JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

DANIEL DUGUÉ

Valeurs extrêmes en statistique

Journal de la société statistique de Paris, tome 101 (1960), p. 194-203
<http://www.numdam.org/item?id=JSFS_1960__101__194_0>

IV

VALEURS EXTRÊMES EN STATISTIQUE

En commençant cet exposé je voudrais vous rappeler un entretien que j'ai eu avec mon maître G. Darmois qui fut président de cette société en 1938 et cela me permettra d'évoquer sa chère mémoire.

Désireux de m'orienter en 1934 vers une discipline qui m'était rendue sympathique par l'attrait de la personnalité que vous connaissez tous, j'étais allé le trouver dans son bureau de l'Institut Henri Poincaré et après m'avoir parlé d'un certain nombre de problèmes encore non défrichés il me dit :

« J'attire votre attention sur l'importance qu'il y a en statistique à bien identifier une variable et en particulier à ne pas la confondre avec la plus grande valeur d'une série » et il me raconta l'anecdote scientifique suivante :

On sait qu'un nombre normal dans le système décimal est un nombre pour lequel la fréquence de chacune des décimales tend vers 1/10 quand leur nombre augmente indéfiniment. D'après les travaux mathématiques de la théorie de la mesure, on sait que la mesure de l'ensemble des points dont les abscisses ne sont pas normales dans le système décimal est nulle, ce qui entraîne que les nombres les plus répandus sont les nombres normaux. C'est une hypothèse non démontrée que le nombre π appartient à l'ensemble des nombres normaux. Si l'on choisit un nombre quelconque de cet ensemble il s'ensuit que la fréquence des décimales ne doit pas s'écarter de 1/10 plus que ne le permet le calcul des probabilités avec un seuil de signification déterminée. Les 1 000 premières décimales de π sont ainsi réparties :

0 —	93	5 —	97
1 —	**116**	6 —	94
2 —	103	7 —	95
3 —	102	8 —	101
4 —	93	9 —	106
			1 000

La fréquence de 1 est donc de 0,116 supérieure de 0,016 à 0,10. L'écart type de cette fréquence sur 1 000 résultats est :

$$\sqrt{\frac{p\,q}{1\,000}} = \sqrt{0,00009} = 0,00948$$

L'écart normé de la fréquence de 1 sera donc :

$$\frac{0,016}{0,00948} = 1,68\ \ldots\ldots$$

La probabilité d'un écart normé supérieur à cette valeur est d'après les tables de la loi de Laplace-Gauss de 0,0465 inférieure au seuil de signification de 1/20. Donc en adoptant ce seuil nous devons rejeter l'hypothèse que π est normal. C'est du moins ainsi que l'on peut être tenté de raisonner et la conclusion serait entachée d'une erreur. 1 ici n'est pas une décimale quelconque, c'est la décimale *la plus fréquente* dans les 1 000 premières décimales de π. Ce que nous observons ici n'est pas la fréquence d'une décimale prise au hasard. Mais bien le fait que toutes les décimales ont une fréquence dont l'écart normé à 1/10 est inférieur à 1,68... La probabilité d'un écart normé inférieur à 1,68 étant 0,9535 il en résulte que la probabilité de dix écarts normés indépendants inférieurs à 1,68... est $[0,9535]^{10}$ (en fait ici les dix écarts ne sont pas indépendants car la somme des dix fréquences est égale à l'unité mais l'erreur entraînée est minime).

$$[0,9535]^{10} = 0,62\ldots$$

La probabilité pour que le plus grand des dix écarts normés soit (sous la réserve faite entre parenthèses) supérieur à 1,68... est donc 0,48 — et nous retrouvons une probabilité tout à fait normale par rapport au seuil de signification de 1/20.

Voilà l'histoire que m'avait racontée M. Darmois et depuis j'ai toujours été intéressé par le comportement des valeurs extrêmes. Je vais vous en dire quelques mots aujourd'hui. Il s'agit là d'un ensemble de résultats très important au point de vue critique de l'expérience.

