LES CAHIERS DE L'ANALYSE DES DONNÉES

J. P. BENZÉCRI

Histoire et préhistoire de l'analyse des données. Partie I La préhistoire

Les cahiers de l'analyse des données, tome 1, nº 1 (1976), p. 9-32

<http://www.numdam.org/item?id=CAD_1976__1_1_9_0>

© Les cahiers de l'analyse des données, Dunod, 1976, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » 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/

Les Cahiers de l'Analyse des Données
Vol.I - 1976 - n°1 - p.9-32

HISTOIRE ET PREHISTOIRE DE L'ANALYSE DES DONNEES

Partie I - La Préhistoire

par J.P. Benzecri [1]

Les progrès de l'analyse des données liés à l'avènement des ordinateurs ne vont pas sans bouleverser toute la statistique. Des écoles naguère toutes-puissantes perdent leur suprématie. Des prémisses philosophiques, qui n'avaient jamais été explicitement proclamées sont ouvertement critiquées. Soucieux de rendre à chacun ce qui lui est dû, nous nous efforcerons de retrouver par quelles voies la statistique est parvenue au point où nous la voyons aujourd'hui ; et, afin d'abaisser l'orgueil humain et d'apaiser les querelles nous nous placerons d'abord dans la contemplation du passé : nous ferons débuter aussi tôt qu'il est possible l'histoire de la statistique et des probabilités (§ 1). Puis, ayant fait aux écoles anglo-saxonnes (§ 2) la place qui leur revient depuis le milieu du XIX^ème^ siècle, nous parviendrons aux recherches qui nous occupent (§ 3).

1. *La science du hasard* :

Au sein de grandes masses de faits, derrière l'imprévu des rencontres individuelles, la statistique ou dénombrement découvre des lois d'ensemble. Et le mathématicien par une énumération ordonnée des possibles, précise et parfois démontre ces lois. Bien que l'abus de calcul des probabilités ait nui à la statistique (cf § 1.7.6), le progrès de ces deux disciplines sera dans ses grandes lignes décrit comme l'histoire d'une seule science : celle du hasard. Sans prétendre condenser en un chapitre ce qu'a pu rassembler l'érudition d'un Isaac Todhunter (cf The History of the Mathematical Theory of Probability ; 1865), nous nous appliquerons à rappeler ce qui importe le plus à notre objet - à l'analyse des données multidimensionnelles.

1.1. *Le monde antique* :

L'administration des grands empires qui, il y a trois ou quatre mille ans existaient déjà en Egypte, en Mésopotamie et en Chine ne se conçoit qu'appuyée sur une sorte de statistique et notamment sur des recensements dont l'Ecriture Sainte fait volontiers mention. Il est hors de doute que les scribes égyptiens étaient de remarquables gestionnaires et si les secrets de leur comptabilité ne nous sont point parvenus, il est juste de faire remonter au moins à leur époque les origines de la statistique descriptive (*).

(*) *Dans sa Statistique, collection Thémis, P.U.F. T. 1 p. 8 André PIATIER signale que "les Egyptiens avaient mis au point le plus ancien baromètre économique connu à ce jour et qui était le Nilomètre : la hauteur de la crue du Nil était un excellent indice de fertilité - et elle servait à fixer le montant des impôts".*

(1) *Laboratoire de Statistique - Université Pierre et Marie Curie - Paris*

Quant aux probabilités, on cherchera leur origine dans la méditation du hasard et de la fortune, certes familière aux hommes depuis des millénaires, ainsi que dans les jeux dont la pratique semble attestée par la préhistoire (*). Un témoignage sur la pensée antique nous est offert par la physique d'Aristote (**). Critiquant d'abord les vues de ses devanciers, Aristote (384-322) nous fait connaître trois attitudes vis-à-vis du hasard et de la fortune. "Certains, dit-il, mettent en question leur existence. Rien évidemment, dit-on, ne peut être l'effet de fortune, mais il y a une cause déterminée de toute chose dont nous disons qu'elle arrive par hasard ou fortune [...]. Pour d'autres, et notre ciel et tous les mondes ont pour cause le hasard car c'est du hasard que proviennent la formation du tourbillon et le mouvement qui a séparé les éléments et constitué l'univers dans l'ordre où nous le voyons [...]. D'autres encore pensent que la fortune est une cause mais cachée à la raison humaine, parce qu'elle serait quelque chose de divin et de surnaturel à un degré supérieur". On reconnaîtra dans ces trois attitudes des pensées qui nous sont toujours présentes. Les vues de Laplace et de Poincaré, selon qui le hasard n'est lié qu'à notre ignorance, le jeu des causes multiples que nous ne parvenons pas à pénétrer pouvant expliquer rigoureusement tout, se retrouvent dans la première école, qui nie l'existence du hasard, voire dans la troisième qui affirme qu'il faudrait une intelligence supérieure pour en pénétrer le jeu. Et les thèses défendues avec passion par des spécialistes éminents de biologie moléculaire quant à l'origine aléatoire des formes les plus perfectionnées de la vie, sont évidemment déjà présentes dans cette théorie des tourbillons familière aux physiciens présocratiques. Aristote quant à lui fonde son exposé du hasard sur la distinction entre <u>cause par soi</u> et <u>cause par accident</u>. C'est, selon Aristote, une cause par soi qu'une statue soit l'oeuvre d'un sculpteur ; c'est une cause par accident qu'elle soit l'oeuvre d'un musicien (si le sculpteu se trouve d'autre part avoir des compétences musicales). De même dit Aristote, les faits que nous appelons faits de fortune ou de hasard sont des rencontres d'ailleurs rares de séries qui n'étaient pas comprises dans un même projet, dans une même intention.

Reconnaissons toutefois que, quelle que soit la richesse de ces pensées, on y cherche en vain ne fût-ce qu'un germe de ce qui constitue la base de notre doctrine probabiliste : la possibilité de calculer sur le hasard et d'en mesurer les effets. L'idée que des causes minuscules peuvent s'additionner dans des proportions déterminées pour produire des effets visibles réguliers semble manquer aux anciens. Ils ont ignoré tout à la fois le calcul différentiel, (dont toutefois on a pu reconnaître les prémices chez Archimède (287-212)), et le calcul des probabilités qui se trouvent, sans doute pour quelque profonde raison être nés ensemble au XVIIème siècle.

(*) *Sur l'usage préhistorique des osselets comme d'un jeu de dés, cf F.N. DAVID Dicing and gaming (a note on the history of probability) <u>Biometrika</u>, T. 42 (1955) ; reproduit dans le recueil : <u>Studies in the History of Statistics a</u> <u>Probability</u>, ed. by E.S. PEARSON & M.G. KENDALL, Londres, Griffin, (1970) ; ce recueil sera désormais cité sous le titre abrégé de <u>Studies</u>.*

(**) *Physique L II ch 4 ; nous utilisons l'édition du texte grec et la traductio française de H. CARTERON, Paris Les Belles Lettres ; ainsi que le texte latin accompagné du commentaire de Saint Thomas d'Aquin : <u>In octo libros physicorum Aristotelis expositio</u> ; ed. Marietti.*

1.2. *Le moyen-âge* :

Ne nous hâtons pas cependant de passer d'Aristote à Leibnitz, d'Archimède à Pascal. Pendant les siècles qui suivirent l'écroulement de l'Empire romain et où se constitua la Chrétienté européenne, l'Orient connut une grande activité scientifique et artistique. Et c'est à cette époque que passa d'Orient en Europe le nom même de hasard. Bien que l'étymologie précise soit difficile à établir, il est avéré que ce mot nous est venu de Syrie. Un dictionnaire arabe moderne donne du nom de zahar الزّهر le sens de "fleur" suivi d'une acception secondaire liée au jeu de dés. Là serait, selon la plupart, l'étymologie de notre nom hasard. Plus précisément, on y peut voir la désignation de figures rares au jeu de dés : par exemple, le "triple six". Cependant d'aucuns prétendent que le nom de hasard est celui d'un château de Syrie. Ce serait pendant le siège de ce château qu'aurai été inventé un certain jeu de dés dont le nom serait passé à tous les jeux, puis à notre science même. Cette étymologie paraîtrait invraisemblable si elle n'avait pour elle l'autorité d'un chroniqueur du XIIème siècle, un Franc d'Orient, Guillaume de Tyr (cf Littré s.v. hasard). Nous n'irons pas plus loin dans ces recherches étymologiques qui nous dépassent, content seulement d'avoir montré le lent cheminement de la pensée depuis l'Antiquité jusqu'à la manifestation des fondateurs de la Statistique moderne et des probabilités.

1.3. *Les Pères du calcul des probabilités* :

Ces hommes sont on le sait, Fermat (1601-1665), Pascal (1623-1662), Huyghens (1629-1695) et Jacques Bernoulli (1654-1705), précédés d'un siècle par les italiens Luca Pacciolo (1445-1514),Tartaglia (1499-1557),Cardan (1501-1576), sans oublier le grand Galilée (1564-1662). Mais il n'est pas sûr ici encore, que parmi tant d'humanistes orientaux à la vie aventureuse (nous pensons à un Omar Khayyâm, poête, mathématicien, astronome, philosophe ; si semblable aux plus étincelantes figures du XVIème siècle italien) des devanciers arabes ou persans ne doivent être un jour découverts à ces recherches inspirées par la folie du jeu ; recherches qui, dans l'Occident aboutirent en un peu moins d'un siècle à l'élaboration d'une analyse combinatoire fondée sur le fameux triangle de Pascal, lequel est, à la fin du XVIIème siècle, parfaitement exploré par nos savants. Et le XVIIIème siècle s'ouvre par la publication d'une remarquable démonstration analytique que donne Jacques Bernoulli de la loi des grands nombres. Au terme d'une méditation soutenue qui, dit-il, l'occupa vingt années (*) cet auteur, l'aîné d'une illustre lignée, parvint à démontrer avec une grande rigueur que la fréquence d'apparition d'un phénomène ayant une probabilité déterminée tend vers cette probabilité lorsque le nombre des essais est multiplié.

1.4. *Le XVIIIème siècle* :

Sans prétendre donner des grands progrès réalisés au cours du XVIIIème siècle par le calcul des probabilités et la statistique une histoire très fidèle à leurs justes proportions, nous signalons quelques travaux où apparaissent des idées dont l'importance est aujourd'hui manifeste (§ 1.4.1) ; et exposons les origines de la théorie de la probabilité des causes (§ 1.4.2) ; avant de saluer la naissance de la statistique moderne (§ 1.5).

