chacun dont 18 soient blanches et 17 noires » – la probabilité d'un intervalle donné assez étroit (7037/14000 ; 7363/14000) autour de cette valeur : elle est de 0,978 chaque année[12]. On sort de cet intervalle seulement 11 fois sur 82, événement qui aurait une probabilité de 1/227 = 0,004[13] d'être dépassé « par hasard ». Avec une telle probabilité critique, nous aurions conclu que cette hypothèse 18/17 est à rejeter, et 16/15 aurait été meilleur. Bernoulli en tire une autre conclusion : un tel resserrement est bien le résultat du hasard. Ce faisant, Bernoulli a en tête un autre hasard qu'Arbuthnot qui s'étonnait lui, que les ratios observés s'écartent tant de la valeur ½. Le hasard de Bernoulli donne 18 chances contre 17 aux garçons. Et sa conclusion ne s'impose pas, puisque De Moivre en tire une conclusion opposée à savoir que « ce rapport lui-même (quel qu'il soit), ou si vous voulez la forme du dé, est un effet de l'intelligence et du dessein ». Arbuthnot et Bernoulli usent bien d'un test de signification à la Fisher avec une hypothèse « nulle » qui correspond chez Arbuthnot à x = 1/2, et chez Bernoulli à 17/16. Ils ne testent donc pas le même hasard. Et les intéresse principalement l'interprétation de cette hypothèse alternative : Autre hasard ? Providence divine ? Miracle ?

Que fait Laplace sur cette question ? Il présente d'abord les chiffres des naissances de Paris, Londres et Naples que l'on peut résumer par le tableau suivant :

Ville	Période	Naissances de Garçons	Naissances de Filles	Ratio G/F
Paris	1745-1784	393 386	377 555	25/24
Londres	1664-1758	737 629	698 958	19/18
Naples	1774-1782	782 352	746 821	22/21

Puis il cherche, pour chacune de ces villes, compte tenu des nombres de naissances observées, « la probabilité que la possibilité de la naissance d'un garçon dépasse ½ ». Il applique le calcul précédent pour Paris, cas le plus défavorable, et trouve pour la probabilité que x > 1/2, une valeur si proche de 1 (l'écart est de l'ordre de 10^{-72}) que l'on doit, dit-il « regarder cette probabilité comme au moins égale à celle des faits historiques les plus avérés ».

Laplace ne reprend pas le même calcul pour Londres et Naples, qui serait encore plus favorable à cette même conclusion, mais un peu plus loin (29), il teste avec la même méthode la significativité de l'écart entre les deux possibilités des baptêmes de garçons à Londres et Paris. Puisque la probabilité que la première dépasse la seconde est de 1 - 1/328 269, il peut écrire qu'il y a donc 328 268 à parier contre un qu'à Londres la possibilité des baptêmes de garçons est plus grande qu'à Paris. Laplace poursuit :

> *Ce résultat parait être une loi générale, du moins en Europe, et si parfois dans quelques petites villes, où l'on a observé un nombre peu considérable de naissances, la nature semble s'en écarter, il y a tout lieu de croire que cet écart n'est qu'apparent, et, qu'à la longue les naissances observées dans ces villes offriraient, en se multipliant, un résultat semblable à celui des grandes villes.*

Laplace illustre ce propos par le cas du bourg de Vitteaux qui a vu naître en cinq années 203 garçons et 212 filles : la probabilité que les naissances de garçons soient

[12] Je trouve prob (7037 < N(7200, 59) < 7363) = 0,994.
[13] Je trouve bien moins avec prob(Bin(82 ; 0,004)) ???

encore d'une possibilité plus grandes que celles des filles ($x > 1/2$) est encore de 1/3, et la probabilité complémentaire de 2/3 lui semble :

> *bien trop faible pour balancer l'analogie qui nous porte à penser qu'à Vitteaux, comme dans toutes les villes où l'on a observé un nombre considérable de naissances, la possibilité de naissance des garçons l'emporte sur celle de naissance des filles.*

L'exemple analogue de la commune de Corcelle-le-Grignon a été évoqué par Laplace dans *l'Essai philosophique*. Il y serait né en cinq ans 983 garçons et 1026 filles. Mais dit Laplace :