On peut dire qu'une partie de l'œuvre de M. Borel tourne autour de cette théorie : la convergence presque certaine en particulier considérée comme différente de la convergence en probabilité, c'est la constatation du fait que la plus grande d'une suite de variables aléatoires n'est pas une de ces variables mais bien un être différent construit à partir de ces variables. Un grand nombre de statisticiens de premier plan se sont intéressés à la question : Sir Ronald Fisher, MM. Fréchet, Gnedenko, Gumbel et tout récemment un jeune et brillant débutant M. Geffroy. Je vais rapidement vous indiquer quelques résultats mathématiques et vous parler ensuite de quelques problèmes pratiques et de quelques remarques soulevées par ces questions.

Plus grande valeur comme élément typique au sens de M. Fréchet.

Une valeur typique est une quantité certaine attachée à une quantité aléatoire et qui donne une information sur le comportement de cette aléatoire, qui permet éventuellement de prendre certaines décisions. La plus connue est la valeur moyenne. Si $F(x)$ est la loi de probabilité des variables indépendantes entre elles dont on s'occupe, cette valeur moyenne est l'intégrale de Stieltjes $\int_{-\infty}^{+\infty} x\, d\, F(x) = a$. Et le phénomène connu sous le nom de loi des grands nombres peut se formuler dans l'énoncé vague suivant :

Sous certaines conditions la moyenne $\frac{x_1 + x_2 + \ldots + x_n}{n} = \overline{x_n}$ *de n résultats aléatoires indépendants ayant* F (x) *pour loi de probabilité converge au sens du calcul des probabilités* (et il y a plusieurs sens différents de cette notion de limite) *vers a., quand n augmente indéfiniment.*

Donc si l'on a n résultats, la moyenne de ces n résultats donne sous certaines conditions une information sur $\int_{-\infty}^{+\infty} x\, d\, F(x)$ qui, à son tour, donne une certaine information sur F (x). Il en est de même pour la plus grande valeur de n résultats $x_1, x_2, \ldots, x_n$ que je noterai x_n^*, mais ici les choses vont être un peu plus compliquées.

F (x) est toujours la loi de probabilité de la variable x_i (i quelconque). La loi de

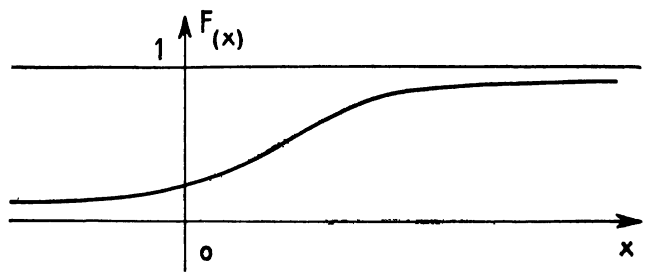

Fig. 1.

probabilité de x_n^* (c'est-à-dire la probabilité pour que *toutes* les variables x_i (i variant de 1 à n) soient inférieures à x sera $F^n(x)$.

En fig. 1 est représentée la loi de probabilité F (x) (fonction non décroissante variant de 0 à 1 quand x varie de 0 à $+\infty$).

En fig. 2 est représenté $F^n(x)$. Quand n augmente indéfiniment, si F (x) satisfait à certaines conditions sur lesquelles je n'insiste pas et pour lesquelles je renverrai à la thèse de M. Geffroy F (x) va se cambrer de plus en plus de manière à ressembler de plus en plus à la forme de la figure 3; l'abscisse *b dépendant de n,* soit b_n.

C'est le phénomène de la stabilité en probabilité dégagé par Kolmogoroff. Et voici l'énoncé dû à Gnedenko :

Sous certaines conditions, la plus grande valeur x_n^* *de n résultats indépendants* x_1, x_2,

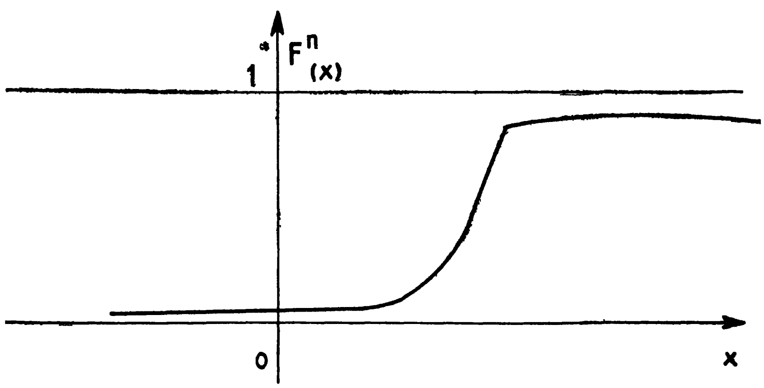

Fig. 2.