(*) *Hoc igitur est illud problema, quod evulgandum hoc loco proposui, postquam jam per vicennium pressi, et cujus tum novitas, tum summa utilitas cum pari conjuncta difficultate omnibus reliquis hujus doctrinae capitibus pondus et pretium superaddere potest (cité d'après J. Bertrand). "Voici donc le problème que j'entends publier ici, après m'y être appliqué pendant vingt ans et dont tant la nouveauté que la suprême utilité jointe avec une égale difficulté peuvent ajouter du poids et du prix à tous les autres chapitres de notre science".*

1.4.1. *Progrès du calcul des probabilités* :

Dans son ouvrage paru d'abord dans les Transactions philosophiques de l'année 1711, Abraham de Moivre (1667-1754) a repris "le théorème de Jacques Bernoulli sur la probabilité des résultats déterminés par un grand nombre d'observations". Mais - nous citons Laplace - "Il ne se contente pas de faire voir, comme Bernoulli, que le rapport des événements qui doivent arriver approche sans cesse celui de leurs probabilités respectives, il donne de plus une extension élégante et simple de la probabilité que la différence de ces rapports est contenue dans les limites données". Cette "expression élégante et simple" n'est autre que la formule de la loi normale (cf §§ 1.5.2 - 3 & 4 ; § 2.2 ; etc.), loi que Moivre atteint ainsi le premier en partant de la formule asymptotique de Stirling pour n !

Le révérend Thomas Bayes (1702-1761) a médité le problème inverse de celui résolu par Moivre : de la fréquence d'un événement sur une série d'essais, inférer la distribution de la probabilité de cet événement. C'était poser la question de la probabilité des causes dont Laplace semble avoir donné le premier la formule générale : nous ferons de cette question un exposé assez détaillé (§ 1.4.2), car elle n'a cessé depuis deux siècles de préoccuper les statisticiens.

Parmi les applications de la probabilité des causes, les recherches sur la probabilité des jugements rendus à la pluralité des voix d'un jury furent très en vogue à la fin du XVIII$^{\text{ème}}$ siècle et longtemps dans le XIX$^{\text{ème}}$; y sont associés les noms de Condorcet (1743-1794), Laplace (1749-1827), Poisson (1781-1840) : mais on a pu contester l'utilité de telles spéculations (cf § 1.6.1).

Les savants continuent de s'intéresser aux jeux de hasard. Citons le jeu de Saint Pétersbourg - proposé par Nicolas Bernoulli (1687-1759) : Pierre et Paul jouent à pile ou face ; le jeu s'arrête quand face sort pour la première fois ; alors Paul donne à Pierre 2^{n-1} si n a été le nombre de jets. D'où :

Question : Quelle doit être la mise de Pierre, (qui est toujours le gagnant ; mais d'une somme variable selon les aléas du jeu de pile ou face) pour que le jeu soit équitable ?

Réponse : L'espérance mathématique du gain de Pierre est $\Sigma(1/2^n) \times 2^{n-1} = \infty$; Pierre doit donc donner à Paul avant la partie une mise infinie.

Cette conclusion offre matière à bien des réflexions morales ... ; mais si nous avons cité le jeu c'est parce qu'en modifiant ses règles pour le rendre réaliste on aboutit à un exemple précoce de ce que nous appelons processus avec temps d'arrêt. Dans le même sens, on citera encore la démonstration de la "ruine du joueur", c'est-à-dire du fait que dans une suite indéfinie de parties d'un jeu équitable, un joueur en viendra sûrement à accumuler des pertes dont le total dépassera sa fortune, ce qui fera sa ruine ; bien que le jeu, poursuivi indéfiniment lui eût aussi apporté non moins sûrement des espérances de gain dépassant toute mesure.

Il faut maintenant citer Buffon (1707-1788), connu de tous pour être l'auteur d'une monumentale Histoire Naturelle, mais qui écrivit aussi un Essai d'arithmétique morale (publié en 1777) où brille un exemple fameux de probabilité géométrique : le problème de l'aiguille. Une aiguille qu'on laisse choir tombe sur un sol horizontal avec une direction uniformément répartie ; si le point d'où on libère l'aiguille est lui-même distribué quasi-uniformément (i.e. avec une densité qui varie peu sur une distance égale à la longueur de l'aiguille) sur un plan horizontal, il en résulte, comme le veut Buffon que vaut $(2/\pi)$ la probabilité pour que l'aiguille se pose en coupant l'interstice entre deux lattes d'un parquet dont l'intervalle est égal à sa longueur. Les probabilités géométriques, où les hypothèses de symétries jouent un rôle essentiel (cf e.g. [Principes] TII A n°1 § 1°) mais difficile à préciser en toute rigueur, offrirent au XIX$^{\text{ème}}$ siècle matière à plus d'un paradoxe (cf infra § 1.6.3).

Les savants du XVIII$^{\text{ème}}$ siècle ont grandement perfectionné les méthodes mathématiques du calcul des probabilités. Dans sa "Théorie analytique des probabilités" Laplace définit les fonctions génératrices et crée la tranformation qui porte son nom. Nous ne tenterons pas d'imiter ici l'illustre géomètre qui dans l'Essai philosophique donne de son analyse un exposé dépourvu de toute formule mathématique ; mais devions au passage saluer un monument. (Pour un exposé des formules, cf § 1.5.3).

1.4.2. *Probabilité inverse et probabilité des causes* :

Avant toute référence historique à Bayes et à Laplace donnons de la formule des probabilités des causes une exposition détaillée qui nous servira dans la suite (cf §§ 2.3.3 & 3.5.1). Soient I et J deux ensembles : I ensemble des causes, J ensemble des effets ; une situation sera pour nous un couple (i,j) formé d'une cause i et d'un effet j. Entre cause et effet, le lien n'est pas supposé rigide (à la vérité la nature du lien causal n'importe pas à ce calcul ; et un grand statisticien a prétendu, abusivement selon nous, bannir la notion de cause, pour ne parler que de lien ; cf § 2.2.7) : mais on désigne par p_j^i ou $p(j/i)$ (lire : probabilité de j quand i) la probabilité (conditionnelle) que la cause i étant posée il résulte l'effet j. Notons de plus p_i la probabilité que la cause soit i ; p_{ij} la probabilité qu'on ait simultanément la cause i et l'effet j. On a la relation : $p_{ij} = p_i \times p_j^i$, ce que nous gloserons : la probabilité que simultanément la cause soit i et l'effet j est égale au produit de la probabilité p_i de la cause i par la probabilité (conditionnelle) que la cause i engendre l'effet j. On peut encore, dans ce modèle remonter de l'effet j à la cause i. L'ensemble $\{(i,j)|i\in I\}$ (cause i quelconque dans I ; effet fixé j) des situations où l'effet est j, a pour probabilité totale la somme $p_j = \Sigma\{p_{ij}|i\in I\}$, (somme des p_{ij} pour i variable, j fixé) ; dans cet ensemble, le poids relatif d'une cause particulière i' est $p_{i'}^j = p_{i'j}/p_j$. Ce quotient $p_{i'}^j$ est la probabilité (conditionnelle) que l'effet étant j la cause soit i. Une fois admise une probabilisation complète de l'ensemble produit I x J des situations (i,j), la symétrie est totale entre ce que nous avons appelé effet (j) et cause (i). Dans les applications en vue, comme le suggèrent les noms d'effet et de cause, il y a dissymétrie : j, l'effet, est directement ou facilement observable ; i, la cause, est plus caché. De plus, les probabilités conditionnelles p_j^i (qu'une cause i étant posée, il résulte l'effet j) sont généralement hors de conteste ; mais les probabilités p_i des causes (dites encore probabilités a priori par opposition aux probabilités a posteriori p_i^j estimées d'après la connaissance de l'effet j) sont souvent estimées en traduisant par des nombres une simple impression qualitative (cf § 1.7.3).

Le mémoire posthume de Bayes "An essay toward solving a Problem in the Doctrine of Chances" (Phil. Trans. 1764-65, réédité dans Biometrika T. 45, 1958 ; et dans Studies) est consacré à un problème, appelé problème de la probabilité inverse, parce qu'il est comme l'inverse de celui résolu par Bernoulli et Moivre. L'événement A a une probabilité π dont la valeur est inconnue. Déterminer la probabilité que π soit comprise dans l'intervalle (a,b) $(0 \leq a < b \leq 1)$ sachant que dans les résultats de n essais indépendants, l'événement A est appar m fois (et m - n fois n'est pas apparu). Sous la condition complémentaire qu'avant expérience (a priori) la probabilité π est distribuée uniformément entre 0 et 1, on a un problème de probabilité des causes posé comme on l'a fait ci-dessus. Mais ici l'ensemble I des causes n'est pas un ensemble fini ; c'est l'intervalle (0,1), la cause i de la série d'essais observés étant la probabilité π (paramètre numérique) de A. C'est pourquoi on ne se propose pas de chercher un p_i^j (probabilité que la cause soit au point i) mais ce qu'on peut noter $p_{(a,b)}{}^j$, (probabilité que la cause soit dans l'intervalle (a,b) : au § 3.5.1 on reviendra sur les transitions probabilistes non entre ensembles finis, mais entre espaces probabilisables). Quant à J, ensemble des effets, il s'identifie à [n] ensemble des entiers compris entre 0 et n (le nombre des essais). Comme nous l'avions annoncé la probabilité p_m^π que, π étant posé, il en résulte m est hors de conteste : on déduit de la formule du binôme que $p_m^\pi = \pi^m(1 - \pi)^{n-m}$ $(n!/(m!(n - m)!))$; mais postuler a priori π uniformément distribué sur (0,1) n'est justifié que par l'ignorance totale où l'on est de sa distribution. Toute la querelle entre statistique dite Bayesienne (postuler des probabilités a prior p_i des causes) et statistique non-Bayesienne (ne travailler que sur les probabilités conditionnelles p_j^i) est ici en germe (cf §§ 1.7 ; 2.3.3 ; 2.3.5).

Fisher rend honneur à Bayes d'avoir posé le premier le problème de l'inférence statistique ; et il ajoute - louange suprême dans sa bouche - que Bayes fut non-Bayesien ; qu'il s'abstint de publier ses recherches de son vivant, parce qu'il connaissait la fragilité de l'hypothèse d'uniforme distribution de la probabilité a priori. Cependant on attribue communément à Bayes la formule générale même des probabilités des causes, formule que Laplace revendique (et qui n'est assurément pas dans Bayes : où l'on trouve seulement la formule des probabilités composées qui avec nos notations s'écrit : $p_{ij} = p_i \times p_j^i$ ou : $p_j^i = p_{ij}/p_i$). Ecoutons Laplace :

"Bayes, dans les Transactions philosophiques de l'année 1763, a cherché directement la probabilité que les possibilités indiquées par des expériences déjà faites sont comprises dans les limites données, et il y est parvenu d'une manière fine et très ingénieuse, quoiqu'un peu embarrassée. Cet objet se rattache à la théorie de la probabilité des causes et des événements futurs, conclue des événements observés, théorie dont j'exposai quelques années après les principes, avec la remarque de l'influence des inégalités qui peuvent exister entre les chances que l'on suppose égales" (Essai Philosophique).