> *ces nombres n'indiquent une plus grande possibilité dans les naissances des filles qu'avec une probabilité 9/10 ; et cette probabilité, plus petite que celle de ne pas amener croix quatre fois de suite au jeu de croix ou pile, n'est pas suffisante pour rechercher la cause de cette anomalie, qui selon toute vraisemblance disparaîtrait si l'on suivait pendant un siècle, les naissances dans cette commune.*

Dans un article rédigé en 1986 et réédité récemment par Thierry Martin, Georges-Th. Guilbaud a repris les données de Buffon et relevé les quelques erreurs dans la transcription qu'en a faite Laplace. L'erreur la plus importante est d'avoir pris en lieu et place des chiffres bien plus faibles de cette commune (36 garçons et 37 filles), ceux de la somme des naissances des 10 communes en tête de la liste de Buffon, et d'avoir arrondi à 9/10 la probabilité que Guilbaud calcule à 0,832. Mais tout cela ne changera pas nos conclusions :

1. Laplace, dans ce 3e exemple, procède à une évaluation directe de la « possibilité » x, c'est-à-dire à un jugement de probabilité sur cette possibilité, sous la forme d'une probabilité *a posteriori* que cette possibilité soit dans un certain intervalle. La procédure des tests laplaciens de ce chapitre VI est bien différente de celle du chapitre précédent. Il n'y a plus comme dans les deux premiers un calcul d'une probabilité critique sous l'hypothèse nulle d'un résultat au moins égal à celui observé. À la limite il n'y a plus de test à proprement parler au sens de l'acceptation ou du rejet d'une hypothèse, mais probabilisation directe par le seul principe bayésien de cette hypothèse $\{x < 1/2\}$.
2. Il est difficile de savoir comment Laplace utilise cette probabilité, et en particulier à partir de quel seuil il considère qu'une hypothèse est vraisemblable. Une probabilité de 90 % pour la supériorité locale des naissances féminines est encore trop petite pour remettre en question la supériorité générale des naissances de garçons. Il est intéressant à cet égard de comparer sur les différentes donnée de sex-ratio cette méthode « bayésienne » et la méthode « fisherienne » de son chapitre précédent, pour laquelle on se contenterait de calculer une probabilité (critique) d'avoir un résultat aussi éloigné de la valeur espérée que celui qui est observé (sous l'hypothèse d'une égalité de possibilité des naissances) :

   L'approche « fisherienne » nous fait rejeter l'hypothèse $x > 1/2$ dans les 3 derniers cas les plus défavorables du fichier de 99 paroisses de Buffon : soit la seule paroisse de Vitteaux, les 10 paroisses qui précèdent Corcelles, et des 42 qui ont moins de garçons que de filles. L'approche « bayésienne », selon les calculs de Laplace ou de Guilbaud (*), ne rejette jamais cette hypothèse bien que dans ces 3 cas $p(x < 1/2)$ soit bien supérieur à $p(x > 1/2)$. Si l'on suit les interprétations de Laplace pour Vitteaux et le faux Corcelles, les probabilités de 0,67 et 0,83 sont jugées bien trop faibles pour invalider $x > 1/2$. Et Guilbaud pense qu'aux yeux de Laplace « 0,0058 n'est pas un

nombre assez petit (et 0,9942 pas encore assez grand) pour faire sérieusement douter de la supériorité des garçons » sur la base des 42 paroisses favorables aux filles.

Sex-ratio			H0: x=1/2				
			H1 : x>1/2				
	Garçons	Filles	Proba critique	Déci	Proba *a posteriori*		Déci
			p(G/N>Obs)		p(x>1/2)	p(x<1/2)	
Paris	393 386	377 555	0	H1	~1	1E-72	H1
Londres	737 629	698 958	0	H1	~1		H1
Naples	782 352	746 821	0	H1	~1		H1
99 paroisses	4 285	3 880	3,49488E-06	H1	~1	3,8462E-06	H1*
Faux Corcelles	983	1 026	0,8256	H0	0,168	0,832	H1*
Vitteaux	203	212	0,6527	H0	0,33	0,67	H1
42 paroisses	1 690	1 840	0,9939	H0	0,0058	0,9942	H1*

