

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

ANDRÉ PALLEZ

Normes de la statistique, du calcul des probabilités et des erreurs de mesure

Journal de la société statistique de Paris, tome 90 (1949), p. 125-133

http://www.numdam.org/item?id=JSFS_1949__90__125_0

© Société de statistique de Paris, 1949, 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/>

VII

NORMES DE LA STATISTIQUE, DU CALCUL DES PROBABILITÉS ET DES ERREURS DE MESURE

L'Association Française de Normalisation vient de publier les cinq normes suivantes :

X 05-001 — Terminologie de la Statistique.

Cette norme comprend trois parties :

— Avant-propos.

— Rappel de quelques définitions arithmétiques appliquées à la statistique.

— Caractère et série statistique à un caractère.

X 05-002 — Terminologie du calcul des probabilités.

X 05-003 — Terminologie de la statistique. — Séries statistiques à plusieurs caractères.

X 05-004 — Terminologie du calcul des probabilités. — Éléments aléatoires à plusieurs dimensions.

X 06-001 — Terminologie des erreurs de mesure.

Elle a publié également le fascicule de documentation suivant :

FD X n° 05-010. — Symboles de la statistique, du calcul des probabilités et des erreurs de mesure.

Ces six documents forment un ensemble cohérent. Ils auraient pu être présentés en un seul fascicule, mais pour faciliter leur diffusion, l'A. F. N. O. R. les vend séparément. Il faut toutefois noter que les normes ont été rédigées dans l'ordre suivant lequel elles sont numérotées et qu'il a été supposé que le lecteur d'une norme, portant un certain numéro, avait connaissance des normes précédentes.

Ces normes sont les résultats des travaux entrepris par la Commission de terminologie statistique, qui a été créée par l'Association Française de Normalisation au début de 1944, sous la présidence du regretté M. Huber, en vue de coordonner les différentes expressions couramment utilisées par ceux qui enseignent, qui étudient et qui utilisent la statistique.

Presque tous les membres de cette Commission font partie de la Société de Statistique de Paris, sans en excepter l'imprimeur à qui a incombé la tâche particulièrement délicate d'en assurer la composition.

Dès la première séance de la Commission, M. Fréchet insistait sur l'urgence de la normalisation des termes employés en statistique et en calcul des probabilités; il faisait remarquer que les travaux sur ces questions devenaient extrêmement nombreux et, comme la terminologie n'était pas bien fixée, il y avait le plus grand intérêt à établir des normes de terminologie.

En plus de M. Huber et M. Fréchet, les personnalités qui ont pris part à la rédaction des normes, sont notamment M. l'Inspecteur général Baticle, président

actuel de la Commission de terminologie statistique de l'A. F. N. O. R., M. Chapelon, M. Darmois, M. Delaporte, M. l'Ingénieur en chef Dumas et M. Krach, ingénieur en chef du Service des instruments de mesures.

Les travaux de la Commission de Terminologie statistique ont commencé au mois de mars 1944 et se sont poursuivis sans interruption. Pour faire un peu de statistique, je dirai que la Commission n'a pas tenu moins de 56 séances, représentant un total de 151 heures et que le temps global passé par les membres de la Commission est de 891 heures, sans compter ni le temps nécessaire à la préparation des projets, ni le temps passé par le personnel de l'A. F. N. O. R.

Il est essentiel pour bien juger ces normes de se souvenir :

d'une part, qu'il ne s'agit pas d'un traité, mais d'une terminologie où l'on devait nécessairement éviter toute position de principe concernant les sujets si discutés relevant des fondements philosophiques du Calcul des Probabilités;

d'autre part, qu'une définition ne doit pas être critiquée isolément, mais en rapport avec l'ensemble des normes, une modification heureuse sur un point pouvant souvent entraîner des changements regrettables sur d'autres points.

Je vais maintenant passer à l'examen détaillé des normes.

Après avoir défini les différentes moyennes : moyennes arithmétique, géométrique, harmonique, moyennes pondérées, la médiane et la médiane arithmétique, les quartiles, déciles, centiles, les médianes, quartiles et déciles pondérés et avoir normalisé le surlignage pour désigner la moyenne arithmétique, la norme X 05-001 — Terminologie de la statistique — fait une innovation en donnant de la « différence » et de l'« écart » deux définitions dont la nécessité s'imposait.

La « différence » de deux nombres a et b est définie comme étant la quantité algébrique $a - b$ et l'« écart » de deux nombres est la valeur absolue de leur différence.

