A propos du texte de Jean Lejeune, Denis Lanier et Didier Trotoux

Claudine Schwartz

Le texte de J.Lejeune, D. Lanier et D. Trotoux (nous le noterons [LLT]) établit des ponts entre les idées de Jules Gavarret et certaines de celles qui sont à l'œuvre actuellement en statistique. Ce texte est conçu pour les enseignants de sciences des lycées et collèges. Passer directement de J. Gavarret au XXIème siècle est efficace pour de nombreux enseignants qui ont à se former rapidement à la pensée statistique. Les réflexions de Gavarret sur la nécessité d'avoir « beaucoup » de données, la prise en compte systématique de la fluctuation d'échantillonnage sont des éléments particulièrement intéressants. Une lecture active de [LLT] permettra au lecteur néophyte d'acquérir de l'expérience statistique sans pour autant risquer de se perdre dans les méandres d'un travail d'histoire des sciences.

Les commentaires, compléments, questions et critiques ci-dessous du document LLT restent dans le cadre de la formation (et de l'auto-formation) des enseignants des lycées et collèges[1].

1- Notations

Les notations dans [LLT] sont celles de Gavarret et de son époque. On appelle ainsi μ la taille d'un échantillon. C'est un peu comme si on parlait du vieux français, c'est parfois difficile à lire. Dans le contexte de la statistique contemporaine, μ est plutôt réservé à une espérance (ou moyenne théorique). Chaque époque et chaque champ disciplinaire a ses notations dont la cohérence décharge la pensée, permettant un accès plus rapide aux problématiques. Pour voir l'importance des notations, il n'est qu'à imaginer un texte où μ serait l'écart-type d'une loi de Gauss, et σ son espérance.

Dans Gavarret et [LLT], on trouve des expressions telles que $\sqrt{\frac{2mn}{\mu^3}}$: certains lecteurs n'auront-ils pas du mal à y voir tout de suite l'expression plus facile à décoder : $\sqrt{2\frac{f(1-f)}{n}}$?

Dans le texte de Gavarret (voir ci-dessous), les pourcentages sont donnés, avec des notations qui ne seraient plus employées aujourd'hui. En plus du « pour cent » ou « pour mille » on utilise le « pour 10 000, pour 100 0000 ». Ainsi, Gavarret écrit qu'entre 1825 et 1830, il y a eu 6094 condamnations sur 10 000 accusés, contre 5388 en 1831, soit une différence de 706 sur 10 000.

[1] On trouvera notamment peu de références explicites et presque uniquement des références téléchargeables gratuitement sur Internet.

1er *Exemple*. Les relevés publiés par l'administration, relativement aux jugements du jury en France, depuis 1825 jusqu'à 1831 inclusivement, abstraction faite des procès politiques, nous fournissent les résultats suivants :

DE 1825 A 1830 INCLUSIVEMENT.	1831
25 777 nombre des condamnés.	4 098 nombre des condamnés.
16 523 nombre des acquittés.	3 508 nombre des acquittés.
42 300 nombre des accusés.	7 606 nombre des accusés.
Ce relevé nous conduit à la proposition suivante : Sous l'empire de la législation criminelle qui a duré de 1825 à 1830, il y avait en France :	Ce nouveau relevé nous conduit à la proposition suivante : Sous l'empire de la législation criminelle de 1831, il y avait en France :
6 094 condamnés sur 10 000 accusés.	5 388 condamnés sur 10 000 accusé

La différence entre les résultats fournis par ces deux relevés s'élève à

706 sur 10 000 accusés.

On peut penser que l'idée de cette écriture est liée au concept d'une probabilité estimée sous la forme de « k chances sur n », issu du « nombre de cas favorables sur nombre de cas possibles » des situations d'équiprobabilité. On peut trouver, jusqu'au 20ème siècle inclus, dans les manuels de mathématiques, la trace de cette réticence à considérer des probabilités qui ne soient pas des nombres rationnels[2]. Réticence qui témoigne aussi de la difficulté à concevoir que la limite d'une suite de rationnels (les fréquences observées sur des échantillons de taille croissante) n'est pas forcément un nombre rationnel[3].

2- **Chiffres et idéologie**

L'exemple 2, page 16, traite des pourcentages de naissances de garçons (taux τ de masculinité[4]) parmi les naissances d'enfants légitimes (τ =51,7%), et parmi les naissances d'enfants illégitimes (τ =51,0%). Cette différence de 0,7% est significative selon les critères de Gavarret (et selon les critères actuels des tests statistiques). Il est intéressant de voir qu'une différence de moins de 1% est significative compte tenu du grand nombre des données.