On lit en effet dans la Théorie Analytique, cet énoncé que nous laissons au lecteur, de traduire en formules !

"Si un événement observé peut résulter de n causes différentes, leurs probabilités sont respectivement comme les probabilités de l'événement tirées de leur existence, et la probabilité de chacune d'elle est une fraction dont le numérateur est la probabilité de l'événement dans l'hypothèse de l'existence de la cause, et dont le dénominateur est la somme des probabilités semblables, relatives à toutes les causes. Si ces diverses causes considérées a priori sont inégalement probables, il faut au lieu de la probabilité de l'événement, résultante de chaque cause, employer le produit de cette probabilité par celle de la cause elle-même".

C'est bien ce qu'on appelle souvent le théorème de Bayes. Théorème dont nous citerons encore un exemple d'application semblable au problème de Bayes, mais assez piquant. Les votes d'un collège de N électeurs sont en cours de dépouillement : on a déjà compté m bulletins en faveur du candidat A et (n - m) bulletins pour son unique adversaire B : quelle est la probabilité que A soit élu ? Evidemment, ici encore, il faut disposer de probabilités a priori sur les intentions des électeurs (à moins que l'un des deux nombres m ou (n - m) ne dépasse N/2 ; en ce cas, il y a certitude !) . Il n'est pas réaliste de postuler, comme dans le problème de la probabilité inverse, que le nombre M des électeurs favorables à A (ce M joue ici le rôle de la cause i) est distribué uniformément (de 0 à N) : on devra donc fixer a priori non sans beaucoup d'arbitraire, une courbe en cloche (ou en créneaux) centrée sur une valeur M estimée la plus probable.

1.5. *Naissance de la statistique moderne* :

C'est par l'introduction d'applications nouvelles que les probabilistes du XVIII^ème^ siècle fondent la statistique mathématique. (Selon A. Piatier, op. laud., le terme même de Statistique fut forgé par G. Achenwall (1719-1772), professeur à Göttingen ; mais il faut remonter au latin status, à l'italien statista, la référence à la notion politique étant claire). En effet aux jeux de hasard s'adjoignent les statistiques démographiques, et à la fin du siècle se développe la théorie des erreurs, c'est-à-dire la recherche de la vraie valeur de grandeurs à partir d'un ensemble de mesure surabondant mais entaché d'erreurs.

1.5.1. *La démographie* (*) :

Cette discipline requiert la collecte d'abondantes données ; oriente vers la découverte de lois empiriques ; suggère la critique de l'échantillonnage. De plus, comme la théorie cinétique des gaz (fondée elle aussi au XVIII^ème^ siècle, vers 1730 : ici revient le nom de Bernoulli ; de Daniel Bernoulli (1700-1782) ; sur le développement ultérieur de cette théorie cf § 1.5.4), la démographie découvre une régularité résultant des mouvements d'une infinité d'individus qui de plus ne sont pas des molécules, mais des personnes. Dans cet ordre global, un statisticien-démographe allemand du XVIII^ème^ siècle, le pasteur Süssmilch (1707-1767) affirme reconnaître une marque providentielle. Citons après Gnedenko (**) son livre de l'Ordre Divin (Die Göttliche Ordnung) ; "Dans le détail il semble que tout apparaisse sans ordre : il faut rassembler bien des faits individuels au long des années, et réunir ensemble bien des provinces, pour mettre ainsi à jour les principes cachés de l'Ordre Divin".

(*) *Selon J. Bertillon le terme même de démographie n'a été introduit qu'en 1855, par Achille Gillard dans son ouvrage intitulé : Eléments de Statistique humaine en Démographie comparée.*

(**) *Cours de théorie des probabilités, Moscou Léningrad 1950 ; le complément historique qui termine cet ouvrage (pp. 340-367) nous a bien servi.*

Sans doute cet ordre suprême est-il plus caché encore : mais tandis que selon le proverbe "Les arbres empêchent de voir la forêt", le pasteur avait raison de noter qu'au statisticien le compte des arbres révèle les lignes de la forêt. Comme le souligne non sans quelque humour noir Emile Borel (1871-1956) (in Le Hasard, P.U.F. § 46) : "Ce sont surtout les statistiques de décès qui ont été étudiées de manière sérieuse, car c'est sur cette étude que repose toute l'organisation des assurances sur la vie". Mais dès le début, on tenta aussi une application à la thérapeutique. Citons ici Laplace (cf. Des tables de mortalité etc ..., in Essai phil.) : "Au milieu de cette dispute [suscitée par l'inoculation d'une forme bénigne de la variole pour se préserver contre la survenue d'une épidémie maligne (*)] Daniel Bernoulli se propose de soumettre au calcul des probabilités l'influence de l'inoculation sur la durée moyenne de la vie. Manquant de données précises sur la mortalité produite par la petite vérole aux divers âges de la vie, il supposa que le danger d'avoir cette maladie et celui d'en périr sont les mêmes à tout âge. Au moyen de ces suppositions, il parvint par une analyse délicate à convertir une table ordinaire de mortalité [telle que celle donnée en son temps par le célèbre Déparcieux (1705-1768)] dans celles qui auraient lieu, si la petite vérole n'existait pas, ou si elle ne faisait périr qu'un très petit nombre de malades ; et il en conclut que l'inoculation [bien que le rapport des individus qu'elle fait périr aux inoculés approchât 1/300 (**)] augmenterait de trois ans au moins la durée moyenne de la vie, ce qui lui parut mettre hors de doute l'avantage de cette opération. J. D'Alembert (1717-1783) attaqua l'analyse de Bernoulli, d'abord sur l'incertitude de ses deux hypothèses, ensuite sur son insuffisance, en ce que l'on n'y faisait point entrer la comparaison du danger prochain quoique très petit de périr par inoculation, au danger beaucoup plus grand mais plus éloigné de succomber à la petite vérole naturelle". En fait, dans ses prémisses, Bernoulli sous-estimait les bons effets de l'inoculation en supposant le risque de la variole égal tout au cours de la vie : au XVIIIème siècle, un nouveau-né sur quatre périssait avant d'avoir atteint un an ; un second quart périssait avant cinq ans ; et la variole était une cause majeure de mortalité infantile (***) ; mais la critique d'Alembert quant à la comparaison de deux risques séparés dans le temps, vaut toujours, et mériterait de retenir notamment les économistes qui calculent les coûts actualisés.

(*) *On se souviendra de distinguer cette inoculation préconisée en Europe au milieu du XVIIIème siècle (à l'imitation de l'Orient), de la vaccination jennérienne (introduite à la fin du XVIIIème siècle) qui préserve de la variole par une pustule d'une autre affection - la vaccine.*

(**) *La vaccination jennérienne au contraire a longtemps été considérée comme d'une innocuité absolue ; on admet aujourd'hui un taux de mortalité par première vaccination de 10^{-6} chez les enfants, 3×10^{-6} chez les adultes cf Lane & Millar et Katz & Krugman, in The New England Journal of Medecine (Nov. 1969) ; cités d'après L. Lebart, Recherches sur le coût de la protection de la vie humaine ... CREDOC, 1970.*

(***) *Dans un rapport de Hallé et Laplace lu à la séance du 26 Floréal an 4 (1795) de l'Académie des Sciences, sur un mémoire du citoyen J.A.Mourgues intitulé "Observations sur les naissances, mariages et morts qu'il y a eu parmi les habitants de la ville de Montpellier pendant 21 années consécutives, de 1772 à 1792 inclusivement etc ..." nous relevons les points suivants :*

"24°) La mortalité des enfants dans le cours de la première année de la vie est à elle seule une moitié de celle des cinq premières, et un quart de la mortalité totale ...

28°) Depuis un an jusqu'à cinq ans la petite vérole est de toutes les causes celle qui a le plus de part à la mortalité ..."

1.5.2. La théorie des erreurs :

Tandis que l'actuariat, science terrestre et mortelle s'il en fût, annonce par la collecte d'abondantes données les recherches qui nous intéressent aujourd'hui, la méthode des moindres carrés née de la mesure des mouvements des astres nous élève jusqu'aux principes de la géométrie multidimensionnelle. C'est pour trouver d'après une série d'observations les cinq paramètres du mouvement d'une comète que fut inaugurée la méthode des moindres carrés, oeuvre de trois mathématiciens que, pour éviter de trancher des controverses de priorité, nous citerons dans l'ordre alphabétique de leurs noms : ce sont Gauss (1777-1855) Laplace,(1749-1827) et Legendre (1752-1833). Remettant au § 1.5.3 les méthodes analytiques, exposons premièrement les principes géométriques selon l'esprit de R. Fisher (§§ 2.3.2, sqq.). Un ensemble I de n mesures $\{x^1, x^2, \ldots, x^i, \ldots, x^n\}$ peuvent être considérées comme les n coordonnées d'un point x^I de l'espace R^n à n dimensions. Si ces n mesures proviennent d'un système physique assujetti à des équations de forme connue, mais où rentrent plusieurs paramètres inconnus, le point de par les équations sera assujetti à être dans un certain sous-espace V de R^n. (Le cas le plus simple, qui n'est pas dépourvu d'intérêt, est celui où les n nombres $x^1, x^2, \ldots, x^i, \ldots, x^n_i$ sont des mesures de la même grandeur : on doit alors avoir $x^1 = x^2 = \ldots = x^i = \ldots = x^n$, c'est-à-dire que V est la diagonale de R^n). En fait les mesures sont entachées d'erreurs ; et le point x^I n'est pas exactement dans V ; on corrigera donc les mesures en substituant à x^I un point y^I situé dans V et aussi proche que possible de x^I. Dans son mémoire de 1806 intitulé "Nouvelles méthodes pour la détermination des orbites des comètes - avec un Appendice sur la méthode des moindres carrés", A.M. Legendre propose de déterminer les paramètres inconnus de telle sorte que soit minima la somme des carrés des restes dans les équations qui devraient lier les x^i(i.e. au lieu de $f_\alpha(x^1, \ldots, x^n) = 0$ on a $f_\alpha(x^1, \ldots, x^n) = \epsilon_\alpha$; on demande que soit minimum $\Sigma\epsilon_\alpha$, somme qui dépend des paramètres du phénomène étudié). En bref, ceci revient à prendre y^I dans V aussi proche que possible de x^I au sens de la géométrie euclidienne associée à une norme $\|x^I - y^I\|$ convenable ; ou encore à projeter orthogonalement x^I sur V. Cette pratique requiert une justification fournie par la loi de la distribution d'un point x^I dont les coordonnées sont entachées d'erreurs autour d'un point central x^I_0. Cette loi est appelée loi normale ou loi de Laplace-Gauss (*). D'abord rencontrée par Moivre pour le cas unidimensionnel (c'est-à-dire le cas d'un nombre unique distribué autour d'une valeur moyenne ; cf § 4), elle fut ensuite considérée pour plusieurs variables simultanément (**).