La norme aborde ensuite l'étude des caractères, puis des séries statistiques à un caractère.

Il ne faut pas oublier que ces normes sont le résultat de discussions prolongées. En bien des cas, l'accord n'a pas pu être réalisé d'une manière unanime et le texte est souvent le résultat d'un compromis. Toutefois, la Commission s'est efforcée de ne pas prononcer d'exclusive contre les termes en usage, même quand elle a jugé qu'il y avait lieu d'en proposer de nouveaux. Par exemple, la norme X 05-001 a limité l'emploi du terme « fréquence » à ce qu'on appelle « fréquence relative » et elle a désigné par « effectif » ou « nombre de répétitions » ce qu'on nomme « fréquence absolue », mais elle a laissé subsister les anciennes expressions de la manière suivante : On dit aussi fréquence absolue, on dit aussi fréquence relative. Il faut souhaiter que l'emploi du mot « effectif » se généralise. Il est bref et euphonique et ne prête à aucune ambiguïté. Par exemple l'effectif d'une classe a exactement le même sens en statistique que dans le langage des militaires.

Le nom d'« étendue » a été donné à l'écart entre la plus grande et la plus petite des valeurs d'une série statistique, généralisant ainsi à la statistique la signification que l'on donne à ce terme en musique.

La question des « limites de classes » s'est montrée particulièrement délicate.

En effet l'aspect scientifique de la question se double d'un aspect juridique, celui des jours francs et des délais. La question des limites est familière aux mathématiciens. Pour le géomètre, c'est un point, pour l'algébriste, c'est un nombre rationnel ou irrationnel que l'on définit dans ce dernier cas par une coupure. Pour le législateur la limite d'un délai n'est pas toujours ponctuelle, c'est quelquefois une durée déterminée, comme un jour. Par exemple tel règlement accorde trois jours francs à partir d'un certain jour, tel contrat spécifie dix jours à partir du 4 janvier. Le 14 janvier est-il ou non compris dans le délai?

Lorsque la limite est ponctuelle, comme par exemple lorsqu'il s'agit de faire repeindre les bornes du kilomètre 30 au kilomètre 40, la question n'est pas plus simple. Dans l'exemple cité, l'entrepreneur a-t-il à repeindre la borne 30 et la borne 40 ou l'une des deux seulement et laquelle?

Il n'a pas été possible de prendre une position définitive; la norme s'est contentée de spécifier qu'il y avait toujours lieu de préciser si les limites font partie de la classe et, dans le cas où une seule des limites en fait partie, de recommander d'inclure dans la classe sa limite inférieure.

Une autre question difficile était la définition des « nombres indécis ». Les définitions qui ont été proposées ont été longuement examinées. La Commission a préparé plusieurs rédactions, mais aucune d'elles ne lui a paru suffisamment satisfaisante pour pouvoir être introduite dans la norme.

Une certaine confusion régnait dans l'appellation des « diagrammes de fréquence ». Les seules expressions normalisées sont :

- diagramme en bâtons des fréquences et des fréquences cumulées;
- histogramme;
- polygone des fréquences cumulées.

La norme définit ensuite les valeurs typiques et les moments. En statistique, ces expressions sont qualifiées d'« empiriques ». Lorsque les moments sont rapportés à la moyenne arithmétique, l'adjectif « apparent » est employé.

L'expression d'« écart-type » a été normalisée pour désigner l'écart moyen apparent quadratique, c'est-à-dire l'écart moyen quadratique par rapport à la moyenne arithmétique. La norme définit ensuite « l'écart médian » l'« interquartile », le « semi-interquartile » et le « coefficient de variation ». L'expression d'« écart probable », dans le sens d'écart médian par rapport à la moyenne arithmétique ou semi-interquartile, a été prohibé en statistique.

Dans la norme X 05-002 sur la terminologie du calcul des probabilités, la définition si controversée de la probabilité n'a pas été donnée. On s'est contenté d'exemples : exemple d'un événement aléatoire, exemple de deux événements aléatoires également probables, exemple de probabilité. Il est à remarquer également que le mot « hasard » ne figure pas dans les normes.

La norme donne des exemples de variables aléatoires et d'éléments aléatoires, qui montrent bien le caractère de généralité que les rédacteurs de la norme ont voulu lui donner. En effet, parmi les exemples figurent la forme d'un fil jeté d'une manière quelconque sur une table et la couleur des yeux d'un individu appartenant à une certaine population.