3. En fait Laplace ne prend pas ici une décision propre à chaque exemple. Il n'y a pas de règle de décision qui s'appliquerait à chaque fois : les probabilités « indiquent », elles ne prouvent pas. Elles constituent des raisons de croire qui se combinent pour former un jugement global. La probabilité de ($x < 12$) dans les cas de Corcelles ou Vitteaux n'est pas assez grande pour « balancer » comme il le dit l'énorme probabilité en faveur des garçons qu'il a calculée pour des grandes villes.
4. Disant cela Laplace tient compte du fait que l'échantillon des valeurs observées n'est pas dans cet exemple le résultat d'un tirage au hasard mais celui d'un choix systématique de cas défavorables. Il y aurait donc une certaine imposture à juger de la question du sex-ratio sur la base d'une seule observation sélectionnée à dessein et aussi extrême. On retrouve ici la critique de Cournot des « coupes arbitraires » et celle de Fisher au sujet des tests appliqués sans procédure de « randomization ».

Cournot (Chap. XIII, p. 195) discute de nombreuses données de sex-ratio en différents lieux et pour différentes sous-populations, cherchant à identifier les facteurs de variation. Il ne retient guère que l'hypothèse de Sadler et Hofacker sur l'influence de l'écart d'âge entre le père et la mère, et rappelle que Laplace en a réfuté une version stratégique, celle d'un joueur :

> *qui prétendrait changer les chances du jeu, et le résultat d'un grand nombre d'épreuves, en adoptant le système de se retirer du jeu après le gain d'une ou plusieurs parties.*

Mais il ne procède à aucun calcul sur ce sujet[14]. Par contre, il propose (n° 170) un test de stabilité par comparaison du sex-ratio entre la période 1817-1839 et l'année 1840 avec des formules établies au début du chapitre VIII consacré aux probabilités *a posteriori*[15]. Sur les 23 premières années, on a observé pour la France entière 22 263 015 naissances dont 11 473 437 masculines, tandis que l'année 1840 a vu 952 318 naissances dont 489 374 masculines. L'écart observé entre les sex-ratios est $F - F' = 0{,}51541 - 0{,}51388 = 0{,}00153 = \delta$. Est-il significatif d'un écart entre les possibilités x et x' des naissances de garçons dans chaque période ? La probabilité qu'un tel écart survienne sous l'hypothèse $x = x'$ est $P = \text{prob}(-\delta < F - F' < \delta \,/\, x = x') = 0{,}1834$. $1 - P = 0{,}8166$ serait alors la probabilité critique d'un test de Fisher, et amènerait à décider que l'écart δ n'est pas significatif et que l'on ne peut rejeter l'hypothèse $x = x'$.

[14] Sujet sur lequel Eric Brian [2005] vient d'apporter un éclairage nouveau en partant des travaux de Halbwachs.
[15] § 99, p. 118 et 108, p. 127.

Mais la démarche de Cournot n'est pas celle-ci. Elle est franchement bayésienne. Il calcule la probabilité *a posteriori* Π que la « chance » en faveur des garçons dans le premier échantillon surpasse cette même chance dans le second ($x > x'$) pour un écart observé δ, qu'il interprète comme une probabilité que cet écart δ « ne soit pas imputable aux anomalies du hasard ». Cette probabilité se déduit de P par la relation $\Pi = (1 + P)/2$[16] = 0,5917. Cette valeur lui apparaît trop faible pour indiquer un changement d'urne, et donc « sans signification possible, quoique l'année 1840 soit, après 1830 et 1828, celle des années de la période où le rapport atteint sa plus petite valeur ». La décision est finalement la même dans ce cas mais la différence des démarches est bien mise en évidence.

4. DE LA MULTIPLICATION DES COUPES

C'est à l'occasion de ce calcul que Cournot développe une inquiétude fondamentale sur le caractère arbitraire des coupes possibles :

> *Rien ne limite, dit-il, le nombre des systèmes binaires de caractères opposés, d'après lesquels on peut faire autant de coupes dans la série totale* (…) *Si l'expérimentateur n'a été induit que par le résultat même du dépouillement [et non a priori] à considérer un système [une coupe] de préférence aux autres* (…) *l'anomalie observée pour ce système pris entre une foule d'autres, peut très bien être fortuite ; il est même très probable qu'on finira par tomber fortuitement sur une telle anomalie, à force de multiplier les coupes et les essais*[17].

