

# JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

I.-J. BIENAYMÉ

## Les grands nombres en statistique

*Journal de la société statistique de Paris*, tome 17 (1876), p. 199-204

[http://www.numdam.org/item?id=JSFS\\_1876\\_\\_17\\_\\_199\\_0](http://www.numdam.org/item?id=JSFS_1876__17__199_0)

© Société de statistique de Paris, 1876, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

## II.



### LES GRANDS NOMBRES EN STATISTIQUE.

Un très-grand nombre de statisticiens, même les plus renommés, considérant les grands nombres comme ayant par eux-mêmes une sorte d'influence sur la constance des résultats en statistique, il nous a paru utile de faire connaître dans ce journal l'opinion de M. Bienaymé, qui fait autorité dans ces matières. On va voir par quelles considérations péremptoires le savant algébriste montre l'inanité de cette prétendue loi.

T. L.

#### Sur un principe que M. Poisson avait cru découvrir et qu'il avait appelé Loi des grands nombres (1).

Il y a environ vingt ans, M. Poisson crut avoir démontré une loi nouvelle destinée à régir toutes les observations statistiques et tous les résultats de même espèce dans toutes les parties des sciences, soit physiques, soit morales et politiques. Il l'appelait *Loi des grands nombres*, et il la fit connaître à l'Académie des sciences (2).

Dès les premières communications de l'illustre géomètre, il me fut facile de reconnaître qu'il n'était point parvenu au but qu'il semblait s'être proposé, et après la publication faite, en 1837, de ses *Recherches sur la probabilité des jugements*, j'indiquai comment il avait dû subir une de ces illusions auxquelles Laplace n'a pas dédaigné de consacrer un chapitre de son *Essai sur le calcul des probabilités*. Mais l'état de la santé de M. Poisson ne permit pas de donner à mes remarques la publicité nécessaire peut-être. La science eut bientôt après le malheur de le perdre prématurément; la négation de la Loi des grands nombres est restée enfouie dans la partie non imprimée des procès-verbaux (3) de la Société philomatique à laquelle j'appartiens; et aujourd'hui ma critique n'est guère connue que des géomètres.

Plusieurs fois on m'a reproché de ne pas l'avoir reproduite, et on a semblé en conclure que peut-être je ne jugeais plus de la même manière un travail que j'ai laissé si longtemps dans l'obscurité, tandis que plusieurs ouvrages ou mémoires publiés dans cet intervalle se sont étayés de la Loi de M. Poisson.

(1) Extrait du Compte rendu de l'Académie des sciences morales et politiques (séance du 10 février 1855).

(2) Voir les Comptes rendus de l'Académie des sciences, séances des 14 décembre 1835, p. 478, et 11 avril 1837, p. 377, etc.

(3) Procès-verbal de la séance du 16 avril 1842.

Je viens, en conséquence, demander à cette Académie la permission de répéter devant elle que la *Loi des grands nombres* n'existe pas, et d'entrer à ce sujet dans quelques explications indispensables. On verra par là que je n'éprouve nul doute sur l'exactitude de ma dénégation; et en même temps elle parviendra plus sûrement aux personnes qui s'occupent des observations et des expériences si multipliées qui ne rentrent pas précisément dans le domaine des sciences physiques ou mathématiques.

Pour être bien compris, il est bon de rappeler que le mot *cause*, quand il s'agit de probabilités, a reçu un sens spécial. Les auteurs, en l'employant, n'entendent pas parler de ce qui produit un effet ou un événement, de ce qui en assure l'arrivée; ils veulent seulement parler de l'état des choses, de l'ensemble des circonstances pendant lequel cet événement a une probabilité déterminée. Ainsi, par exemple, si la probabilité mathématique de la naissance d'un garçon, dans un certain pays, restait numériquement la même, on dirait que la cause ou les causes de la naissance d'un garçon sont constantes. Ce qui n'empêcherait nullement qu'une grossesse prise au hasard amenât une fille, puisque la probabilité constante de la naissance d'un garçon laisse subsister la probabilité contraire, alors constante aussi, de la naissance d'une fille.

Ce mot ainsi défini, on voit sans difficulté ce que signifie le théorème de Jacques Bernouilli sur les événements soumis à des probabilités constantes. En voici l'énoncé : lorsque les causes sont constantes, et que par suite la probabilité d'un événement ou d'un fait quelconque reste la même pour chaque épreuve ou chaque observation, le nombre des répétitions de cet événement sur un grand nombre d'épreuves est à peu près proportionnel à la fraction qui exprime sa probabilité; en d'autres termes le rapport des nombres de répétitions d'un événement et de l'événement contraire s'écarte peu du rapport de leurs probabilités respectives.