L'expression : « loi de probabilité », a été normalisée pour désigner ce qui est souvent appelé loi de répartition ou loi de distribution.

La norme définit ensuite la « probabilité intégrale », la « fonction de répartition », la « probabilité élémentaire ».

L'appellation et le symbole à adopter pour l'espérance mathématique ont fait l'objet de nombreuses et vives discussions. La Commission a dû procéder à un referendum dans lequel on rappelait les expressions déjà employées :

- espérance mathématique;
- valeur probable;
- valeur moyenne;
- valeur centrale.

Si l'espérance mathématique était adoptée, on proposait deux formes pour le symbole.

Le referendum a donné une grosse majorité en faveur de l'expression : « espérance mathématique », symbole E.

La question des crochets après le symbole, n'ayant pas été soulevée dans le referendum, fit l'objet d'une seconde enquête. Finalement la Commission adopta le symbole E [X] qui permet aux actuaires de distinguer l'espérance mathématique de la représentation d'un capital différé. La norme précise que les crochets peuvent être omis; elle signale également que l'on emploie quelquefois le symbole $\mathcal{M} X$. Quant au symbole \bar{X} , il est réservé à la statistique.

Quoique la norme ait été rédigée avec le concours d'éminents mathématiciens, l'appareil mathématique auquel elle fait appel est très réduit. En particulier l'intégrale, dite de Stieltjes, n'est mentionnée qu'en note à propos de l'espérance mathématique.

La norme définit ensuite la « valeur la plus probable » dans le cas d'une variable aléatoire discrète et la valeur « dominante » dans le cas d'une variable aléatoire continue. Le mot « mode », qui est pourtant un mot français adopté par les Anglais, n'a pas été retenu, mais il a toutefois été mentionné. Ces deux notions sont symbolisées par \hat{X} , qui s'énonce grand X circonflexe.

Le symbole normalisé pour la valeur équiprobable est $\bar{\bar{X}}$ qui s'énonce grand X égalité. Ces symboles nouveaux sont clairs et expressifs. Il est à souhaiter que leur emploi se généralise.

Viennent ensuite les moments. Les moments rapportés à l'espérance mathématique sont qualifiés de centrés. L'ancienne expression d'« écart probable » a été rejetée et remplacée par une expression nouvelle, plus imagée : « écart équiprobable ».

En ce qui concerne la « loi de Laplace », la Commission, considérant que Laplace a découvert la loi qui devrait porter son nom et qui porte celui de Gauss, à une époque où Gauss était encore un jeune enfant, a rétabli la vérité en rendant à Laplace l'hommage qui lui était dû.

La norme se termine par un chapitre intitulé : Erreurs envisagées du point de vue du calcul des probabilités, qui donne la correspondance entre la terminologie du calcul des probabilités et celle de la théorie des erreurs.

La même largeur de vues, qui avait présidé à la rédaction de la norme relative à la statistique à un caractère se retrouve dans la norme X 05-003 concernant la statistique à plusieurs caractères.

Le terme « mesure » a été évité parce qu'il n'a pas été jugé suffisamment précis. Il a en effet de nombreux sens (Littré en mentionne vingt et un) : en particulier l'instrument de mesure, l'action de mesurer et le résultat de cette action. Dans un but de précision, le mot mesure dans le sens d'action de mesurer a été remplacé par « mesurage », qui est évidemment un peu lourd, mais qui, lui, n'a qu'un seul sens.

Pour éviter qu'une confusion se produise entre les expressions « erreur absolue » et « valeur absolue de l'erreur », puisque dans la norme l'erreur est algébrique, l'expression, qui peut paraître un peu surprenante, d'« erreur vraie » a été normalisée.

La norme définit également les autres erreurs : relative, systématique, parasite et accidentelle.

La définition de l'erreur systématique a donné lieu à une difficulté. Certains membres de la Commission considéraient que toute erreur a une cause, connue ou inconnue, et qu'en conséquence toute erreur est systématique. D'autres membres au contraire estimaient que les erreurs systématiques sont les erreurs dont la cause est connue et dont par conséquent on se débarrasse avant toute interprétation du résultat. Un accord a heureusement pu être réalisé et le texte de la norme a reçu l'approbation générale.

L'« erreur quadratique moyenne » et l'« erreur probable » n'ont pas été définies, puisque la correspondance entre ces notions et les notions correspondantes du calcul des probabilités fait l'objet du dernier chapitre de la norme X 05-002.