S'occupant au chapitre suivant de la statistique, Cournot assigne à cette discipline une plus noble mission que la seule compilation des faits ; elle doit nous faire connaître la chance d'arrivée d'un événement et nous faire comprendre les raisons d'une différence significative. Et c'est ici qu'il revient sur la même remarque à propos des coupes arbitraires :

> *tandis que le nombre des coupes augmente ainsi sans limite, il est a priori de plus en plus probable que par le seul effet du hasard, l'une des coupes au moins [par exemple l'un au moins des départements opposé à tous les autres dit-il plus loin] offrira pour le rapport du nombre des naissances masculines au nombre total des naissances, dans les deux catégories opposées, des valeurs sensiblement différentes* (…) *La probabilité que nous avons appelée Π et qui correspond à l'écart δ perdra pour la personne étrangère aux essais qui ont manifesté cet écart toute consistance objective*[18].

Cournot poursuit en faisant remarquer que, dès lors, le jugement probable sur la significativité d'un écart observé devient la combinaison de deux éléments forts différents : l'un qui est bien le calcul de probabilités mathématiques P relatives à chaque coupe, l'autre qui est un jugement *a priori* sur l'intérêt de telle ou telle coupe, et qui n'est pas de nature mathématique. Il y a ici comme un péché originel de la statistique mathématique, qui la déconsidère à ses yeux suffisamment pour lui empêcher de jouer un rôle majeur dans les sciences sociales. Le résultat d'un test de signification n'a pas du tout la même interprétation selon qu'il a été produit sur un cas [un département] pris

[16] Formule apparemment simple qui n'est pas évidente à établir (Laplace II-29, dit Bru).
[17] Cournot, 1843, § 102, Chap. VIII, p. 121.
[18] Cournot, 1843, § 111, Chap IX, p. 130.

au hasard ou sur celui qui a été retenu au bout d'une longue séquence d'essais aboutissant à sa sélection.

Or ce renforcement arbitraire de la probabilité *a posteriori* d'un écart significatif au travers de la pratique répétée des tests de comparaison est précisément ce que dénoncent les économètres anglais de l'école de Hendry dans les années 1980[19] sous le nom de *Data Mining :* utiliser le même jeu de données observées (et non expérimentales) pour tester systématiquement plusieurs modèles linéaires concurrents jusqu'à valiser un de ces modèles (= un ensembles de variables exogènes « significatives ») conduit à sous estimer gravement le niveau de significativité réel, et donc à trouver des effets significatifs trop facilement : si, par exemple, 10 variables orthogonales choisies au hasard (H0) sont testées successivement comme causes possibles de y au niveau α = 5 %, la sélection de deux d'entre elles comme significatives constituera un modèle accepté avec un risque d'erreur reel α^* = 22,6 %. Le biais de Lovell, dû au *data mining*, est exactement ce que Cournot a stigmatisé 140 ans plus tôt ! Il disqualifie finalement la pratique des tests d'hypothèse, fisheriens ou bayésiens, chaque fois qu'ils sont pratiqués sans randomisation, or c'est le cas de toutes les données d'observation qui ne résultent pas d'un plan d'expérience, dans lequel le hasard neutralise les coupes « remarquables ».

Une autre façon de contrevenir à l'exigence de la randomisation, et d'obtenir des résultats significatifs, sans trop se fatiguer dans une multitude d'essais systématiques, est de choisir à dessein une situation totalement exceptionnelle, la pire de toutes. C'est un tel choix que Laplace a institué avec la commune de Corcelles et d'autres exemples du même type.

C'est ce qui amène Cournot à distinguer au chapitre XVII la probabilité philosophique de la probabilité mathématique et à critiquer la notion d'événement remarquable :

> *En admettant qu'on connaît le nombre des événements remarquables et celui d*es événements non remarquables, *M. Poisson assigne, d'après les règles communément admises dans la théorie des probabilités a posteriori, la probabilité que l'apparition d'un événement remarquable n'est point l'effet du hasard. Mais le défaut du raisonnement consiste à supposer d'une part qu'on peut tracer une ligne de démarcation entre les événements remarquables et les événements non remarquables ; d'autre part que les événements réputés remarquables sont remarquables au même degré* (…) *La combinaison de caractères d'imprimerie dans laquelle un voyageur reconnaîtrait quelques mots parlés par une peuplade sauvage doit elle être comptée parmi les combinaisons remarquables, et serait elle pour nous aussi remarquable que celle qui nous offrirait des mots usuels de notre langue ?* (…) *Toute application des notions de la probabilité mathématique au jugement de probabilité dont il est question serait fausse et illusoire.*