(*) *Selon K. Pearson (Biometrika, T. 13, 1920 ; et Studies), "Laplace anticipated Gauss by some 40 years" car en 1783 il proposa de tabuler l'intégrale :*

$$(2\pi)^{-\frac{1}{2}} \int_x^\infty e^{-x^2/2}\, dx.$$

(**) *Sur la loi normale multidimensionnelle chez Laplace, cf infra § 1.5.4. Disons tout de suite qu'au XIXème siècle, cette loi bien reconnue, au moins implicitement, par les plus grands, est ignorée des praticiens des moindres carrés. La méthode de Laplace, fondée sur le théorème central limite, ne requiert pas la normalité de x^I ; car il démontre la normalité de son estimateur.*

Les titres des mémoires de Gauss (qui ne publie qu'après Legendre, mais qui, selon Laplace, "avait eu, plusieurs années avant la publication [de Legendre] la même idée dont il faisait un usage habituel") valent d'être traduits, car ils suggèrent le progrès de sa pensée : Theoria motus corporum coelestium, 1809 - Théorie du mouvement des corps célestes ; Disquisitio de elementis ellipticis Palladis, 1810 - recherche sur les éléments elliptiques de Pallas (une petite planète) ; Theoria combinationis observationum erroribus minimis obnoxiae, 1821 - Théorie de la combinaison des observations sujette aux moindres erreurs. Cependant Laplace dans sa Théorie analytique des probabilités (1° éd. : Paris 1812) expose indépendamment de Gauss et de Legendre la méthode des moindres carrés, en l'appuyant de théorèmes sur les moyennes d'erreurs qui annoncent les recherches contemporaines sur les sommes de variables aléatoires indépendantes (théorème central limite). A la fois grand analyste, grand mécanicien, grand probabiliste et ce qui - singulièrement dans la langue de ce temps - dit plus que tout : grand géomètre, Laplace était tout désigné pour de tels travaux (cf 1.5.3).

La méthode des moindres carrés a servi grandement l'élaboration des mesures numériques : d'abord en astronomie (*), puis en géodésie (où la somme des angles d'un triangle n'est jamais exactement ce qu'elle devrait être !) et dans toute la physique. La loi normale multidimensionnelle a été prise - non sans abus croyons-nous - pour le modèle universel de tous les êtres complexes jusqu'en biologie (cf § 2.2) et en psychologie (cf § 2.4). Mais elle a fourni à F. Galton et à K.Pearson le cas typique sur lequel leur imagination a saisi corrélation, régression et analyse factorielle.

(*) *Bessel que nous citons d'après J. Bertrand, a écrit : "In der Astronomie ist die Praxis eine Aufgabe der Warscheinlichkeits - Rechnung, die Theorie eine Aufgabe der hoheren Mechanik". Un exemple saisissant montrera que la confrontation des mesures astronomiques requérait impérieusement une théorie des erreurs. A la fin du XVIIème siècle, Halley avait proposé pour mesurer la parallaxe solaire p (angle sous lequel est vu du soleil un rayon équatorial de la terre) une méthode dont la mise en oeuvre semblait très simple, fondée sur la mesure des temps de passage de Vénus sur le disque solaire, vus de deux points éloignés de la terre. La durée de ces passages pouvant être de quelques heures, on espérait sur p (8"< p < 9") une erreur inférieure à 0,01". En 1761, l'occasion se présenta enfin d'observer un passage de Vénus sur le soleil. Les astronomes recueillirent non deux mais soixante observations de temps de passage. "60 observations au lieu de 2, dit J. Bertrand, faisaient espérer par leurs combinaisons 1770 (C_{60}^{2}) déterminations identiques. La déception fut grande ; les résultats variaient entre 7" et 11" ". Les astronomes tentèrent en vain de réduire l'aléa en choisissant les observations qui semblaient les plus appropriées. "En combinant les observations sans règle et sans méthode, les calculateurs du XVIIIème siècle n'en purent montrer que l'incertitude". Encke appliquant en 1821 la méthode des moindres carrés à l'ensemble des observations de 1761, trouva p = 8,49" ; la valeur admise aujourd'hui est 8,80" ; son écart à 8,49" représente plus de trois fois l'écart-type estimé par Encke : ce qui est considérable. J. Bertrand propose d'améliorer l'estimation de p en éliminant a posteriori les observations qui conduisent aux résultats les plus excentriques, attribuant celles-ci à des erreurs manifestes, non à des imprécisions : c'est avant la lettre ce qu'on appelle aujourd'hui estimation robuste (cf § 2.3.3 in fine) méthode conçue pour s'affranchir des perturbations par lesquelles la nature s'écarte le plus du modèle normal.*

Et il nous importait de voir conclue dès le début du XIXème siècle, entre la statistique et la géométrie euclidienne multidimensionnelle, une union dont le calcul électronique accroîtrait un jour la fécondité.

1.5.3. *Explication mathématique de la loi normale* :

Chez Legendre, la règle de rendre minima la somme des carrés des restes n'a pas de justification probabiliste : c'est un principe géométrique. Chez Gauss et Laplace, la méthode des moindres carrés est fondée sur la distribution normale des erreurs de mesure. Il est curieux de s'arrêter à voir comment ces deux illustres auteurs justifient une telle loi dont la fortune devait être grande dans l'histoire de la statistique (sur l'usage de l'hypothèse de normalité cf § 2 ; notamment § 2.2.5) ; "loi à laquelle tout le monde croit" car, selon une boutade de l'astronome Lippman citée par Poincaré, "les expérimentateurs s'imaginent que c'est un théorème de mathématiques, et les mathématiciens que c'est un fait expérimental".

Gauss part du problème de l'estimation de la vraie valeur z d'une grandeur d'après une série de mesures $x_1, \ldots, x_n$. Supposant que la loi de probabilité de la mesure x d'une grandeur dont la vraie valeur est z_0, est de la forme $\varphi(x-z_0)dx$ (où φ est une fonction à déterminer qui résulte du processus de mesure), Gauss décide d'estimer z par ce que nous appelons après Fisher (cf 2.3.3) le principe du maximum de vraisemblance (*) : i.e. de choisir z rendant maximum le produit :

$$\varphi(x_1 - z) \times \varphi(x_2 - z) \times \ldots \times \varphi(x_n - z).$$

De plus Gauss estime naturel le <u>postulatum</u> que l'optimum z est exactement la moyenne des nombres $x_1 \ldots x_n$. Le <u>postulatum</u> écarte l'éventualité d'une erreur systématique constante : Gauss dit explicitement : Errorum regularium consideratio proprie ab instituto nostro excluditur. En demandant que le <u>postulatum</u> s'accorde avec la règle du maximum de vraisemblance quels que soient les $x_1, \ldots, x_n$, il obtient par un calcul simple que $(\text{Log}\,\varphi)'$ est une fonction linéaire donc Log φ une fonction quadratique. Il retrouve ainsi une formule d'abord rencontrée par Moivre (cf § 1.4.1). Mais bien que les considérations de Gauss nous intéressent fort, elles ne démontrent nullement la nécessité physique de la loi normale ; aussi ne s'étonnera-t-on pas que selon J. Bertrand (1822-1900) (<u>Calcul des Probabilités</u> ; 1° éd. 1899) "Ni le succès près des observations de sa loi de probabilité des erreurs, ni la simplicité des conséquences, ni leur accord constant avec les faits n'ont décidé son illustre auteur à y voir une vérité démontrée".

Laplace, au contraire donne en substance de la loi normale la justification qui demeure la meilleure que l'on puisse proposer aujourd'hui : le théorème central limite. Ce théorème ne fut aucunement goûté des contemporains ; et aujourd'hui c'est la démonstration de Laplace qui semble illisible...

(*) *Principe auquel D. Bernoulli fut sans doute le premier à recourir : cf <u>Biometrika</u> T. 48 pp 1-18, et <u>Studies</u> un texte de D. Bernoulli (publié en 1777) suivi d'un appendice d'Euler ; et commenté par M.G. Kendall.*

Il faut reconnaître avec J. Bertrand que "le Traité analytique du calcul des probabilités commence par deux cents pages, au moins, dans lesquelles l'exposition des théories mathématiques qui doivent servir au calcul des chances est complètement indépendante de toute application ultérieure". Mais nous nous garderons de suivre J. Bertrand quand il ironise : "Laplace, après avoir trouvé des méthodes nouvelles devait leur donner la préférence : les problèmes sont choisis et les solutions proposées de manière à mettre en évidence l'utilité des fonctions génératrices". Lisons Laplace :

"Soit y_x une fonction quelconque de x ; si l'on forme la suite infinie $y_0+y_1\ t+y_2\ t^2+y_3\ t^3+\ldots+y_x\ t^x+y_{x+1}\ t^{x+1}+\ldots+y_\infty\ t^\infty$ on peut toujours concevoir une fonction de t qui, développée suivant la puissance de t, donne cette suite : cette fonction est ce que je nomme fonction génératrice de y_x".

Laplace est un analyste d'avant Cauchy : "il peut toujours concevoir" : il n'ignore certes pas que toute série n'est pas convergente ; mais il ne s'arrête pas à cette question ; un contemporain dirait qu'il calcule sur des séries formelles. Laplace voit plus loin : avec une ambiguité calculée il utilise la lettre x, réservée d'ordinaire à une variable continue, pour désigner ce qui semble n'être qu'un indice entier positif ; mais il écrit y_x et non y(x) sa fonction quelconque. Lui-même a professé dans son §1: "tel est l'avantage d'une langue bien faite, que ses notations les plus simples sont devenues souvent la source des théories les plus profondes ..." ; c'est une telle langue que veut écrire Laplace. Poursuivons :

"... u étant la fonction génératrice de y_x, celle de y_{x+r} est u/t^r ; ... le coefficient de t^x dans u((1/t)-1) est donc égal à $y_{x+1}-y_x$, ..., différence que nous désignerons par Δy_x, Δ étant la caractéristique des différences finies".

De la suite des valeurs entières de x, Laplace peut passer à une progression arithmétique de raison quelconque ; puis à une progression arithmétique de raison infinitésimale dx : le calcul des différences finies se résoud dans le calcul différentiel et celui-là fournit à celui-ci un calcul symbolique que retrouvera Heaviside (1850-1925). Làplace écrit par exemple :

$$'\Delta y_x = e^{\alpha dy_x/dx} - 1 ;$$

l'accent devant Δ signifie qu'il s'agit de la différence $y_{x+\alpha}$(et non $y_{x+1} - y_x$) ; c'est notre formule : $f(x+\alpha) - f(x) = (\exp(\alpha d/dx) - 1)\ f(x)$; cette formule s'obtient par passage à la limite à partir de $\Delta \approx ((1/t) - 1)$; $\Delta^n \approx ((1/t) - 1)^n$; $'\Delta = ((\alpha/t) - 1)$ etc ... Laplace sait qu'en développant son exponentielle il retrouve la formule de Taylor.