Le plus intéressant de cet exemple donné par Gavarret est l'interprétation qu'il en propose. Elle est recopiée dans [LLT] et reproduite ci-dessous :

---

[2] Une exception classique est celle de l'aiguille de Buffon.

[3] Et pourtant…tout nombre réel peut être obtenu comme limite de nombres rationnels, c'est-à-dire de nombres qui admettent une écriture sous forme de quotient de deux nombres entiers.

[4] Le taux de masculinité est de l'ordre de 0,5 ; le sex-ratio, quotient du nombre de naissances de garçons sur le nombre de naissances de filles et est de l'ordre de 1.

> D'après ce que nous avons dit jusqu'ici, nous devons en conclure que les enfants légitimes ont plus de chance que les enfants illégitimes de naître garçons. Cette proposition, loin de surprendre, pouvait être en quelque sorte prévue *a priori*; car des documents authentiques prouvent que partout où existe la *monogamie*, les enfants mâles naissent plus nombreux que ceux du sexe féminin, tandis que le contraire a lieu dans les pays où existe la *polygamie*.

Le texte de Gavarret ne dit pas que la polygamie est la cause d'un taux de masculinité faible, mais simplement qu'on a observé des concomitances entre taux de masculinité et régime matrimonial. Et cependant cette concomitance semble faire fonction de cause dans le texte de Gavarret : les enfants illégitimes ont « plus de chances de naitre fille » que les enfants légitimes dans la mesure où la situation légitime-illégitime serait assimilée à celle de monogamie-polygamie. A tel point que la significativité de la différence observée est in fine jugée par Gavarret prévisible a priori.

La quasi constance du sex-ratio et ses variations faibles mais significatives constitue un sujet d'étude[5] toujours d'actualité et qui a fait l'objet en Europe, notamment au XVIII siècle, de débats nombreux et de controverses passionnées. Montesquieu, dans « l'esprit des lois »(1748), partant d'un refus d'admettre une explication morale ou religieuse (un ordre divin, la Providence) pour la concomitance « polygamie – sex- ratio faible », argumente en sens inverse de Gavarret et bien d'autres[6]. Selon lui, le sex-ratio faible dans les pays chauds conduit à un excès de femmes qui rend la polygamie nécessaire. Autrement dit, le sex-ratio serait la cause et non l'effet de la polygamie. En termes modernes, on peut aussi dire que selon Montesquieu le climat serait une variable d'influence sur le sex-ratio, le statut matrimonial étant alors une variable de confusion.

Pour des grands groupes de population, on a calculé sur de très nombreuses données des valeurs du sex-ratio et on les a utilisées comme valeurs de références à l'aune desquelles on essaye d'expliquer les fluctuations qui dépassent largement ce qu'on peut attendre de la fluctuation d'échantillonnage. La valeur de référence pour le sex-ratio en France est 1,05 (souvent noté « 105 garçons pour 100 filles »), soit un rapport de masculinité de 51,2% (105 garçons pour 205 naissances). Il est étonnant que le taux de masculinité [7] parmi les naissances légitimes soit supérieur au même taux pour les naissances illégitimes !Si ce n'était le cas que pour une période données dans un pays donné, la question aurait été vite enterrée. En effet, si on regarde en effet le taux de masculinité sous tous les angles, en fonction de très nombreux critères, on trouvera logiquement que toutes les époques offrent quelques résultats « significativement exceptionnels » dont on ne saura jamais s'ils témoignent d'une fluctuation rare ou d'une cause autre. Mais le phénomène a ici été observé en France et dans d'autres pays, pendant longtemps et a donc fait couler beaucoup d'encre. Parmi les explications (raisonnables …) avancées, il y aurait un taux de masculinité des enfants mort-nés(en fin de grossesse) plus élevé que celui des naissances vivantes (on ne sait pas du tout pourquoi il en est ainsi) associé à un taux de mort-nés plus élevé chez les enfants illégitimes (conditions de grossesse plus difficiles). La différence d'âge des parents aurait aussi (selon M. Halbwachs mais cela n'a pas été confirmé) une influence sur le taux de masculinité et sa répartition pourrait être différente entre les naissances légitimes et les naissances illégitimes. Enfin, à la naissance, les garçons pourraient avoir été plus facilement reconnus (légitimés) que les filles.