Cela ne veut pas dire que cette petite différence des rapports se réalise nécessairement. Elle n'est que très-probable. Si le nombre des observations ou épreuves est excessivement grand, elle devient excessivement probable. Mais elle ne sera jamais certaine. Il en est toujours ainsi dans tout calcul de probabilités; et c'est là souvent ce qui rend si difficile de suivre un long raisonnement sur des choses probables, et d'échapper à toute illusion.

Du théorème de Bernouilli il ressort une conséquence fort remarquable, c'est que les résultats moyens de causes constantes ne sauraient offrir, le plus souvent, que de très-petits écarts. De sorte que la part du hasard est très-petite : il faut que la cause change pour que les valeurs moyennes des faits observés en grand nombre viennent à subir de grandes variations.

Par exemple : sur un million d'épreuves assujetties à la même probabilité mathématique, l'écart entre la moyenne observée et la moyenne calculée d'après cette probabilité n'excédera pas mille probablement. Mais sur 10,000 épreuves un écart de 100 aurait la même probabilité, du moins à très-peu près.

La petitesse de l'étendue de ces écarts reste la même, quelle que soit la nature des faits observés; et elle reste encore la même si l'on se représente la probabilité constante non plus comme absolument fixe, mais comme étant la valeur moyenne constante d'un certain nombre de probabilités qui résultent de causes variables, dont chacune peut se présenter à chaque épreuve indifféremment, suivant une loi de possibilité assignée d'avance.

On conçoit facilement que les deux cas reviennent absolument au même. Il n'est pas besoin de le démontrer, et sans doute ceux qui en ont pris la peine n'y avaient vu qu'un exercice de calcul. L'identité d'une probabilité constante et de la probabilité moyenne d'un certain nombre de probabilités qui peuvent toutes régir une épreuve quelconque, avait paru jusqu'à ces derniers temps d'une évidence complète. C'est même ainsi que Jacques Bernoulli a entendu sa probabilité unique. On peut s'en assurer en lisant ce qu'il en a dit dans la préambule de son théorème (1). Il donne comme exemple la probabilité résultant des causes multiples des maladies mortelles; et, qui plus est, la probabilité constante déduite des causes météorologiques : *Innumerabili complexionum varietate industriam nostram æternum lusuris.*

On sait que l'ouvrage posthume de Bernoulli est brusquement interrompu après la démonstration de son théorème. De sorte qu'il ne subsiste aucune trace des applications réelles dont il promettait l'exécution, et qu'il est impossible de savoir comment il les aurait dirigées et entendues. Mais quand on fait des recherches scientifiques vraiment sérieuses, qu'on ne se borne pas à de petits nombres d'expériences ou d'observations, et que l'on a à comparer les faits de plusieurs années, il est difficile qu'on ne s'aperçoive que les écarts assignés par le théorème de Bernoulli sont loin d'égaliser les différences considérables qui se rencontrent entre les rapports des nombres des phénomènes naturels recueillis avec le plus d'exactitude. Il y a surtout un fait qui doit frapper un esprit attentif : c'est que le nombre des observations nécessaires pour obtenir une moyenne offrant un certain degré de précision, n'est pas le même, à beaucoup près, dans tous les genres de recherches. Dans certaines recherches, des nombres petits relativement suffisent pour que les résultats moyens diffèrent peu les uns des autres. Pour d'autres, cette espèce de constance des résultats moyens exige des nombres d'observations beaucoup plus grands.

Ainsi les comptes de la justice criminelle attestent que le rapport moyen des condamnations aux acquittements a peu varié d'une année à l'autre (sous la même législation), bien que le nombre total des accusations n'excède pas 7,000 ou 8,000.

Au contraire, pour que le rapport des naissances de garçons aux naissances de filles prenne le même degré de fixité, il faut des nombres 10 ou même 100 fois plus considérables.

Il serait permis, jusqu'à un certain point, d'attribuer ces contrastes à ce que les causes varient bien plus pour une classe de faits que pour une autre. Et c'est ce qui résulterait de la simple application du théorème de Bernoulli. Mais cette explication n'est point satisfaisante, parce qu'en étudiant le mieux possible les circonstances des choses, on ne voit pas clairement qu'elles aient dû réellement beaucoup changer d'un recueil de faits à un autre. Souvent même, on a lieu de constater des écarts considérables alors qu'on est presque persuadé de la constance de l'ensemble de causes qui a régi les diverses séries de recherches.

Cette discordance n'avait pas échappé à M. Poisson, et il avait senti la nécessité d'une nouvelle explication de la manière dont se combinent les probabilités pour que les résultats moyens finissent par devenir constants, en exigeant toutefois des nombres d'observations très-différents,

---

(1) *Ars conject.*, pars 4<sup>a</sup>, p. 224 et 226.