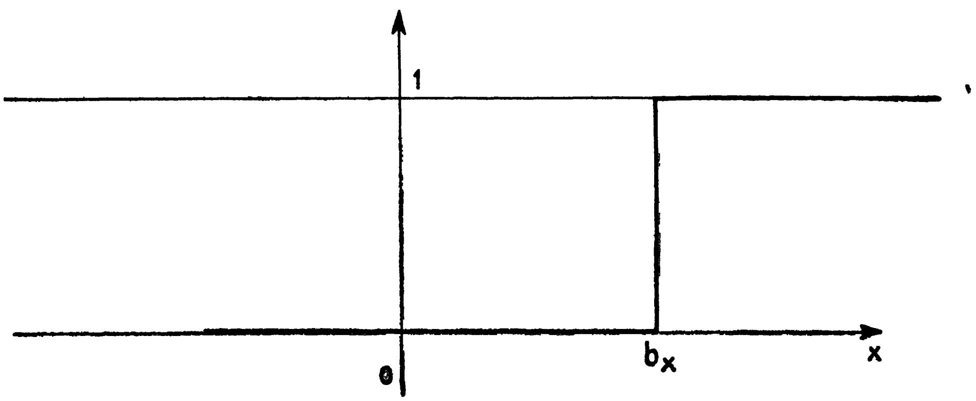

Fig. 3.

... x_n *ayant même loi de probabilité* F (x) *est stable au sens du calcul des probabilités, c'est-à-dire qu'il existe une suite de nombres certains* a_1, a_2, ..., a_n, — *tels que* $x_n^* - a_n$ *tend vers* O *au sens du calcul des Probabilités* (comme je l'ai déjà dit, il y a plusieurs sens pour cette convergence).

Donc comme la moyenne, la plus grande valeur dans ces cas très généraux donne des informations sur une variable aléatoire mais les choses sont moins simples.

« En gros » la moyenne c'est une quantité certaine et fixe (indépendante du nombre des résultats), la plus grande valeur c'est aussi une quantité certaine *mais variable avec n.* C'est une fonction de n $\Phi(n)$. $\Phi(n)$ étant caractérisée par la forme de F (n). $\Phi(n)$ peut être déterminée dès que l'on a F (x). Un exemple va éclairer davantage ces affirmations. Considérons une population de N êtres humains nés au même instant et pour lesquels la loi de probabilité de vie est la même. Quand N augmente, la moyenne des vies de chacun de ces êtres humains va tendre vers un nombre fixe indépendant de N. Le « record » de longévité de cet ensemble va pouvoir être *à l'avance* déterminé d'une manière de plus en plus précise si N augmente de plus en plus (et si l'on connaît la loi de probabilité de vie des êtres humains) mais ce record variera avec N, ce sera une fonction évidemment croissante de N.

Autant la convergence de la moyenne vers un nombre certain est un fait que l'on réalise facilement, autant le résultat sur la plus grande valeur surprend à première vue.

* * *

La plus grande valeur dans certains cas d'ailleurs très généraux donne donc une information sur la variable que l'on étudie, encore faut-il qu'elle soit correctement maniée.

Une enquête américaine voulait établir que l'aptitude sportive décroît avec l'âge. Cette enquête est résumée par l'information suivante :

Pour 3 012 sujets de moins de 15 ans, la cote de la meilleure épreuve était 29.
Pour 58 sujets de 60 à 65 ans, la cote de la meilleure épreuve était 4.

De la comparaison des deux chiffres 29 et 4, l'enquêteur concluait à une réduction de l'aptitude sportive, alors que les deux chiffres (l'un « record » sur 3 012, l'autre record sur 58) ne sont absolument pas comparables.

L'aptitude sportive pour les sujets de moins de 15 ans est une variable aléatoire comme pour les sujets de 60 à 65 ans. Soit $F_1(x)$ la loi de probabilité de la première et $F_2(x)$ la loi de probabilité de la seconde (x étant la cote).