Le commentaire expéditif que fait Gavarret montre que la manipulation des chiffres peut être mise au service de l'idéologie la plus pure : si c'est ici immédiatement visible, cela l'est parfois moins dans d'autres situations. Certains diront que cela discrédite l'usage de la statistique. Mais cet exemple n'illustre-t-il pas au contraire la nécessité absolue du débat et de l'argumentation pour commenter et « expliquer » des données ?

---

[5] Voir *Le sexisme de la première heure* : Hasard et sociologie. Eric Brian, Marie Jaisson

[6] Le titre originel du chapitre qui traite de ce sujet était « comment la Loi de Polygamie est une affaire de calcul ».

[7] Un synthèse de certains résultats est dans un numéro de la revue « population et société » publiée par l'INED http://www.ined.fr/fichier/t_publication/516/publi_pdf1_pop_et_soc_francais_61.pdf

Aujourd'hui, les articles des revues de médecine doivent contenir au moins trois paragraphes nettement séparés : le premier d'entre eux est un paragraphe « matériel et méthode » qui décrit entre autre la population ciblée, le mode d'échantillonnage, les techniques statistiques qui seront mises en œuvre sur les données recueillies. Le deuxième paragraphe imposé est le paragraphe « résultats » qui fournit une description des données recueillies et des résultats issus directement de l'emploi des méthodes statistiques mises en œuvre pour les analyser. Un paragraphe séparé est souvent intitulé « discussion » ; on y débat des résultats trouvés, en les mettant notamment en regard d'autres résultats, qu'ils soient quantitatifs ou qualitatifs et sans écarter des interprétations nouvelles et l'explicitation de nouvelles hypothèses ou conjectures. On sépare ainsi ce qui relève de la technique et ce qui relève de la pensée hypothético-déductive classique.

3- Intervalle de confiance et test

Dans [LLT], on lit page 62 qu'il est *erroné* de regarder si les intervalles de confiance de deux probabilités ont une intersection vide ou non pour décider que la différence observée entre deux fréquences est significative ou non. Page 94, les auteurs proposent même de « tordre le cou » à une telle pratique. La raison à cela, selon les auteurs, est qu'elle n'a pas de justification mathématique. Mais cette méthode par intervalles de confiance (notons la MIC) est naturelle, et aussi « justifiable mathématiquement » que la méthode classique par test statistique (notons la MT). Ces deux méthodes sont des choix de stratégie dont on peut comparer les qualités et les limites, mais dire qu'un choix est *juste* et l'autre *faux* ne me parait…pas juste.

Détaillons un peu : On cherche à définir si l'écart absolu entre deux fréquences f_n et $f'_{n'}$ peut être assimilé à une fluctuation d'échantillonnage ou non. Avec MIC, on construit pour chacune des deux séries de données un intervalle de confiance au niveau de confiance 1-α. Ces intervalles I_α et I'_α donnent pour chacune des séries, l'ensemble des valeurs des probabilités qui sont α–compatibles avec les données, au niveau de confiance 1-α. Il est logique de dire que si ces intervalles sont disjoints, il n'y a pas de modèle α*–compatible* avec les deux séries de données, donc que l'écart des fréquences est significatif (pour cette méthode). Et si les intervalles ne sont pas disjoints, on conclue que la différence n'est pas, disons significative pour MIC.

Il y a un lien entre la α-significativite avec MIC et la α -significativité définie par la méthode des tests (notons MT) :

Si la différence est jugée α-significative pour MIC, elle le sera aussi pour MT.

Par contre, il se peut que la différence soit jugée α-significative pour MT, et pas pour MIT

Nous allons dorénavant dire que la α-significativité avec MIC est une α-significativité forte.

Ce qui serait mathématiquement faux serait de dire que si on simule deux séries de données à partir d'un même modèle, la probabilité de conclure à une différence α-significativement forte est α. Cette probabilité α' est nettement inférieure à α[8]. La méthode MIC peut-être re-formulée en terme de tests, avec un risque de première espèce plus petit que α.

Prenons l'exemple des naissances illégitimes :

	Enfants légitimes	Enfants illégitimes	
garçons	g=939 641	g'=71 661	1 011 302
filles	877 931	68 905	946 836
total	N=1 817 572	N'=140 566	1 958 138

Les intervalles de confiance du taux de masculinité pour les enfants légitimes et illégitimes sont respectivement :

$$I_\alpha \approx [g\text{-}1/\sqrt{N} \ ; \ g+1/\sqrt{N}\ ) \approx [51{,}6\% \ ; \ 51{,}8\%]$$

[8] Voir par exemple : http://www.statistix.fr/spip.php?article43

$$I'_{\alpha} \approx [g'-1/\sqrt{N'} ; g'+1/\sqrt{N'}) \approx [50,7\% ; 51,2\%].$$

Ces intervalles sont disjoints : la différence est significative, au risque 0,05, au sens fort comme au sens des tests.