C'est encore ce que Gini a magnifiquement illustré par sa parabole du gendarme :

> *En jouant à la loterie, vous avez gagné un quaterne et vous allez retirer votre lot. Vous trouvez au guichet un agent de police qui vous apostrophe ainsi : « la probabilité que vos numéros sortent par l'effet du hasard est de 1/511038. Il y a donc 511037 chances contre une que votre succès soit l'effet d'une supercherie. C'en est assez pour que je vous arrête ».*

[19] Cf. les travaux de Leamer [1983] mais plus encore ceux de Lovell [1983], Charemza et Deadman [1997].

La probabilité d'un résultat fortuit donné *a posteriori* est toujours très petite et conduit toujours à rejeter une hypothèse de hasard, donc à admettre une raison particulière. Autre exemple : je produis le nombre 432567 et prétend l'avoir obtenu au hasard (H0) ; or sous cette hypothèse la probabilité d'obtenir ce nombre de 6 chiffres est de 1/100000 ; donc je rejette cette hypothèse et conclus à une supercherie.

Joseph Bertrand, secrétaire de l'Académie des Sciences, et auteur du principal traité de Calcul des probabilités de la fin du XIX[e] siècle[20] , n'a pas ménagé ses sarcasmes à propos des dérapages du raisonnement probabiliste sous la plume de ses prédécesseurs, que ce soient Condorcet, dont il rejette l'assimilation d'une raison de croire subjective à une probabilité objective, ou Laplace dont il rejette l'assimilation de toute répétition d'épreuves à un tirage dans une urne de composition fixe :

> *L'application du calcul des probabilités à l'étude des erreurs d'observation repose sur une fiction dont il ne faudrait pas faire une réalité. Les erreurs sont supposées tirées au sort dans une urne dont la composition est définie par la loi de probabilité acceptée. Si l'observateur est malade, si l'on a dérangé les fils de sa lunette, exposé à l'humidité les poids de sa balance, changé son thermomètre habituel, il fait, ce jour-là, le tirage dans une autre urne, les résultats échappent à toute théorie.*[21]

Bertrand brocarde allègrement les résultats de Laplace sur la masse de Jupiter :

> *Les principes du calcul des chances permettaient de parier suivant Laplace 999 308 contre 1 que l'erreur n'est pas la centième partie de la valeur trouvée [une masse de 1/1070 de celle du Soleil]* […] *On aurait eu tort de risquer dix sous ; on les aurait perdus, les perturbations de Junon l'ont prouvé*[22].

Mais foin des railleries. Quelle est la pratique « testuelle » de Joseph Bertrand ? On trouve d'abord plusieurs exemples – sur le sex-ratio, sur les résultats d'un tir – où l'hypothèse à tester est celle d'une distribution binomiale, et pour lesquels aucun calcul de probabilité n'est fait : il suffit de vérifier que le rapport des moments empiriques σ^2/μ^2 est à peu près égal à $\pi/2$, ou encore de comparer les fréquences de classes de valeurs aux probabilités théoriques, pour valider cette hypothèse. Au n° 216, p. 277, on trouve finalement un test de proportion par la probabilité des causes. Observant – ou plutôt supposant – qu'une pièce lancée un million de fois a donné 500 391 fois *Face*, Bertrand évalue facilement le rapport entre la probabilité de cet événement sous l'hypothèse extrême p(Face) = 0,499609 et la probabilité du même événement sous l'hypothèse la plus plausible p(Face) = 0,500 391. Ce rapport vaut 0,294332. Or, dit-il :

> *le rapport des probabilités de deux causes également probables a priori est exactement celui des probabilités observées. La seconde valeur de p n'est donc pas quatre fois plus probable que la première, que nous avons choisie absolument contraire.*

C'est une stricte application du 6[e] Principe (bayésien) de Laplace, et cela doit nous convaincre que l'écart à 0,5 est ici bien peu significatif. Mais ce n'est pas dit comme cela parce que ce n'est pas un test de significativité.