Dire que l'aptitude sportive décroît avec l'âge pourrait se manifester par la constatation que pour F_2 les valeurs faibles de x sont plus probables que pour F_1. Cela entraînera que $\Phi_1(n)$ au sens où nous l'avons depuis tout à l'heure va se trouver au-dessus de $\Phi_2(n)$ pour toutes valeurs de n. Mais le fait que

$$\Phi_1(3.012) > \Phi_2(58)\ ,$$

n'entraîne nullement que $\Phi_1(n) > \Phi_2(n)$ pour tout n. La figure 4 montre un exemple du contraire. L'argument fourni par l'enquête est donc sans valeur.

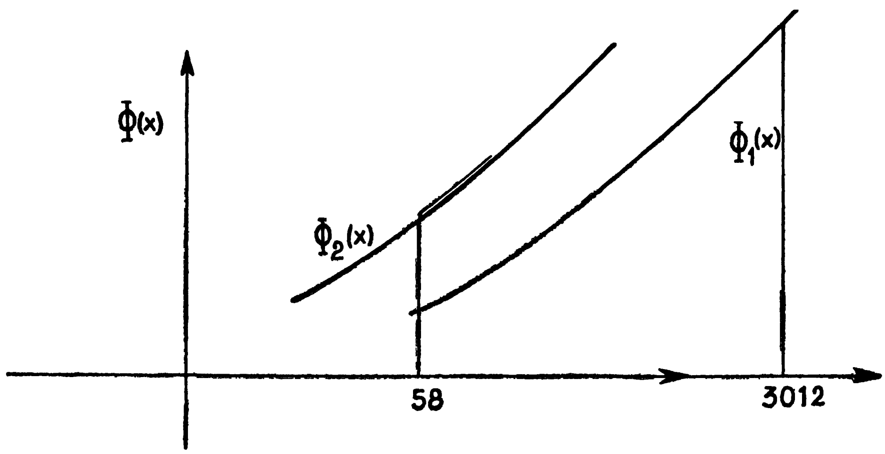

Fig. 4.

Ajoutons également une source de difficultés : c'est que la plus grande valeur peut être entachée dans le cas d'une enquête d'une erreur beaucoup plus gênante que la moyenne.

Imaginons que la cote donnée à chaque sujet soit non pas une note sous contrôle mais le résultat d'une enquête (on demande à chaque sujet par exemple la hauteur de saut dont il est capable). Il suffit d'un *seul* sujet ayant le désir de fausser sa note dans des proportions considérables (par exemple de la doubler) pour que toute l'enquête soit grossièrement erronée. Il n'en est pas de même pour la moyenne.

*
* *

La plus grande valeur me paraît jouer un rôle important dans toutes les théories de sécurité en comprenant ce terme au sens le plus large de manière à y faire entrer une théorie de la mortalité.

L'étude de l'usure des matériaux par fatigue a récemment fait l'objet de la thèse de M. Bastenaire qu'il a communiquée au Séminaire de Statistiques en Mars dernier. Dans cette étude la loi de Weibull qui est une loi de plus grande valeur joue un rôle énorme. C'est la simple constatation du fait que la résistance d'une chaîne est celle de son maillon le plus faible et que pour qu'une chaîne survive à un effort il est nécessaire que tous ses maillons survivent. Nous retrouvons le phénomène de plus grande (ou plus petite) valeur. De même la résistance d'un organisme est celle de son organe essentiel le plus fragile. J'ai toujours été très intéressé et je souhaite que cette étude se tente, par le fait que la durée de vie d'organismes animaux est étonnamment variable d'une espèce à l'autre (chien 15 ans, chat 20 ans etc.) Je sais que l'on peut m'objecter que cette étude a en partie été faite et que les courbes de mortalité sont les intégrlaes d'équations différentielles que l'on peut justifier. J'avoue ne pas être entièrement satisfait car il me semble que le problème n'est que déplacé et assez faiblement déplacé. Pour arriver à une théorie qui me paraisse valable, il faudrait que devant un organisme déterminé (celui du cheval ou de l'éléphant par exemple), on puisse calculer les constantes qui entrent dans ces équations différentielles et on en est encore très loin.