On attribue à Laplace la tranformation qui à la fonction α associe φ définie par

$$\varphi(s) = \int_0^\infty e^{-st}\ d\alpha(t) ;$$

on pourrait aussi lui attribuer la fonction caractéristique d'une loi de probabilité de densité p(x) :

$$f(t) = \int e^{itx}\ p(x)\ dx.$$

Car Laplace considère la fonction génératrice d'une loi de probabilité (alors $y_x = p_x$), il passe à la limite (loi $p(x)\,dx$ et non variable aléatoire discrète p_0, p_1, ...), et donne à l'argument t de sa suite $\Sigma p_x t^x$ des valeurs complexes de module 1 : $t = \exp(i\varphi)$. C'est par cette voie qu'il démontre que le quotient par $(n^{\frac{1}{2}})$ de la somme de n erreurs $x^1, \ldots, x^n$ chacune distribuée suivant la même loi quelconque (de variance finie), tend pour n infini à être distribuée normalement. C'est bien notre théorème central limite (cf Théorie analytique des pr. L II ch 4).

Mais est-il vrai que toute mesure physique, étant elle-même résultante additive d'une infinité de petites causes soit distribuée normalement ? Nous en douterons avec Poincaré ; qui sous le dessin d'une courbe en cloche écrit : "Ce résultat se vérifie, paraît-il. Ainsi Bessel (1784-1846) a pu représenter les résultats d'un grand nombre d'observations ..." (cf infra § 2.2.6 la critique que fait Pearson d'une assertion imprudente d'Airy).

1.5.4. *La loi normale multidimensionnelle chez Laplace* :

Pour un statisticien quelque peu géomètre et qui écrit après Fisher (cf § 2.3), justifier la méthode des moindres carrés requiert impérieusement le recours aux variables normales multidimensionnelles. Pourtant la pratique et la théorie des moindres carrés furent largement reçues dans le monde scientifique près de cent ans avant la loi normale multidimensionnelle. En 1846, Auguste Bravais (cf infra § 2.2.3) écrit que "... l'illustre Laplace ... s'est borné à l'appréciation de la possibilité des erreurs de la mesure finale d'une longueur" (entendez d'une seule longueur) ; et il croit être le premier à traiter de la loi conjointe de plusieurs grandeurs. A la fin du XIXème siècle Galton ($\approx$ 1885) redécouvre la loi normale multidimensionnelle ; et selon K. Pearson (cf § 2.2.3) il est le premier à lui donner sa juste place dans la philosophie naturelle.

Or la loi normale multidimensionnelle - sinon la philosophie qu'on en peut faire - est indubitablement dans la Théorie analytique des Probabilités : mais elle y est dans la justification de la méthode des moindres carrés (au Livre II) ; et des prétentions probablement sincères de Bravais, nous devons conclure que les démonstrations de l'illustre Laplace étaient plus admirées que lues ! Sans nous astreindre, ô lecteur, à lire Laplace comme il le mériterait, considérons au moins quelques citations qui suffisent à démontrer que cet auteur a connu la loi normale multidimensionnelle.

"Supposons maintenant que l'on ait deux éléments à corriger pour l'ensemble d'un grand nombre d'observations ..." Laplace considère que l'on dispose pour ces deux grandeurs d'une valeur approchée qu'il adopte pour origine et note z et z' les corrections qu'il convient d'apporter à ces valeurs ; corrections qu'il déterminera d'après un ensemble de mesures qu'il note $\beta^{(i)}$. Si la précision des mesures était parfaite, $\beta^{(i)}$ serait une fonction exacte des deux grandeurs donc des variables z et z'. Dans le domaine de l'étude, z et z' sont censés être assez petits pour qu'on puisse linéariser cette fonction ; donc poser :

$$\beta^{(i)} = h^{(i)} + p^{(i)}\, z + q^{(i)}\, z' ;$$

en fait, $\beta^{(i)}$ est à affecter d'une correction inconnue $\epsilon^{(i)}$; et la relation véritable est :

$$\beta^{(i)} + \epsilon^{(i)} = h^{(i)} + p^{(i)}\, z + q^{(i)}\, z'$$

en notant $\alpha^{(i)} = \beta^{(i)} - h^{(i)}$, Laplace obtient ce qu'il appelle les équations de condition :

$$\epsilon^{(i)} = p^{(i)}\ z + q^{(i)}\ z' - \alpha^{(i)}\ ;$$

il importe de noter que dans une telle équation, $p^{(i)}$, $q^{(i)}$, $\alpha^{(i)}$ sont connus ; $\epsilon^{(i)}$ est inconnu, mais Laplace suppose que tous les $\epsilon^{(i)}$ sont des variables aléatoires indépendantes de même loi ; enfin z et z' doivent être estimés d'après le système indicé par i des équations de condition.

Ce système étant surabondant, Laplace propose de le ramener à deux équations, obtenues par combinaison linéaire : conservons - comme on garde à un tableau de maître son cadre - les notations de Laplace, qui écrit S $a^{(i)}$ là où nous écrivons $\Sigma\{a^{(i)} | i\epsilon I\}$; le système réduit est :

$$S\ m^{(i)}\ \epsilon^{(i)} = z.\ S\ m^{(i)}\ p^{(i)} + z'.\ S\ m^{(i)}\ q^{(i)} - S\ m^{(i)}\ \alpha^{(i)}\ ;$$

$$S\ n^{(i)}\ \epsilon^{(i)} = z.\ S\ n^{(i)}\ p^{(i)} + z'.\ S\ n^{(i)}\ q^{(i)} - S\ n^{(i)}\ \alpha^{(i)}\ ;$$

"Si l'on suppose nulles les deux fonctions S $m^{(i)}\ \epsilon^{(i)}$ et S $n^{(i)}\ \epsilon^{(i)}$,..., les deux équations finales précédentes donneront les corrections z et z' des deux éléments. Mais ces corrections sont susceptibles d'erreurs, relatives à celle dont la supposition que nous venons de faire est elle-même susceptible. ... Il faut maintenant déterminer les facteurs m, $m^{(i)}$,..., n, $n^{(i)}$,..., de manière que l'erreur moyenne à craindre sur chaque élément soit un minimum".

Laplace applique sa méthode des fonctions caractéristiques et écrit pour la densité de la loi conjointe de S $m^{(i)}\ \epsilon^{(i)}$ et S $n^{(i)}\ \epsilon^{(i)}$ désignés respectivement par l et l', la formule même de la loi normale, formule que nous citons telle quelle :

$$\frac{1}{\frac{4\ k''\ \pi}{k}\ a^2\sqrt{E}}\ c^{-\frac{k}{4k''a^2}\ \frac{l^2.S\ n^{(i)2} - 2\ ll'\ S\ m^{(i)}n^{(i)} + l'^2.S\ m^{(i)2}}{E}}$$

où c est la base des logarithmes népériens (notée e aujourd'hui) k",k et a^2 des constantes qui lui donnent la variance commune à toutes les erreurs $\epsilon^{(i)}$ et E dépend de m et n.

$$E = S\ m^{(i)2}.\ S\ n^{(i)2} - (S\ m^{(i)}\ n^{(i)})^2.$$

... "pour avoir la probabilité que les erreurs u et u' des corrections des éléments seront comprises dans les limites données, il faut substituer... au lieu de l et l' leurs valeurs en u et u' ". Laplace n'a pas de peine à tirer ces valeurs du système linéaire à deux inconnues auquel il a réduit les équations de conditions ; il obtient donc pour la loi conjointe de u et u' une loi normale où figurent en paramètres les $m^{(i)}$, $n^{(i)}$: il montre alors que la supposition

$m^{(i)} = \mu\, p^{(i)}$; $n^{(i)} = \mu\, q^{(i)}$ (où μ est constant) rend minima l'erreur à craindre sur z et z' ; et calcule avec ces coefficients des corrections dont il écrit :

"Il est facile de voir que ces corrections sont celles que donne la méthode des moindres carrés des erreurs des observations, ou du minimum de la fonction

$$S(p^{(i)}z + q^{(i)}z' - \alpha^{(i)})^2 ;$$

d'où il suit que cette méthode a généralement lieu, quel que soit le nombre des éléments à déterminer ; car il est visible que l'analyse précédente peut s'étendre à un nombre quelconque d'éléments".

Un, deux, l'infini : l'illustre Laplace avait aux pieds les bottes de sept lieues de l'analyse ! Pour ceux qui sont moins agiles, voici une interprétation géométrique de sa démonstration. Considérons dans l'espace R^I le vecteur $\alpha^I = \{\alpha^{(i)} | i \in I\}$, dont les coordonnées sont les mesures primaires (à un changement d'origine près). S'il n'y avait pas d'erreur d'observation, le point α^I serait sur la sous-variété linéaire V^I, d'équation paramétrique en (z,z')

$$\forall\, i \in I : \alpha_0{}^{(i)} = p^{(i)}\, z + q^{(i)}\, z' ;$$

en fait, α^I est distribué normalement autour du point central $\alpha_0{}^I$; en déterminant les paramètres z et z' par les deux équations linéaires :

$$0 = z \,.\, S\, m^{(i)}\, p^{(i)} + z' \,.\, S\, m^{(i)}\, q^{(i)} - S\, m^{(i)}\, \alpha^{(i)}$$

$$0 = z \,.\, S\, n^{(i)}\, p^{(i)} + z' \,.\, S\, n^{(i)}\, q^{(i)} - S\, n^{(i)}\, \alpha^{(i)} ,$$

Laplace effectue une projection de α^I sur V^I ; avec le choix $m^{(i)} = p^{(i)}$ et $n^{(i)} = q^{(i)}$ il réalise exactement une projection orthogonale.

1.5.5. *La loi normale en théorie cinétique des gaz* :

Anticipant sur le cours du XIXème siècle, montrons ici par quels arguments mathématiques, de très grands physiciens ont justifié l'introduction de la loi normale dans un domaine où sa validité est acceptée : la théorie cinétique des gaz.

Désignons par p(u,v,w) du dv dw la probabilité que le point ayant pour coordonnées les trois composantes de la vitesse d'une molécule de gaz soit contenu dans l'élément de volume du dv dw centré au point (u,v,w). Pour démontrer que le vecteur vitesse est distribué normalement, Maxwell (1831-1879) affirme d'abord par raison de symétrie que la fonction de densité p ne dépend que du carré du module de la vitesse : $p(u,v,w) = F(u^2+v^2+w^2)$; pour préciser la forme de F, Maxwell fait encore une hypothèse dont la justification physique est beaucoup plus difficile : il suppose que les trois composantes de la vitesse sont des grandeurs aléatoires indépendantes entre elles.