La présentation des résultats approchés a fait l'objet de longues discussions. Pour le physicien, le résultat d'un mesurage, donné sous la forme $A \pm \epsilon$, signifie qu'en recommençant la même mesure un certain nombre de fois, les écarts avec la mesure A ne dépasseront que rarement ϵ , mais en général, il ne précise pas si les écarts se rapportent aux valeurs individuelles ou aux moyennes calculées sur des groupements de valeurs individuelles; il n'indique pas non plus quelle est la probabilité pour qu'une valeur individuelle dépasse exceptionnellement ϵ .

Pour le rédacteur d'un cahier des charges ou pour un contrôleur, $A \pm \epsilon$ signifie que les pièces qui s'écartent de plus de ϵ de la valeur A doivent toutes être rebutées, de manière qu'en définitive toutes les pièces reçues soient, toutes sans exception, comprises entre $A - \epsilon$ et $A + \epsilon$.

La Commission n'a pas voulu imposer de règle stricte. Elle s'est contentée, d'une part de définir l'erreur à craindre correspondant à une probabilité donnée et d'autre part d'exiger que, lorsqu'un résultat est donné sous la forme $A \pm \epsilon$, la signification de la quantité ϵ soit toujours précisée.

Les normes sont complétées par un fascicule de documentation n'ayant pas le caractère d'une norme et où sont rassemblés les différents symboles figurant dans les normes.

Ces symboles sont disposés sous forme d'un tableau qui peut servir également de table de matières et d'aide-mémoire.

Ce fascicule fait ressortir en particulier l'emploi qui a été fait dans les normes des lettres capitales, des lettres minuscules accentuées ou non et des lettres grecques.

Les caractères et les variables aléatoires sont représentés par des capitales. Par exemple X, Y.

Les valeurs distinctes, pouvant être répétées n_1, n_2, \dots fois, d'un caractère ou les valeurs distinctes d'une variable aléatoire sont représentées par des minuscules non accentuées. Par exemple $x_1, x_2, \dots, y_1, y_2, \dots$

Les valeurs empiriques distinctes ou non, d'un caractère sont représentées par des minuscules accentuées $x'_1, x'_2, \dots, y'_1, y'_2, \dots$. Il en est de même des moments empiriques.

Par contre, les moments apparents, c'est-à-dire les moments par rapport à la moyenne arithmétique, sont désignés par des lettres grecques, qui sont accentuées quand il s'agit de moments empiriques.

Après avoir été imprimées, les normes ont fait l'objet de deux enquêtes publiques. De nombreuses réponses ont été reçues, soit d'utilisateurs, soit de statisticiens n'ayant pas pu prendre part aux travaux de la Commission, soit d'organismes étrangers. Elles ont permis la mise au point des normes.

Avec ces normes, l'A. F. N. O. R. abordait un terrain particulièrement redoutable, la normalisation des termes scientifiques. L'enquête publique a montré que cette façon de concevoir la normalisation avait de nombreux détracteurs.

On nous a en effet écrit :

D'une part :

« Je crois que la matière visée ne relève pas de la normalisation. Il s'agit en effet d'un travail scientifique pour lequel une grande latitude doit être encore laissée aux intéressés ».

D'autre part :

« Les projets sont alourdis de nombreuses définitions que je juge inutiles parce que je n'imagine pas que ces expressions puissent être employées dans des sens différents de ceux indiqués. Je reconnais cependant que le fait de citer ces expressions doit avoir pour résultat d'éviter que d'autres expressions plus ou moins heureuses soient employées dans le même sens ».

Enfin :

« Par leur caractère abstrait, les matières dont il s'agit (statistique, calcul des probabilités, erreurs de mesure) me paraissent difficilement se concilier avec l'objet de la normalisation. Pour les matières en question, le but qu'il convient d'atteindre me semble être en effet de réaliser l'accord sur une terminologie qui prête encore aujourd'hui à interprétations souvent divergentes. Or, pour la mise au point d'une terminologie appropriée, la procédure de normalisation n'est pas, à mon avis, suffisamment souple, ce qui s'explique parce qu'elle n'a certainement pas été envisagée pour des objets de caractère aussi abstrait. Je ne pense pas en conclusion que la normalisation projetée soit véritablement génératrice de progrès; je crains plutôt qu'elle risque de cristalliser la terminologie, peut être même les concepts, dans une forme définitive et, par conséquent mal adaptée à l'évolution qu'il y a lieu de prévoir pour ce genre d'étude ».