C'est là du moins ce qu'on peut conclure des divers passages où il a parlé de sa loi des grands nombres (1).

Si donc il n'avait publié aucun calcul, aucune formule, bien que ses idées ne paraissent pas avoir été arrêtées très-nettement, et qu'il semble avoir changé parfois de point de vue, on devrait présumer qu'il avait donné le nom de *Loi des grands nombres* à quelque théorème nouveau. On pourrait sans doute penser que ce nom offre un sens plus large que le sens possible d'un nouveau théorème : car il est bien clair qu'il y a des nombres très-grands qui n'offrent pas de résultats constants, tandis que les mots *Loi des grands nombres* doivent exprimer quelque règle applicable à tous les nombres, dès qu'ils auraient le caractère d'être *grands* : sauf à s'entendre sur leur grandeur.

Mais les calculs de M. Poisson ne permettent aucun doute. Après avoir lu ses *Recherches sur la probabilité*, si l'on a suivi l'analyse qu'il développe et les formules qui en résultent, on acquiert la certitude qu'il a simplement démontré le théorème de Jacques Bernoulli, dans l'hypothèse où la probabilité constante est la valeur moyenne d'un ensemble de probabilités variables qui peuvent s'offrir toutes à toutes les épreuves : hypothèse si évidente qu'il n'est pas nécessaire de la démontrer.

Les formules de M. Poisson assignent en effet les mêmes écarts, très-petits, qui se déduisent de la règle de Bernoulli. On n'y trouve pas ce qu'il avait annoncé dans l'introduction de son ouvrage et antérieurement, savoir : des écarts différents pour des nombres d'égale grandeur, quand les phénomènes diffèrent ; rien n'y marque non plus la nécessité de nombres d'observations plus grands dans certains cas que dans d'autres, pour arriver à des résultats moyens à peu près fixes. Si M. Poisson eût bien examiné l'étendue si restreinte des écarts de la formule de Jacques Bernoulli ou de la sienne, il aurait sans doute abandonné toute idée d'une découverte réelle. A cet égard, il est resté dans une illusion singulière. Car il a bien été obligé de reconnaître (2) que la probabilité moyenne jouait dans sa formule le rôle de la probabilité unique dans celle de Bernoulli, de manière que les formes étaient véritablement identiques. Mais au moment même où il admet cette identité de forme, il soutient qu'il y a différence au fond. Et l'explication qu'il croit en donner n'est pas exacte, puisqu'elle ne repose que sur les diversités de la succession possible des causes, et qu'il a soigneusement fait remarquer lui-même que cette succession est à chaque épreuve également possible dans tous les sens.

Toute son analyse, de même que ses résultats, établissent qu'il en est du fond comme de la forme : et, d'après ses calculs, rien ne nécessite, pour parvenir à des moyennes constantes, de plus grands nombres que n'en exige la formule de Bernoulli.

On est donc en droit d'affirmer que la *Loi des grands nombres* n'a pas d'existence réelle, et qu'il ne faut plus s'appuyer de ces mots dans le sens que M. Poisson leur avait donné.

Le bon sens suffisait d'ailleurs à faire naître des doutes sérieux sur la prétendue découverte. Comment pourrait-il se faire que la grandeur des nombres dispensât de la constance des causes ? On ne comprend pas l'existence de résultats constants, sans

---

(1) Voir les passages des Comptes rendus de l'Académie des sciences cités plus haut, p. 1, et les *Recherches sur la probabilité des jugements*, Introd., et ch. 2 et 4.

(2) Voir page 146 des *Recherches sur les Jugements*.

remonter à l'existence de quelque réunion constante de circonstances qui les produit. On a d'ailleurs des exemples assez fréquents de très-grands nombres dont les relations ne cessent de varier. Enfin M. Poisson lui-même admettait que les causes ne devaient pas varier d'une manière progressive. Mais par le fait, comme il n'en prenait que la moyenne, il ne les faisait pas varier du tout.

J'ai supprimé ici tout calcul à ce sujet : je dirai seulement que toute l'algèbre du monde n'y ajoute que très-peu de clarté.

Je citerai un résultat particulier auquel j'étais parvenu il y a longtemps, bien que je ne l'aie publié qu'en 1839 (1). En cherchant à expliquer comment d'un ensemble constant de causes fixes, il peut résulter des écarts considérables dans les résultats moyens, j'ai trouvé le théorème que voici : Quand des causes formant un ensemble invariable peuvent se présenter indifféremment, mais quand la probabilité due à celle qui vient à s'offrir régit un très-petit nombre d'épreuves successives, trois ou quatre, par exemple, l'étendue des écarts probables au-dessus ou au-dessous de la probabilité moyenne peut devenir double de ce qu'elle était dans la formule de Bernoulli, de sorte qu'il faut alors un nombre d'observations quatre fois plus grand pour arriver à des moyennes qui jouissent du même degré de fixité probable. On conçoit que si la probabilité variable d'une cause à l'autre durait pendant plus de trois ou quatre épreuves, l'étendue des écarts pourrait être très-grande. Ce serait alors une véritable variation des causes.