Si donc $f(u^2)$ du est la loi de probabilité d'une composante on a :

$$F(u^2 + v^2 + w^2) = f(u^2)\ f(v^2)\ f(w^2) = (f(0))^2\ f(u^2 + v^2 + w^2)\ ;$$

des propriétés multiplicatives de la fonction f, on déduit alors que :

$f(u^2) = \exp(\alpha + \beta\ u^2)$ (où pour assurer la convergence β doit être négatif). C'est bien la densité de la loi normale que Maxwell a caractérisée dans l'espace par deux propriétés simples : la symétrie sphérique ; et l'indépendance en probabilité des composantes dans des directions orthogonales. Depuis lors (1859), on appelle loi de Maxwell, l'assertion que dans un gaz le vecteur vitesse d'une molécule est distribué normalement.

Boltzmann (à partir de 1868) entreprend de donner de la loi de Maxwell une démonstration physique plus profonde. Ayant admis la symétrie sphérique, il donne l'équation d'évolution temporelle de la fonction F. Ignorant tout des lois de l'interaction des molécules, Boltzmann adopte d'abord le modèle des molécules sphériques se déplaçant en ligne droite et changeant de direction par choc élastique les unes contre les autres (ou contre les parois) ; non qu'il croit à ce modèle, mais il préfère fixer une analogie mécanique, plutôt que de s'abandonner à la généralité indéfinie des formules mathématiques. Boltzmann vérifie que la loi de Maxwell est une solution stationnaire de l'équation d'évolution, (c'est-à-dire que si à un instant donné la distribution des vitesses est normale, les chocs élastiques qui modifient les vitesses des molécules individuelles, respectent dans l'ensemble cette distribution). Puis il démontre qu'au cours du temps l'intégrale $H = -\int F \log F$ du dv dw ne peut décroître. Compte tenu de ce que F est une loi de probabilité ($\int F$ du dv dw = 1) et de ce que l'énergie cinétique est constante dans le temps ($\int (u^2+v^2+w^2)F$ du dv dw = Cte), Boltzmann établit alors par un calcul de variation très simple que le maximum de H est atteint pour la distribution normale des vitesses qui est donc la loi limite que tendent à instaurer les chocs entre molécules (parce que H ne peut décroître). Enfin Boltzmann relie l'intégrale H à la probabilité de la distribution F, entendue comme le nombre de répartitions des molécules entre classes de vitesses, compatibles avec une valeur fixée de l'énergie totale (cf <u>supra</u>) : (il s'agit ici de passage à la limite infinie sur des calculs d'analyse combinatoire ; calculs que la mécanique quantique a éclairés et assurés en montrant que les états possibles d'un système de molécules forment un ensemble discret (non continu) qui se prête donc à des dénombrements) ; on débouche sur la définition de l'entropie.

Einstein (1905) considère le mouvement au sein d'un fluide d'une molécule (ou d'une particule microscopique, petite mais visible, donc bien plus grosse qu'une molécule) agitée par des chocs thermiques incessants. Sous l'hypothèse que les déplacements effectués par cette particule au cours d'intervalles de temps successifs (de longueur petite : τ) sont aléatoirement indépendants, Einstein obtient l'équation différentielle d'évolution de la distribution de probabilité de la particule : il en déduit que le vecteur espace parcouru pendant un temps T fixé a une distribution normale sphérique dont la variance est proportionnelle au temps T. Ce que Jean Perrin a vérifié dans ses expériences extraordinairement ingénieuses sur le mouvement brownien.

Gauss, Laplace, Maxwell, Boltzmann, Einstein : combien de génies une formule écrite par Moivre n'a-t-elle pas arrêtés ! Et nous verrons cette même formule accompagner la biométrie (§ 2.2) et la psychométrie (§2.4) dans la conquête de l'espace multidimensionnel.

1.6. *Un siècle stupide* :

Tel fut dit-on le XIXème siècle. Il ne convient pas d'invectiver un âge qui enseveli dans le passé, ne peut nous répondre : mais stupide vient de stupor, et il est véritable qu'une sorte de stupeur, de doute sembla frapper pour un temps la science mathématique du hasard.

1.6.1. *Us et abus des probabilités* :

Le lecteur des oeuvres de Laplace admirera la puissance des méthodes analytiques qu'il a fondées. Il s'étonnera de la patience du calculateur compilant les observations les plus antiques et conjuguant mécanique et statistique pour démêler l'écheveau des mouvements des astres. Témoin l'étude de l'Equation de la Lune que son auteur expose ainsi (cf Essai philosophique). "Ce fut encore par la considération des probabilités que je reconnus la cause de l'équation séculaire de la lune. Les observations modernes de cet astre, comparées aux anciennes éclipses, avaient indiqué aux astronomes une accélération dans le mouvement lunaire ; mais les géomètres, et particulièrement Lagrange, ayant inutilement cherché, dans les perturbations que ce mouvement éprouve, les termes dont cette équation dépend, ils la rejetèrent. Un examen attentif des observations anciennes et modernes et des éclipses intermédiaires observées par les Arabes, me fit voir qu'elle était indiquée avec une grande probabilité. Je repris alors sous ce point de vue la théorie lunaire, et je reconnus que l'équation séculaire de la lune est due à l'action du soleil sur ce satellite, combinée avec la variation séculaire de l'excentricité de l'orbe terrestre ; ce qui me fit découvrir les équations séculaires des mouvements des noeuds et du périgée de l'orbite lunaire, équations qui n'avaient même pas été soupçonnées par les astronomes. L'accord très remarquable de cette théorie avec toutes les observations anciennes et modernes l'a portée au plus haut degré d'évidence". Tout est à méditer dans ce texte pour le statisticien ; même s'il ne voit pas distinctement les révolutions de la lune décrivant autour de la terre des ellipses dont la forme et l'orientation dérivent lentement. Soulignons la part de la statistique dans le faisceau des preuves qui convergent vers l'évidence. Laplace possède déjà les notions de risque et d'utilité telles qu'elles fleuriront après lui en statistique. Il fut encore un éminent démographe (nous avons cité au § 1.5.1 son rapport du 26 Floréal an IV) : on n'a pas vu depuis (on ne reverra peut-être jamais plus) un mathématicien de cette grandeur interroger si passionément les humbles chiffres. Mais Laplace ne fut pas absolument heureux en tout. Quand entreprenant l'application des Probabilités aux sciences morales, il traite de la Probabilité des témoignages, des choix et des décisions des assemblées, de la Probabilité des Jugements des tribunaux (titre et sous-titres sont de Laplace lui-même dans son Essai Philosophique ...) ses calculs nous semblent plus ingénieux que ses hypothèses ne sont sûres. Laplace, il est vrai, ne s'abandonnait pas aveuglément à l'illusion de contribuer au progrès moral, mais il concluait à l'utilité de ses spéculations. Ecoutons-le : "La plupart de nos jugements étant fondés sur la probabilité des témoignages, il est bien important de la soumettre au calcul. La chose il est vrai devient souvent impossible ... Mais on peut dans plusieurs cas résoudre des problèmes qui ont beaucoup d'analogie avec les questions qu'on se propose, et dont les solutions peuvent être regardées comme des approximations... Une approximation de ce genre lorsqu'elle est bien conduite, est toujours préférable aux raisonnements les plus spécieux". C'est cependant parmi ces raisonnements que certains critiques n'hésitent pas à ranger les approximations de Laplace ! Et J. Bertrand leur trouve chez Rabelais ce divertissant précédent

"Cela fait, comment sentenciez-vous mon ami ?

"Comme vous autres, messieurs, répondit Bridoye, pour celuy je donne sentence duquel la chance livrée par le sort des dez judiciaires premier advient".

Selon Bertrand, "Tout se passe, suivant Condorcet comme si les magistrats imitant Bridoye, juge de Myrelingues sentenciaient par le sort des dés. L'assimilation de l'opinion d'un juge au tirage d'une boule dans une urne de composition déterminée est pour lui une identité. Si la boule est blanche, la décision sera bonne. Le juge se trompera s'il tire une boule noire ... Le difficile est de trouver la composition de l'urne. ... Laplace ... assimile, comme Condorcet, l'opinion d'un juge à un tirage fait dans une urne, mais il repousse l'hypothèse d'une probabilité invariable".

1.6.2. Deux sévères verdicts :

Parvenu au seuil du XIXème siècle dans son chapitre sur l'histoire de la théorie des Probabilités, Gnedenko écrit : "Nous ne pouvons taire le fait que Laplace et Poisson (*) qui eurent un grand rôle dans le progrès de la théorie des probabilités, furent en même temps la cause indirecte de sa longue stagnation, caractéristique en Europe Occidentale de la deuxième moitié du XIXème siècle et des premières décennies du XXème. Ces deux illustres savants préconisèrent largement l'application de la théorie des probabilités aux <u>sciences morales</u> ... L'engouement pour ces soi-disant applications, faites au mépris de la nature des faits, sans égard aux conditions qui les déterminent pesa lourdemant sur le progrès de la théorie des probabilités. Toutes ces applications fondamentalement erronées furent qualifiées dans la suite de "scandale mathématique" (**). A cause de ces échecs, l'engouement pour la théorie des probabilités fit place au désenchantement ; et l'opinion se répandit largement parmi les mathématiciens de l'Europe Occidentale que cette théorie n'était qu'une sorte de divertissement mathématique sans applications importantes et ne méritant guère l'attention des savants sérieux. Même les succès de la théorie dans ses applications sérieuses - théorie cinétique des gaz (***), théorie des erreurs, théorie du tir etc ... - ne purent en Europe Occidentale venir à bout de cette erreur de jugement. Il fallut le génie de Tchebycheff (1821-1894) et la création par lui de l'école russe, ainsi que le progrès impétueux de la physique recourant plus qu'auparavant aux mathématiques en général et à la théorie des probabilités en particulier pour refaire de cette théorie une grande science étudiant par ses méthodes propres de grands ensembles de faits du monde matériel. Le progrès ultérieur de la théorie est indissolublement lié aux succès de notre propre science. Car c'est dans notre pays que furent obtenus les résultats fondamentaux dans cette étape du progrès de la théorie des probabilités".

Faisons dans cette vision du XIXème siècle et de l'Europe Occidentale la part du patriotisme russe exalté par une guerre formidable (Gnedenko écrit peu avant 1950) ; il reste qu'un autre auteur (dont le chauvinisme mathématique rivalise il est vrai avec plus d'un nationalisme) ne s'exprimerait pas autrement.