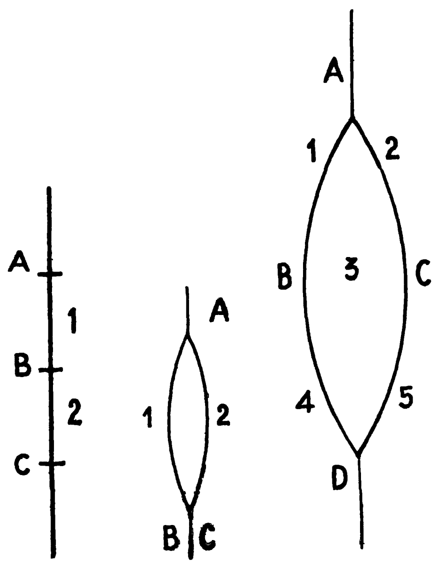

Fig. 5. Fig. 5 *bis*. Fig. 5 *ter*

Je voudrais ici dire un mot d'une application possible de la théorie des graphes à ces études de durée de vie. Supposons tout d'abord une chaîne composée des 2 tronçons 1 et 2 (fig. 5). (1 de A à B, 2 de B à C.)

Si p_1 est la probabilité de résistance du tronçon 1, p_2 celle du tronçon 2, la probabilité de résistance de la chaîne est $p_1\, p_2$.

Si l'on avait la disposition en parallèle (fig. 5 *bis*) au lieu de la disposition en série de la figure 5, la résistance de la chaîne serait avec la même définition de p_1 et p_2, $p_1 + p_2 - p_1\, p_2$, comme on le voit aisément.

Imaginons que l'on ait maintenant une chaîne constituée par 5 tronçons reliés entre eux selon le « graphe » de la figure 5 *ter*. p_1, p_2, p_3, p_4, p_5 étant les probabilités de résistance de chacun des tronçons, quelle sera la probabilité de résistance de la chaîne?

La formule de Poincaré permet de donner son expression.

$$p_1\, p_4 + p_2\, p_5 + p_1\, p_3\, p_5 + p_2\, p_3\, p_4 - p_1\, p_2\, p_4\, p_5 - p_1\, p_3\, p_4\, p_5 - p_2\, p_3\, p_4\, p_5$$
$$- p_1\, p_2\, p_3\, p_4 - p_1\, p_2\, p_3\, p_5 + 2\, p_1\, p_2\, p_3\, p_4\, p_5$$

On voit donc que la structure des liaisons à l'intérieur des matériaux et à l'intérieur de l'organisme (ici cela pose l'existence de circuits de remplacement) modifie la probabilité de survie. Ce serait je crois un point fort intéressant à élucider que celui de la représentation par graphe de la structure de résistance d'un organisme.

*
* *

Enfin le dernier point auquel on peut me semble-t-il appliquer la théorie des plus grandes valeurs est celle de l'opinion (avec toutes les incidences possibles sur la théorie des sondages). Jusqu'à présent à ma connaissance les spécialistes de l'opinion publique se sont surtout attachés à la définir à un instant donné. Ils n'ont pas cherché à analyser scientifiquement ses causes ni à préciser ses lois d'évolution. Il serait peut-être intéressant de définir une physique de l'opinion. Ce pourrait d'ailleurs être aussi terriblement dangereux. A mon sens cette théorie se rattache à la théorie des valeurs extrêmes de la façon suivante : l'opinion d'un groupe est non pas comme on le croit souvent l'opinion moyenne de toutes les opinions, mais l'opinion d'un seul membre des groupes, du leader. Je pense ici à une enquête récente effectuée dans une grande école sur le succès d'un certain mode d'enseignement. La direction avait posé une question aux élèves en précisant qu'ils aient à en discuter en salle (une promotion de 300 élèves est répartie en 30 salles de 10 chacune ayant à sa tête un élève classé dans les 30 premiers). L'enquête avait donné un pourcentage d'environ 70 % de réponses affirmatives à la question : « Tirez-vous profit de ce mode d'enseignement? »

Une contre-enquête que j'ai effectuée par sondage non chiffré m'a convaincu que la presque totalité des élèves ne tirait aucun profit de ce mode d'enseignement. Comment expliquer ces deux résultats contradictoires. A mon avis l'explication est très simple. Dans chacune des salles où avait lieu l'enquête le chef de salle qui est un très bon élève impose son point de vue à tous ses camarades de salle. L'enquête révèle donc selon moi simplement que 70 % des *meilleurs élèves* tirent profit du mode d'enseignement au sujet duquel l'enquête avait lieu.

Je ne sais quel est le principe qui a amené à faire décider par un jury de la culpabilité d'un homme accusé d'un crime. J'imagine que l'on peut y voir une trace de l'antique adage « vox populi, vox dei ». Encore faudrait-il que cette enquête par sondage ne soit pas faussée par la délibération.