Aussi la formule qui exprime cette loi spéciale peut-elle servir à expliquer la marche irrégulière de certains phénomènes. Telle est, par exemple, la quantité moyenne de pluie qui tombe chaque année. Comme les causes qui amènent la pluie durent pendant plusieurs jours consécutifs, que d'ailleurs le nombre des jours de pluie n'atteint guère que de 100 à 200 par année dans nos climats, il faut un temps très-long pour obtenir une moyenne qui ait un certain caractère de fixité. Jusqu'à présent les observations ne comprennent guère plus d'un siècle; il n'a été possible de former que deux ou trois moyennes d'environ 30 ou 40 années chacune et la connaissance des lois que suivent les quantités d'eau dans les diverses saisons est très-peu avancée.

Il y a quelques années, par exemple, on plaçait Poitiers et ses environs sous le régime des pluies d'été, d'après dix années d'observations qu'un médecin, M. de La Mazière, avait envoyées au Père Cotte, qui recueillait toutes les notions statistiques qu'il pouvait se procurer. Mais il convient, au contraire, jusqu'à plus ample informé du moins, de ranger Poitiers dans la zone des pluies d'automne. C'est la conséquence qui ressort de l'extrait que j'ai fait, il y a dix ans, d'un manuscrit du même médecin, M. de La Mazière. Ce manuscrit, qui m'a été communiqué par le bibliothécaire de la ville de Poitiers, contenait 40 années d'observations assidues, dont 30 postérieures à celles que M. de La Mazière avait envoyées au Père Cotte (2). Ainsi, 10 années avaient fourni une moyenne opposée à la véritable moyenne

---

(1) Voir les procès-verbaux de la Société philomatique du 4 mai 1839, et le journal *l'Institut* du 6 juin 1839.

(2) Cette conséquence est beaucoup plus importante qu'elle ne le paraît au premier abord. Le repeuplement des bois après les coupes, si facile en Allemagne et dans certaines parties de la France, ne paraît résulter de ce que ces contrées ont le *maximum* des pluies en été. Dans une très-grande partie de la France, le *maximum* est en automne. Il en résulte que très-souvent les jeunes recrues sont grillées par le soleil, et périssent faute d'eau. Il y a donc des précautions spéciales à prendre lors des coupes de bois, pour parer à cet accident naturel qui accroît les chances de dépeuplement.

des quantités d'eau tombées dans les différentes saisons. Il en a fallu 40 pour faire ressortir cette vraie moyenne. Et il en faudrait 40 ou 80 autres pour la confirmer.

Malgré l'application heureuse qui a pu se faire dans ce cas du théorème dont je viens de parler, je ne le considère nullement comme toujours capable d'expliquer les grands écarts que présentent des faits statistiques journaliers. Je pense qu'il n'y a point ici de loi générale à chercher, que c'était une véritable illusion que de se proposer un tel but. Il y a manifestement variation dans les causes quand les résultats varient : je crois donc qu'il faudra étudier chaque résultat pour deviner en quelque sorte par quelle combinaison variable de causes constantes, ou même de causes variables, on pourra les expliquer. La variété infinie des arrangements employés par la nature pour produire très-simplement des effets très-complexes fait présumer qu'on découvrira des théorèmes particuliers qui régiront diverses classes de faits. Mais il ne semble pas qu'une formule analogue à celle de Jacques Bernoulli puisse embrasser toutes les circonstances possibles.

Je me borne à cette indication pour ne point m'écarter davantage du seul point que je désirais constater ici, la non-existence de ce qui avait été appelé : *Loi des grands nombres*. Je dirai en terminant que, si quelques statisticiens habiles ont montré sur ce sujet une juste défiance et ne s'y sont pas mépris, il n'en n'est pas moins regrettable que ce prétendu théorème ait reçu un nom, et surtout un nom aussi propre à saisir les oreilles. Lorsque quelque idée est nommée, il semble qu'elle existe, à tel point qu'en Allemagne on a soutenu que le néant était quelque chose, uniquement parce qu'il faut bien le désigner par un mot. J'espère que le mot qui vient de m'occuper ne sera pas conservé dans les applications scientifiques, et que l'on s'en tiendra au théorème de Jacques Bernoulli, sous la forme si remarquable et si simple que Moivre avait donnée le premier, et qui a été étendue si heureusement par Laplace.

I.-J. BIENAYMÉ,  
*membre de l'Institut*

---