---

[20] Bertrand [1889] ; cf. aussi l'article de B. Bru sur ce traité dans le *Journal électronique d'histoire des probabilités et de la Statistique,* 2-2, décembre 2006.
[21] Bertrand, 1889, p. 222.
[22] Bertrand, 1889, p. XXXIX.

> *le rapport des probabilités de deux causes également probables a priori est exactement celui des probabilités observées. La seconde valeur de p n'est donc pas quatre fois plus probable que la première, que nous avons choisie absolument contraire.*

C'est une stricte application du 6e Principe (bayésien) de Laplace, et cela doit nous convaincre que l'écart à 0,5 est ici bien peu significatif. Mais ce n'est pas dit comme cela parce que ce n'est pas un test de significativité.

Concluons. Laplace a pratiqué, après plusieurs érudits du XVIIIe comme le célèbre Arbuthnot, les tests de signification de ce que je n'affirmerais pas être une hypothèse nulle à la Fisher mais qui s'en écarte très peu. « Concevons que la cause n'existe point » disait-il et calculons la probabilité qu'un résultat au plus aussi extrême se produise sous cette condition assimilée au hasard. Mais quelle forme du hasard ? Nous avons vu qu'il y en avait plusieurs de disponibles. Plus cette probabilité s'approche de l'unité, plus elle « indique » l'existence d'une cause. Dans cette méthode on ne la mesure pas directement. Mais l'interprétation dérape facilement vers un discours de corroboration directe d'une loi de cause à effet, loi de la nature ou grande Loi du Divin Dessein. L'hypothèse alternative non formulée est l'objet de toutes les convoitises, de toutes les audaces interprétatives. Neyman est déjà là. Cette corroboration peut-elle s'exprimer par une probabilité ? C'est tentant. Mais cela ne peut se faire que par un lent glissement du plaisir… de mesurer directement la vraisemblance de cette hypothèse. Ce que finalement Laplace semble préférer avec l'évaluation directe – par la voie de son théorème bayésien de proportionnalité – d'une probabilité *a posteriori* de la possibilité de cette cause. Mais bien sûr la probabilité d'une possibilité c'est une chose compliquée, qui mélange le calcul et la philosophie. Surtout si le résultat est choisi délibérément. Cournot l'a bien vu. Bertrand et Gini aussi. L'hybride est bien là, déjà, avec en plus une interrogation fondamentale sur la validité du test unique sur un seul résultat : dans quelle série faut-il le replacer pour que ce test prenne une valeur générale ? La réponse en terme de randomisation que donnera Fisher est déjà pressentie mais ne vaut que pour des situations expérimentales et des résultats produits par une certaine forme de hasard. Pour des données choisies, isolées par l'observateur arbitrairement, dans un flux d'informations non contrôlé, quel peut-être le sens d'un test d'hypothèse ? On n'a pas fini de se le demander.

BIBLIOGRAPHIE

ARBUTHNOT J., "An argument for Divine Providence, taken from the constant regularity observed in the births of both sexes", *Phil. Trans. of the Royal Society* 27, 1710, p. 186-190.

ARMATTE M., « La construction des notions d'estimation et de vraisemblance chez Ronald A. Fisher », *Journal de la Société de Statistique de Paris*, tome 129, n° 1-2, 1988.

ARMATTE M., « Sur les tests d'hypothèse : la véritable nature d'une méthodologie hybride », *Journal de la Société Française de Statistique*, Vol. 145, n° 5, 2005, p. 27-36.

BARNETT V., *Comparative Statistical Inference*, John Wiley & Sons, 1982.

BERTRAND J., *Calcul des probabilités*, 3e édition, New York, Chelsea, 1971, XLIX + 317 p. [1ère édition, 1889 ; 2e édition, 1907].

BRIAN E., JAISSON M., « Le critère du fait social. Retour sur le taux de masculinité à la naissance », in Maurice Halbwachs et al., *Le Point de vue du nombre (1936),* éditions de l'INED, 2005, p. 169-197.

BRU B., « Estimations laplaciennes », *Estimation et Sondages*, J. Mairesse, (éd.) *Economica*, Paris, 1988, p. 7-48.