La délibération d'un jury a pour résultat de cristalliser l'opinion d'un groupe autour de l'opinion du leader.

L'article 312 de l'ancien code d'instruction criminelle (je pense qu'il est toujours en vigueur) prescrivait aux jurés de prêter le serment de ne communiquer avec personne au sujet de l'affaire en jugement jusqu'après la délibération.

Si ce que l'on souhaite est la manifestation par sondage de la « vox populi » je crois que dans la circonstance il vaudrait mieux laisser les jurés parler dans leurs groupes sociaux de l'affaire qu'ils ont à juger et les empêcher ensuite de délibérer, que de leur permettre de délibérer en s'abstenant de communiquer avec l'extérieur.

Enfin le phénomène de popularité et l'instabilité qui y est attachée s'explique aussi très bien par la théorie de la plus grande valeur. L'opinion d'un groupe sur une personne est encore l'opinion d'un leader. Il suffit que ce leader disparaisse pour que l'opinion du groupe soit totalement modifiée. C'est ce que l'on constate au sujet de la popularité d'un professeur dans une promotion d'élèves. Cette popularité ou cette impopularité dépend de l'opinion sur le professeur du leader de la promotion.

J'espère vous avoir convaincus de l'utilité fondamentale de ces études sur les plus grandes valeurs et souhaite que les problèmes qu'elles soulèvent fassent l'objet de recherches qui me paraissent très importantes.

Daniel Dugué

DISCUSSION

M. Aubenque. — Puisque l'on a évoqué des situations biologiques, on pourrait rappeler qu'en biométrie notamment les valeurs extrêmes des distributions expérimentales peuvent être aberrantes. Il est classique de citer, en l'occurrence, l'exemple des géants et des nains qui, rencontrés si l'on considère de très larges échantillons de population, donnent une allure discontinue aux extrémités de la distribution des tailles. Mais interviennent alors des facteurs nettement morbides, de sorte que l'on a affaire à des « populations » de tailles pathologiques distinctes de la population de tailles supposées physiologiques. On peut dire que l'on ne mesure plus des choses ayant exactement la même nature (bien que le passage du physiologique au nettement pathologique soit bien incertain). Il ne saurait donc être question d'invoquer ici une exception à la foi mathématique des valeurs extrêmes. On peut penser, au contraire, qu'une application judicieuse de la théorie pourrait avoir en l'occurrence une précieuse utilité (discrimination). Il s'agissait simplement d'attirer l'attention sur l'éventualité d'anomalies au moins apparentes de valeurs extrêmes expérimentales, anomalies dont la critique et l'analyse statistique doivent être toujours faites.

M. Dugué. — Je suis tout à fait d'accord avec M. Aubenque. Il est bien certain que devant un cas extrême on peut toujours se demander s'il ne s'agit pas d'un individu en dehors de l'échantillon (c'est-à-dire selon la terminologie mathématique n'obéissant pas à la loi de probabilité du lot). Cela pose la question de la discrimination à laquelle se sont attachés beaucoup de bons auteurs (en particulier M. Ivanovîc).

M. Metz. — Je m'excuse de prendre la parole : je suis ici au nom de mon ami François Divisia, avec qui je travaille depuis plusieurs années sur des problèmes de statistiques portant sur des produits industriels fabriqués en grande série.

Il n'a pu venir et m'a demandé de le remplacer ici.

Nos études nous ont amenés à des conclusions analogues à celles de M. Aubenque qui a pris la parole tout à l'heure à propos des statistiques portant sur la taille des hommes : il a fait remarquer que les « valeurs extrêmes » dans cet ordre d'idées correspondent à des « nains », et que le nanisme est un phénomène tout différent de ceux qui provoquent des variations de taille chez les individus normaux ou voisins de la normale.

De la même façon, dans les fabrications industrielles, on constate que le plus grand nombre des objets fabriqués s'écartent peu de la moyenne, et obéissent à des lois voisines de celle de Laplace-Gauss. Mais il y a — et bien souvent en nombre beaucoup plus grand que ne l'indique le calcul d'après cette loi — des « valeurs extrêmes », qui correspondent à des objets *à rebuter*. Ces accidents ne résultent pas de la superposition aléatoire de nombreux petits écarts (ils seraient alors beaucoup plus rares) mais de phénomènes tout à fait différents, tels que l'apparition de « pailles », de « criques » ou de défauts de ce genre.