Le problème majeur de cette notion de différence fortement significative est qu'elle ne se généralise pas, ne serait-ce qu'à la comparaison globale de plusieurs fréquences.

4- A. Quetelet

A. Quetelet [9], un des précurseurs de la sociologie quantitative, est un contemporain de Gavarret (il était son aine d'une dizaine d'années). Il y a des recoupements forts entre leurs œuvres, des différences et une communauté de pensée ; on peut imaginer qu'ils ont été en contact : est-ce le cas ?

L'exemple 2 (LLT, page 16, Gavarret, page 92) transcrit ci-dessus est à ce propos amusant si on compare ce qu'en dit Gavarret en 1840 et ce qu'en dit Quetelet[10] en 1835 : ils comparent et traitent les mêmes données concernant le nombre des condamnés en France entre 1825 et 1831 (voir ci-dessus page 2). Quetelet, n'ayant pas de formule pour étudier les différences de pourcentages se contente de parler d'un nombre d'acquittements un peu plus grand en 1831 que les années précédentes tandis que Gavarret insiste sur l'effet significatif du changement de la législation.

Cet exemple illustre aussi le fait que l'ouvrage de Gavarret est à visée didactique, celui de Quetelet étant de communiquer les résultats de ses recherches et de ses réflexions. La visée didactique de l'œuvre de Gavarret apparait dans le fait qu'il compare d'une part ce qui se passe avant 1830 (il groupe les données des six années entre 1825 et 1830 inclus) et ce qui se passe en 1831 : il trouve qu'il y a une différence significative et en déduit que la législation doit avoir changé. Or s'il compare ces six années précédents à l'année 1831, c'est qu'il a une bonne raison de le faire (sinon pourquoi juste cette comparaison là)[11] : il sait que la législation a changé. Il a donc l'explication toute faite dans sa poche de la différence significative qu'il constate. Si on replace les calculs que fait Gavarret dans un contexte de pratique de la statistique (et non de son enseignement), il faudrait dire qu'il teste si la nouvelle législation a eu ou non un effet significatif. Après tout, cette nouvelle législation aurait pu ne pas avoir d'effets.

5- Claude Bernard

Bien que le cadre et les limites délibérément choisies dans [LLT] soient compréhensibles, signalons que Claude Bernard a vécu à la même époque que Gavarret[12],

---

[9] http://www.insee.fr/fr/ffc/docs_ffc/cs104a.pdf

[10] Essai de physique sociale, tome 2, page 169 (édition gratuite Googlebook).

[11] C'est un bon réflexe, face à des données dont on ne dit pas le pourquoi de leur étude avant de se poser des questions sur « pourquoi celles la et pas d'autres ».

[12] A cette époque vivait aussi Quetelet, qui a travaillé sur des données semblables (voire les mêmes ?) que celles de Quetelet. On peut se demander pourquoi Quetelet n'est jamais cité par J. Gavarret.

mais était un « opposant à l'usage de la moyenne» et plus généralement à la statistique … pour de bonnes raisons, en partie reprises dans J. Gavarret en particulier :

-le recueil des données qui mènent à des résultats statistiques lui partait fort critiquable.

La qualité des données est un premier souci pour toute étude statistique ; le choix du protocole d'échantillonnage fait l'objet d'un travail en amont, et le recueil lui-même est l'objet d'attention et de contrôles soignés.

-Considérer des groupes de malades et non des individus pose problème (ce qu'on retrouve chez Gavarret) car les malades ne sont pas comparables.

Les définitions de « groupes homogènes de malades » (GHM) font l'objet depuis des années de nombreux travaux de recherche. C'est un chapitre toujours vivant, les maladies évoluent et les moyens de diagnostic aussi.

- L'usage de la moyenne seule, sans mesure de dispersion lui parait à juste titre critiquable. J. Gavarret , de son coté, fait de la variabilité un thème central de son travail.

-Au 19ème siècle, ce qui n'est pas certain est conjectural et donc pas « vraiment scientifique ». J.Gavarret, par ses liens- avec S.D. Poisson semble en avance sur son temps à ce sujet.