(*) *L'illustre mécanicien dont on se bornera à citer ici les "Recherches sur la probabilité des jugements en matière criminelle et en matière civile" (1837) et les "Mémoires sur la probabilité du tir à la cible" (ce dernier titre évoquant immédiatement la <u>loi</u> à laquelle nous donnons le nom de Poisson).*

(**) *L'expression est de Stuart Mill ; nous y reviendrons avec Joseph Bertrand.*

(***) *Dont on a traité au § 1.5.5, à propos de la loi normale.*

Dans le paragraphe final de la substantielle note historique au ch V de son livre sur l'Intégration, N. Bourbaki, place dans une parenthèse la condamnation des probabilistes du XIXème siècle et l'éloge des contemporains qui en sont les successeurs mais non les continuateurs, quand il affirme : ... "c'est du côté des applications qu'il faut chercher les progrès réalisés par la théorie de l'Intégration depuis 1920 : théorie des probabilités (autrefois prétexte à devinettes et paradoxes, et devenue une branche de la théorie de l'Intégration depuis son axiomatisation par Kolmogoroff, mais branche autonome avec ses méthodes et ses problèmes propres) ; théorie ergodique ; théorie spectrale et analyse harmonique, ..."

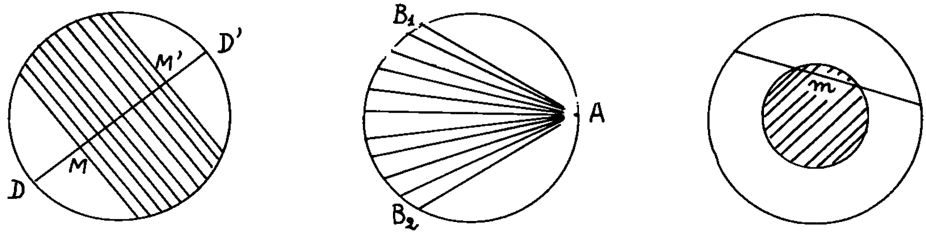
Raturant un siècle de leur plume acérée, Gnedenko et Bourbaki nous conduisent jusqu'aux recherches probabilistes contemporaines fondées sur la théorie de la mesure, sans même citer l'école statistique anglo-saxonne ; ce dont celui-ci est excusable vu son propos tout mathématique, mais non celui-là ! Gardons-nous de cette hâte : interrogeons quelques témoins : Joseph Bertrand (1822-1900), Henri Poincaré (1854-1912) ; Emile Borel (1871-1956) aussi, qui excelle à évoquer bien des faits et des pensées dans son modeste ouvrage sur Le Hasard (P.U.F. ; 1948 ; 1° éd. en 1914). Prenons divers exemples de ces scandaleux paradoxes.

1.6.3. *Les paradoxes du calcul des probabilités* :

Paradoxe de Le Dantec (biologiste : 1869-1917 cf Borel, op. laud. § 20) "Imaginons ... un joueur qui suit la règle suivante. Il joue une première partie ; s'il la gagne, il encaisse l'enjeu, disons 1 franc, et il s'arrête. Il reprend ensuite avec un autre joueur ; s'il perd la première partie, il continue jusqu'à ce que l'équilibre soit rétabli, puis il s'arrête ... Il peut continuer indéfiniment ... tantôt il gagnera 1 franc, tantôt il se retirera après une partie nulle. Il réalisera donc sûrement un gain croissant et même, affirmerait Le Dantec, proportionnel au temps". A ce paradoxe, on avait répondu par avance (cf § 1.4.1) que la ruine du joueur l'empêcherait d'attendre "jusqu'à ce que l'équilibre soit rétabli". Répétons que les jeux indéfinis dans le temps ont introduit l'étude des processus aléatoires : et on sait que le terme de martingale est passé des salles de jeu aux séminaires de mathématiques.

Les séquences au jeu de pile ou face (Borel, § 2.1) : "Nous appellerons séquence une série de coups gagnés par un même joueur précédée et suivie d'au moins un coup gagné par l'autre joueur. Toute partie se décompose ainsi en une suite de séquences successives ... Nous appelons valeur moyenne des séquences le quotient de la somme de toutes leurs longueurs [nombre total de coups] par leur nombre". Mais ajoute aussitôt Borel, "on pourrait donner de la longueur moyenne d'une séquence une définition différente qui se justifierait dans certaines circonstances pratiques". En bref chaque instant, c'est-à-dire chaque coup c est inclus dans une séquence ; dont a posteriori on peut connaître la longueur n(c) : on peut calculer la moyenne de n(c) sur l'ensemble des coups, ce qui revient à donner à chaque séquence dans le calcul de la moyenne, un poids égal à sa longueur.

La nécessité de préciser les poids relatifs des événements est plus cruciale encore dans les problèmes géométriques où il s'agit non d'ensembles finis mais de continuum. Voici un problème dont Joseph Bertrand donne trois solutions qui conduisent à des résultats différents : "On trace une corde au hasard dans un cercle ; quelle est la probabilité pour que sa longueur soit supérieure au côté du triangle équilatéral inscrit ?" Première solution, par raison de symétrie toutes les directions sont également probables pour la corde ; et la direction Δ étant fixée, il y a une chance sur deux pour que la corde dépasse le côté du triangle équilatéral (car Δ étant fixée, le milieu de la corde est sur le diamètre DD' ;



1ère solution 2ème solution 3ème solution

Figure 1-1 : un paradoxe du calcul des probabilités. déterminer la probabilité pour que la longueur d'une corde soit supérieure à celle du côté du triangle équilatéral inscrit

et la corde a la longueur requise si son milieu est sur la moitié centrale MM' de DD'). Deuxième solution : par raison de symétrie, tout point A du cercle a même chance d'être une extrémité de la corde ; et cette extrémité A étant fixée, il v a une chance sur trois pour que la corde dépasse le côté du triangle équilatéral (on le voit en considérant soit l'autre extrémité B, soit la direction de la corde). Troisième solution : tirer au sort la corde, revient à tirer au sort un point quelconque du disque pour être son milieu m ; la corde dépassera le côté du triangle équilatéral si le milieu est dans le disque de rayon moitié ; disque dont l'aire est (1/4) de l'aire du grand cercle ; donc, probabilité 1/4. "Doit-on penser, interroge E. Borel (op. laud. § 3.4) que ces trois solutions sont également bonnes et, par suite, également mauvaises ? Nullement, il s'agit simplement de préciser le mode d'après lequel se fera la vérification expérimentale, c'est-à-dire comment on s'y prendra pour tracer une corde au hasard dans un cercle". Borel conclut toutefois que "la plupart des procédés naturels ... conduisent à la première" [solution]. Pour nous cette solution s'impose : notre "raison de symétrie" est de munir l'ensemble (bidimensionnel) des droites du plan, d'une mesure invariante par déplacement : dès lors la mesure de l'ensemble des sécantes déterminant une corde de longueur supérieure à celle du côté du triangle équilatéral apparaît comme étant la moitié de celle de l'ensemble de toutes les sécantes du cercle ...

Généralisant jusqu'à l'extrême les paradoxes géométriques de Joseph Bertrand, Poincaré observe qu'il n'y a aucune raison de supposer qu'une certaine variable continue x est uniformément répartie (sur le segment (0,1) par exemple) plutôt que son carré x^2, ou toute autre fonction f(x) ; et a fortiori n'y-a-t-il pas de loi de probabilité naturelle pour un point sur une surface, ou dans un domaine multidimensionnel. "Arrivé là, dit Borel, tout autre que Poincaré se serait arrêté, semble-t-il, puisque le résultat est entièrement négatif. C'est ici au contraire, que Poincaré introduit une idée neuve et très intéressante". Partons de ce problème : "Un cadran circulaire est divisé en 2n parties égales peintes alternativement en blanc et en noir ; une aiguille mobile autour du centre du cadran est lancée avec une force suffisante pour faire avant de s'arrêter, plusieurs fois le tour du cadran ; quelle est la probabilité pour qu'elle s'arrête en face d'une division blanche ?

La réponse est (1/2), indépendamment de la position de départ, et de la loi précise du lancer, sous cette seule condition générale que l'aiguille fait plusieurs tours avant de s'arrêter. Le résultat d'un calcul de probabilités est certes d'autant plus précieux que comme dans l'exemple de cette roulette, il résulte nécessairement de simples propriétés qualitatives des données, quelles que soient les valeurs numériques précises des probabilités élémentaires postulées. Poincaré démontre par exemple que sous des hypothèses très générales quant aux probabilités des diverses permutations effectuées en battant les cartes, le battage répété tend à rendre équiprobables tous les rangements de celles-ci quel qu'en soit l'ordre initial.

Ces exemples montrent que les paradoxes où s'arrêtaient les savants du XIXème siècle (quels que fussent l'ironie ou le scepticisme de ces savants ; cf § 1.7.5) étaient bien autre chose que des divertissements ou des devinettes peu sérieuses. Pour les dépasser il restait à élaborer la théorie des processus ; à préciser les probabilités sur des espaces (et non seulement sur des ensembles finis ou discrets), et particulièrement la mesure invariante sur les groupes et les espaces homogènes ; à démontrer l'existence de distributions limite. Quant aux applications aventureuses du calcul des probabilités aux sciences morales, ne doit-on pas avec Joseph Bertrand, en excuser le calcul lui-même ? "L'application du calcul aux décisions judiciaires est, dit Stuart Mill, le scandale des Mathématiques. L'accusation est injuste. On peut peser du cuivre et le donner pour or, la balance reste sans reproche. Dans leurs travaux sur la théorie des jugements, Condorcet, Laplace et Poisson n'ont pesé que du cuivre". On a fait pire depuis : de nos jours dans les applications vulgaires des mathématiques aux sciences humaines, la balance est aussi peu fidèle que le métal est vil.

1.7. *Définition des probabilités* :

La boutade de Joseph Bertrand nous incite à analyser la locution familière de Calcul des Probabilités : d'abord des probabilités, puis un calcul adapté à ces données.