J'arrête ici ces citations. En bref, nos correspondants nous reprochaient :

- d'avoir fait des normes;
- de définir des expressions que tout le monde est censé connaître;
- de cristalliser la terminologie, peut-être même les concepts.

¶ Nous avons fait des normes parce qu'un grand nombre de personnalités éminentes nous ont demandé de le faire. Les difficultés rencontrées au cours de leur élaboration ont montré que certaines notions, d'emploi quotidien, donnaient lieu à des interprétations divergentes et quelquefois opposées. La préparation des normes a été l'occasion pour les rédacteurs d'éprouver leur terminologie.

Dire que les normes définissent des expressions que tout le monde est censé connaître n'est pas tout à fait exact. Les statisticiens les plus éminents ont commencé par ignorer ce qu'était une moyenne arithmétique ou un écart-type. Il est bon que l'étudiant ait sous la main une norme qui lui indique le sens précis qu'un aéropage de spécialistes est convenu de donner aux termes qu'il est amené à employer.

D'autre part, si l'on veut traduire avec exactitude un ouvrage étranger, il n'y a pas de plus sûr moyen, pour rendre fidèlement la pensée de l'auteur, que de disposer, dans les deux langues, de définitions claires permettant de s'assurer que le mot français et le mot étranger, qu'on a l'intention de lui faire correspondre, ont exactement le même sens.

Les normes sont-elles susceptibles de cristalliser la terminologie? J'ai montré tout au long de cet exposé l'esprit dans lequel les normes ont été rédigées. Chaque fois qu'une expression nouvelle a été normalisée, l'expression traditionnelle a été conservée en deuxième position. Les normes ne cristallisent pas plus la terminologie que ne le fait un dictionnaire et l'on sait que les dictionnaires, même celui de l'Académie, n'ont jamais empêché les langues d'évoluer. Leurs éditions successives en sont la preuve.

Reste la dernière objection. Les normes sont-elles susceptibles de cristalliser les concepts?

Définir c'est toujours en un certain sens cristalliser. Cependant les mots doivent être définis, puisque seules les onomatopées ont un sens par soi-même. Mais définir n'a jamais empêché d'évoluer. Le concept des parallèles a survécu à travers les âges dans le sens que lui avait donné initialement Euclide. Cette cristallisation n'a pas empêché Lobatchevsky et Riemann d'inventer des concepts différents.

Ce serait un grand malheur que la moyenne arithmétique ne soit pas définie et que n'importe quel auteur puisse définir sa propre moyenne arithmétique. Cette cristallisation n'a cependant pas empêché l'école italienne d'inventer toutes sortes de moyennes. Mais ce ne sont pas des moyennes arithmétiques.

Je terminerai en remerciant les membres de la Société de Statistique de Paris, et en particulier MM. Darmois et Fréchet qui ont bien voulu attirer l'attention des statisticiens étrangers sur ces normes. Cette propagande a valu à l'A. F. N. O. R. de nombreuses commandes provenant des États-Unis, du Royaume-Uni et même d'Australie.

La diffusion à l'étranger des normes françaises de statistique a contribué au maintien du prestige de la science française. Je n'en voudrais pour preuve que la phrase suivante, intercalée dans un bon de commande provenant de Londres :

« Il est fort à propos que le pays de Laplace, de Demoivre et de Poisson continue d'avoir la préséance dans ce domaine ».

André PALLEZ.

DISCUSSION

M. BELGODÈRE demande s'il a été établi et publié, soit des textes en langue anglaise analogues aux normes de terminologie de l'AFNOR, soit un glossaire anglais-français et français-anglais des termes de la statistique et du calcul des probabilités.

MM. DARMOIS et PALLEZ répondent en citant des exemples de difficultés de traduction, M. Darmois annonce la parution prochaine d'une traduction française de l'ouvrage de F. YATES, *Manuel de Sondages*, préparé à la demande de la Sous-Commission des Sondages statistiques des Nations Unies. La traduction sera, autant que possible, conforme à la terminologie codifiée par l'AFNOR, et, de fait, rendra dans une certaine mesure, et pour les questions de Sondages, les mêmes services qu'un tel dictionnaire. La question générale de la terminologie internationale du Calcul des Probabilités et de la Statistique mathématique sera très probablement étudiée par l'Institut International de Statistique, à sa prochaine session, en septembre 1949.