Pour apprécier la fréquence de ces phénomènes, il faut étudier des « populations » beaucoup plus nombreuses que celles dont on se contente le plus souvent dans les contrôles statistiques. En effet, les « méthodes de contrôle statistique » généralement admises prescrivent des essais portant sur 10 à 20 échantillons, d'où l'on déduit toutes sortes de conclusions. Certaines de ces conclusions portent sur les proportions de pièces à rebuter, en appliquant sans le dire, l'hypothèse que toute la distribution est « normale ». Il y a là une erreur qui peut donner lieu à de graves mécomptes.

M. Dugué. — Cette question est bien intéressante. Un phénomène aléatoire n'est évidemment pas forcément la somme de « petits phénomènes » mais souvent une autre fonction de ces phénomènes élémentaires (par exemple le plus grand). Comme l'a remarqué M. Frechet cela écarte la 2^{e} loi de Laplace (ou loi normale ou de Laplace Gauss) pour donner une plus grande importance à la première loi de Laplace.

M. Baticle souligne l'intérêt de la théorie des valeurs extrêmes pour les ingénieurs, en ce qui concerne les questions de sécurité (par exemple, débits des crues des rivières, fatigue des métaux, etc.). Les résultats expérimentaux ne donnent que des indications imprécises sur la fréquence des valeurs extrêmes et il est souhaitable que la théorie en soit développée en suivant la voie indiquée par le conférencier.

Il faudrait, autant que possible, échapper à l'incertitude qui règne sur l'estimation des paramètres des lois des valeurs extrêmes actuellement connues qui induisent à des probabilités notablement différentes selon les hypothèses sur lesquelles reposent ces lois.

L'idéal serait de trouver pour les valeurs extrêmes une formule analogue *mutatis mutandis*, à celle de Bienaimé-Tchebicheff qui ne fait pas intervenir la loi dont dépend la variable aléatoire en cause.

M. Dugué. — M. Baticle a parfaitement raison de soulever la question du problème de Bienaimé-Tchebicheff pour les valeurs extrêmes. Tous les résultats, « law-free », indépendants de la loi initiale, sont du plus haut intérêt.

Je répète ma conclusion : Ce domaine des valeurs extrêmes est un très vaste champ de recherches.

M. Paul Vincent. — La communication de M. Dugué me paraît d'autant plus intéressante, qu'en soulignant le comportement des valeurs extrêmes, elle met en garde contre des erreurs de jugement que d'excellents statisticiens n'ont pas toujours su éviter.

Je pense ici spécialement à Gumbel, dont le nom a été cité tout à l'heure. Dans une communication présentée en 1949 à l'*Institut international de statistique* (I. I. S.), j'ai été amené à critiquer son « paradoxe du dernier âge » (1), selon lequel les records de longévité seraient d'autant plus élevés que la mortalité serait plus forte.

Ce « paradoxe » résulte de l'emploi abusif d'une loi d'ajustement hors de ses limites de validité. On constate bien, en effet, que la répartition des décès par âge, au sein d'une génération, s'effectue approximativement selon une loi de Laplace-Gauss au-delà de l'âge « normal » (ou modal) au décès, lequel augmente lorsque la mortalité décroît. Mais du fait de la décroissance concomitante de l'écart quadratique moyen de la distribution de l'âge au décès par rapport à cette valeur modale, les courbes de Laplace-Gauss qu'on peut ajuster aux distributions observées sont telles, que les courbes correspondant aux mortalités les plus fortes se situent inévitablement au-dessus des autres pour des âges suffisamment élevées.

Gumbel a cru pouvoir en déduire le « paradoxe » évoqué. C'est exactement comme si, du fait que les erreurs de mesure se répartissent approximativement selon une loi « normale », laquelle présente des branches infinies, on en concluait qu'il y a une probabilité non nulle pour qu'un rectangle dont on a déterminé la longueur par une série de mesures, ait une longueur négative !

(1) Cf. « Quotients de mortalité aux âges élevés — Tables de mortalité de vieillards ». *Bulletin de l'I. I. S.*, **32** — 2, *Berne*, 1950, p. 377 à 388.