1.7.1. *Probabilités uniformes et principe de symétrie* :

Pour les fondateurs du XVIIème siècle, les probabilités de base sont celles d'un ensemble fini d'éléments équiprobables (pile et face ; ou les six chiffres d'un dé) ; le calcul dénombre les combinaisons de ces éléments et la probabilité de la réalisation d'une issue complexe est définie comme le rapport du nombre des combinaisons favorables à celui des combinaisons possibles. Dans les probabilités géométriques considérées dès le XVIIIème siècle (e.g. dans le problème de l'aiguille de Buffon cf § 4.1), l'ensemble de base est infini (e.g. toutes les positions d'un segment - l'aiguille - dans le plan) : chaque élément (e.g. telle position du segment) a donc une probabilité individuelle nulle, c'est par intégration d'une densité qu'on pourrait définir la probabilité d'un domaine (e.g. les segments dont la direction est comprise entre θ_1 et θ_2, et dont le milieu est dans tel disque du plan). Comme de poursuivre avec quelque rigueur de tels calculs entraîne fort loin, les mathématiciens résolurent d'abord les problèmes en les simplifiant par des considérations de symétrie très puissantes, mais on l'a vu, d'un maniement parfois délicat. Cependant la combinatoire discrète des jeux et la géométrie du hasard se ressemblent en ce qu'elles calculent sur des probabilités dont les valeurs s'imposent a priori : à l'équivalence des faces du dé, répond l'invariance par rotation et translation du problème de l'aiguille : dans ces problèmes les probabilités n'ont pas à être mesurées (par des essais répétés de jet de dé ou d'aiguille) ; le bon géomètre sait ce qu'elles doivent être. Tout autre sont les probabilités du statisticien : quand Daniel Bernoulli (cf § 1.5.1)) suppute l'influence de l'inoculation sur l'espérance de vie, il part d'une table de mortalité fondée sur des dénombrements, des cas réels ; et s'il postule que le risque variolique est uniforme au cours de la vie, c'est seulement faute de mieux non par foi en un principe d'invariance.

1.7.2. La mesure des fréquences :

Mesurer les probabilités d'une série d'issues deux à deux exclusives (e.g. mourir avant l'âge d'un an, entre un et deux ans ; entre deux et trois, etc ...) par leur fréquence sur un ensemble d'observations requiert une grande patience ; et ces mesures restent entachées d'un type d'erreur qui leur est propre : les fluctuations d'échantillonnage. Trouver que sur cent patients atteints de telle maladie trente ont bénéficié de tel remède, ne signifie pas que la probabilité d'efficacité de celui-ci soit exactement 30% ; même si les diagnostiques et la conduite du traitement sont irréprochables.

1.7.3. Probabilité subjective :

Aussi à la probabilité définie objectivement par la statistique comme limite idéale d'une fréquence observée, substituera-t-on souvent une simple estimation subjective. D'aucuns acceptent d'enthousiasme ces probabilités subjectives. "Pour introduire la notion de probabilité, on se réfère d'ordinaire à la notion de fréquence. Pourquoi ? interroge B. de Finetti (in Revue internationale de Statistique ; pp 117-130 ; T. 42 ; n°2 ; 1974) ... Ma réponse est très simple et naturelle, même si beaucoup la jugent paradoxale : la persistance de la définition en fréquence de la probabilité s'explique parce qu'elle est la pire possible". Bien qu'il cultive le paradoxe, B. de Finetti ne se croit pas dispensé de proposer en faveur de sa thèse les meilleurs arguments. Déterminer les probabilités par les fréquences est difficile, voire impossible. Sans probabilités a priori (cf supra § 1.4.2) on ne peut appliquer la formule de Bayes : que peut être la statistique non bayesienne (cf infra §§ 2.3.3. et 2.3.5.) ? Pour la dépeindre Finetti cite une boutade : "Ce sol n'est pas assez ferme : c'est du sable ; enlevons le sable et fondons la bâtisse sur le vide' Le sable : les probabilités subjectives ; le vide : ne plus utiliser la formule de Bayes. Souvent, comme dans le cas considéré par Henri Poincaré (cf § 1.6.3) les conclusions cherchées ne dépendent pas des valeurs précises des probabilités a priori mais seulement de leurs grandes lignes (c'est le cas de la théorie de l'estimation ; pourvu que l'on dispose d'un nombre suffisant d'observations ; mais alors, selon nous, l'abondance de celles-ci équivaut un peu à une estimation par les fréquences). Les probabilités subjectives sont-elles si imprécises que d'aucuns le disent ? Finetti emprunte un exemple à Lindley : "Un ingénieur chimiste, tout en admettant la possibilité d'une défaillance de l'appareillage dont il était responsable, hésitait à chiffrer une probabilité. Mais comme il connaissait les incidences financières d'une telle défaillance, je lui demandai : si je vous offrais un dispositif éliminant tout aléa, combien le payeriez-vous : mille dollars, dix mille ? Cette dernière somme lui parut ridiculement exagérée, mais la première l'intéressa plus sérieusement. Après marchandage on s'arrêta à 750 ; nombre que (connaissant la perte en jeu) on peut convertir en une probabilité". (A ce marchandage, fondé en effet sur une certaine perception des probabilités, nous objecterons que les expériences des psychologues mettant des sujets humains aux prises avec un processus aléatoire font voir chez ces sujets un comportement qui n'est pas optimum). Enfin pour faire aux éléments subjectifs la part belle, Finetti conclut en rappelant que son ami, notre regretté collègue le Professeur Giuseppe Pompilj aimait à ce propos citer Pirandello : "Un fait est comme un sac : s'il est vide, il ne tient pas debout".

1.7.4. *L'existence objective des probabilités* :

Ceux qui avec nous s'inquiètent de ce subjectivisme désinvolte doivent ici s'interroger : objectivement, les probabilités existent-elles ? Certainement oui, s'il s'agit de probabilités égales sur un ensemble fini (les dés, pile ou face), ou de probabilités invariantes par un groupe continu de transformations : seuls restent dans l'ombre les écarts à l'uniformité (e.g. les irrégularités qui sont cause que l'expérience de l'aiguille de Buffon, cf § 4.1, ne fournit que deux ou trois décimales de π, même si on la prolonge longtemps). Mais une probabilité entendue comme limite de fréquence n'est observable avec précision que si les conditions sont assez stables : ce qui n'est pas le cas notamment si l'expérimentation transforme rapidement le phénomène étudié ; par exemple en psychologie où le sujet est en butte à la fatigue, et bénéficie, en revanche, d'un apprentissage (par le seul fait qu'il se soumet à l'expérience). Bien plus, la probabilité n'existe objectivement pas sans conditions stables. Selon nous, le modèle mathématique susceptible de rendre compte de l'instauration d'un équilibre probabiliste au sein d'un système complexe est celui de la théorie ergodique : on peut dire qu'on étudie un point (figurant le système décrit par de multiples coordonnées) qui se meut dans un espace de dimension très élevée ; sous certaines hypothèses, toute trajectoire tend à remplir uniformément tout l'espace, (muni d'un élément de volume convenable) ; ainsi on retrouve le cas initial des probabilités uniformes. Or d'une part l'équilibre probabiliste ne s'établit qu'en un temps dit temps de relaxation, à l'échelle duquel il importe que les conditions évoluent peu. D'autre part, parler d'équilibre, d'espace de configuration multidimensionnel ... présuppose une définition précise de l'objet qu'on étudie : et il faut répéter que la probabilité est la mesure de notre ignorance ; ignorance quant aux déterminations individuelles des faits, cela va de soi (comment le dé a-t-il été jeté ?) ; mais en plus, ignorance souvent du plan général, ce qui interdit de poser correctement le problème probabiliste, de définir l'objet. Ainsi objectivement, les probabilités dont nous parlons peuvent être floues : tous les arguments de l'école subjectiviste (B. de Finetti) sont donc bien-venus pour légitimer leur emploi.

1.7.5. *La méthode axiomatique* :

Pour les auteurs du XIXème siècle, douter de la mesure et même de l'existence objective des probabilités portait à douter des calculs dont celles-ci sont l'objet. Joseph Bertrand avec l'humour incisif qui lui est propre place en exergue à son traité une citation latine de Daniel Bernoulli d'une ambiguïté non fortuite : "Facile videbis hunc calculum esse saepe non minus nodosum quam jucundum". On verra facilement que souvent ce calcul n'est pas moins noueux [est-ce difficile, est-ce embarrassé ?] qu'agréable [veut-il dire fécond ou réjouissant ?]. Et Poincaré conclut en sage désabusé ses profondes leçons : "Le calcul des probabilités offre une contradiction dans les termes mêmes qui servent à le désigner [calculer sur notre ignorance même], et, si je ne craignais de rappeler ici un mot trop souvent répété, je dirai qu'il nous enseigne surtout une chose : C'est de savoir que nous ne savons rien".

Donc il importe de séparer la détermination des probabilités des règles de leur calcul. C'est ce qu'accomplit la théorie axiomatique proposée par A.N. Kolmogorov, et aujourd'hui très largement reçue. Gnedenko écrit (cours ; § 8) : "A la base de la théorie des probabilités entendue comme une science mathématique, on doit poser quelques prémisses, expression générale de l'expérience séculaire de l'homme.

Le développement ultérieur doit se faire par déduction à partir de ces prémisses sans recours à des représentations concrètes ni aux évidences du sens commun". Il est vrai qu'à tout ce que l'expérience séculaire nous offre comme problèmes probabilistes, on peut donner pour base un ensemble U d'événements élémentaires muni d'une structure d'espace mesurable, et d'une mesure positive de masse totale 1 (*). Les calculs ainsi fondés (qu'il s'agisse d'analyse combinatoire, de familles de lois dépendant de paramètres numériques ou de processus réquérant tout l'appareil de l'analyse fonctionnelle) sont rigoureux et clairs, et ils portent loin. On regrettera seulement que les jeunes générations formées à l'austère discipline de la théorie de la mesure, ne reçoivent plus dans les paradoxes à la Joseph Bertrand, la tradition de l'expérience séculaire et qu'on puisse être breveté probabiliste ou même statisticien, sans avoir appris par un prudent usage selon quels justes principes de dissection on découvre dans un problème concret un espace probabilisé.

1.7.6. *Probabilité et Statistique* :

L'Analyse des Données, quant à elle, reçoit du Calcul des Probabilités son inspiration mais non ses méthodes. Considérons le problème typique : découvrir le système des rapports entre lignes et colonnes d'un tableau rectangulaire Ce problème est fini dans ses données actuelles, même s'il débouche sur l'ensemble potentiellement infini des cas analogues à ceux recensés effectivement dans le tableau traité. C'est en ce sens qu'il faut entendre le slogan : Statistique n'est pas Probabilité (cf T II A n°1 § 1°). Parvenu au seuil du XXème siècle nous ne tenterons donc pas de poursuivre, même en gros, l'histoire des progrès si divers qu'a connus la science du hasard : nous ne dirons rien des principes probabilistes de la mécanique ondulatoire qui a pénétré les atomes et leurs noyaux, et de plus renouvelé la physique statistique (cf § 1.5.5) ; nous ne ferons que rencontrer (cf § 3.5.1) les méthodes d'analyse fonctionnelle propres au calcul des probabilités édifié désormais sur les fondements axiomatiques. Mais avant d'exposer la suite des recherches auxquelles nous avons participé (§ 3), nous tenterons d'illustrer un siècle de pensée statistique (§ 2).

(*) *Rappelons que la théorie de la mesure fut créée par Borel et Lebesgue au début du XXème siècle ; et que Borel en vit aussitôt l'importance pour le calcul des probabilités.*