BRU B., « Remarques sur l'article de Denis », *Journal de la Société Française de Statistique*, Vol. 145, n °5, 2005, p. 37-38.

BRU B., 2006, « Les leçons de calcul des probabilités de Joseph Bertrand », *Journal électronique d'histoire des probabilités et de la statistique,* 2-2, Décembre 2006.

CAPEL R., MONOD D., MÜLLER J.-P., « De l'usage perverti des tests inférentiels en sciences humaines », *Genèse* 26, 1997, p. 123-142

CHAREMZA W. et DEADMAN D.F., *New Directions in Econometric Practice*, Edward Elgar, 1997.

COURNOT A. A., (1843), « Exposition de la théorie des chances et des probabilités », *Oeuvres Complètes*, B. Bru (éd.), Paris, Vrin, 1984.

FISHER R.A., *Statistical Methods for Research Workers*, 1925. [14^{e} édition, Edinburgh, Oliver and Boyd, 1970. Traduction française I. Bertrand sur la 10^{e} édition, Paris, 1947].

FISHER R. A., "The logic of inductive inference", *Journal of the Royal Statistical Society* 98, 1935, p. 39-82, [avec discussion, Collected Papers 124].

FISHER R. A., *The Design of Experiment*, Edinburgh, Oliver and Boyd, 1935.

FISHER-BOX J., *R.A. Fisher, the life of a scientist*, New York, John Wiley & sons, 1978.

GIGERENZER G., MURRAY D. J., *Cognition as Intuitive Statistics*, Hillsdale / London, Lawrence Erlbaum associates, 1987.

GIGERENZER G., "The Superego, the Ego, and the Id in Statistical Reasoning", in Keren G. et Lewis C., (eds), *A handbook for data analysis in the behavorial sciences : Methodological Issues*, Londres, Lawrence Erlbaum Ass. Ed., 1993.

GILLISPIE C., "Laplace", *Dictionary of Scientific Biography*, Supplément, 1978, p. 273-403.

GUILBAUD G.-TH., « La ténébreuse affaire de Corcelles », *Arithmétique politique dans la France du XVIII^{e} siècle*, T. Martin (dir.), 2003, p. 249-258.

GUILFORD J.P., *Fundamental Statistics in Psychology and Education*, New York, McGraw-Hill, 1942.

HAAVELMO T., "The probability Approach in Econometrics", Supplément à *Econometrica* 12, 1944, 115 p.

HUBERTY C.J., "Historical origins of testing practices : the treatment of Fisher versus Neyman-Pearson views in textbooks", *Journal of Experimental Education* 61 (4), 1993, p. 317-333.

LAPLACE P. S., « Théorie analytique des probabilités », *Oeuvres Complètes,* VII, 1886. [1ère édition, 1812, 2^{e} édition, 1814 avec préface, 3^{e} édition, 1820 avec trois suppléments. Version numérique de la BNF (Gallica) sur la 3^{e} édition].

LEAMER E. E., "Let's take the con out of econometrics", *American Economic Review* 23, 1983, p. 31-43.

LECOUTRE M., LECOUTRE M. P., POITEVINEAU J., "Uses, Abuses and Misuses of Significance Tests in the Scientific Community: Won't the Bayesian Choice be Unavoidable ?", *International Statistical Review* 69, 2001, p. 399-417.

LOVELL M.C., "Data mining", *Review of Economics and Statistics* 65, 1983, p. 1-12.

MEUSNIER N., *Dr Arbuthnot et Mr Hidden*, Cahiers du CAMS, n° 162, Paris, EHESS, 1999.

POITEVINEAU J., « L'usage des tests statistiques par les chercheurs en psychologie : aspects normatif, descriptif et prescriptif », *Math et Sciences humaines* 167, 2004, p. 5-25.

SHEYNIN O. B., "Newton and the classical Theory of Probabiliy", *Arch. Hist. Exact Sci.* 7, 1971, p. 217-243.

SHEYNIN O. B., "Laplace's Work on Probability", *Arch. Hist. Exact Sci.* 16, 1976, p. 137-187.

SPANOS A., "Probability Theory and Statistical Inference", *Econometric Modelling with Observational Data*, Cambridge, Cambridge University Press, 1999.