La statistique est en voie de reconnaissance comme champ scientifique majeur pour le 21ème siècle. Il reste néanmoins toujours vrai que si elle permet de conjecturer des causes possibles de certains phénomènes physiologiques et d'apporter des éléments de preuve éventuellement très solides, elle ne peut pas à elle seule fonder une preuve définitive.

Ci-dessous quelques extraits de l'Introduction à la médecine expérimentale (1865) :

> *Une autre forme d'application très fréquente des mathématiques à la biologie se trouve dans l'usage des moyennes ou dans l'emploi de la statistique qui, en médecine et en physiologie, conduisent pour ainsi dire nécessairement à l'erreur. Il y a sans doute plusieurs raisons pour cela ; mais le plus grand écueil de l'application du calcul aux phénomènes physiologiques, est toujours au fond leur trop grande complexité qui les empêche d'être définis et suffisamment comparables entre eux.*
>
> ………………….
>
> *Quant à la statistique, on lui fait jouer un grand rôle en médecine, et dès lors elle constitue une question médicale qu'il importe d'examiner ici. La première condition pour employer la statistique, c'est que les faits auxquels on l'applique soient exactement observés afin de pouvoir être ramenés à des unités comparables entre elles. Or, cela ne se rencontre pas le plus souvent en médecine. Tous ceux qui connaissent les hôpitaux savent de quelles causes d'erreurs grossières ont pu être empreintes les déterminations qui servent de base à la statistique. Très souvent le nom des maladies a été donné au hasard, soit parce que le diagnostic était obscur, soit parce que la cause de mort a été inscrite sans y attacher aucune importance scientifique, par un élève qui n'avait pas vu le malade, ou par une personne de l'administration étrangère à la médecine. Sous ce rapport, il ne pourrait y avoir de statistique pathologique valable que celle qui est faite avec des résultats recueillis par le statisticien lui-même.*
>
> ……………………

La statistique ne saurait donc enfanter que les sciences conjecturales ; elle ne produira jamais les sciences actives et expérimentales, c'est-à-dire les sciences qui règlent les phénomènes d'après les lois déterminées. On obtiendra par la statistique une conjecture avec une probabilité plus ou moins grande, sur un cas donné, mais jamais une certitude, jamais une détermination absolue.

Sans doute la statistique peut guider le pronostic du médecin, et en cela elle lui est utile. Je ne repousse donc pas l'emploi de la statistique en médecine, mais je blâme qu'on ne cherche pas à aller au-delà et qu'on croie que la statistique doive servir de base à la science médicale ; c'est cette idée fausse qui porte certains médecins à penser que la médecine ne peut être que conjecturale, et ils en concluent que le médecin est un artiste qui doit suppléer à l'indéterminisme des cas particuliers par son génie, par son tact médical. Ce sont là des idées antiscientifiques contre lesquelles il faut s'élever de toutes ses forces, parce que ce sont elles qui contribuent à faire croupir la médecine dans l'état où elle est depuis si longtemps.

6- Au fil de la lecture de [LLT]

-Page 11 : Il est tentant de mentionner la notion de données censurées à propos des données médicales dont parle Gavarret et d'en faire un précurseur des études de survie, mais ici, justement ce ne semble pas être le cas. Gavarret parle du fait qu'il ne peut pas tenir compte de « tous les malades qui se présentent après l'étude » et que s'il ne traite que peu de cas, la considération de nouveaux cas peut changer beaucoup l'estimation, alors que sur de longues séries, ce ne sera plus le cas. Les séries de données qu'il considère sont toutes complètes : pour chaque malade de la série, on sait s'il est guéri ou non (voir son ouvrage page 70, donnes reproduites dans [LLT] page 11).

-Page 25 : on parle, ([LLT], page 25), d'une formule de Liapounov. Quelle est cette formule ?

-L'exemple 2 de Gavarret discuté dans [LLT] page 30 est intéressant parce-que les tests sont choisis a postériori, une fois qu'on a regardé les données, ce qui est une pratique très peu recommandable. Ainsi, si on lance 1000 fois un dé, il se peut très bien qu'en regardant juste l'écart entre les fréquences extrêmes (face la plus obtenue et la moins obtenue) et sans tenir compte que ce sont les extrêmes, on trouve cet écart significatif

-L'activité 1, page 44, a un défaut majeur : les échantillons avant et après traitement sont appariés. En tirant au hasard deux fois 5 personnes parmi 20, on peut tomber deux fois sur la même personne : les deux échantillons ne seront alors plus indépendants et les calculs faits sont incorrects.