JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

RENÉ PORTET

L'utilisation de la statistique dans les recherches pharmacologiques ou biologiques

Journal de la société statistique de Paris, tome 103 (1962), p. 107-128
<http://www.numdam.org/item?id=JSFS_1962__103__107_0>

© Société de statistique de Paris, 1962, 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/

VI

L'UTILISATION DE LA STATISTIQUE DANS LES RECHERCHES PHARMACOLOGIQUES OU BIOLOGIQUES

A — L'expérimentation en biologie

I — *Caractères particuliers*

Le développement actuel de la biologie et des sciences qui lui sont rattachées nécessite, de plus en plus, l'utilisation de méthodes quantitatives. Les progrès accomplis en chimie organique et en chimie biologique ont permis de connaître quelques-unes des réactions se produisant dans l'organisme vivant, d'isoler les produits intermédiaires et terminaux de nombreuses réactions biologiques et surtout de mettre au point des méthodes de dosages quantitatifs des équilibres organiques. Que ce soit en bactériologie, physiologie, pharmacologie ou toxicologie, on ne peut plus admettre, actuellement, comme valables des résultats basés sur l'observation et son interprétation qualitative. Si, dans le domaine de la clinique médicale, ces méthodes sont encore courantes, ceci est dû aux difficultés d'expérimentation sur l'organisme humain; il est vrai que, de plus en plus, des tests quantitatifs précis viennent se substituer à des observations plus ou moins subjectives. Dans le domaine de l'expérimentation biologique sur les animaux, par contre, il est en général possible d'obtenir des résultats numériques qui pourront être interprétés, dans la mesure, toutefois, où l'expérimentation aura été correctement conduite. C'est d'ailleurs dans ces deux domaines qu'intervient le calcul statistique : planification de l'expérience et interprétation des résultats.

En effet, les expériences sont entreprises pour recueillir un certain nombre d'informations; si elles ont été mal organisées, il sera impossible d'en tirer des résultats satisfaisants. L'analyse statistique de mauvais résultats ne peut apporter de conclusions valables; il en est de même, d'ailleurs, pour une mauvaise analyse de résultats correctement obtenus. Ceci veut dire que la préparation d'une série expérimentale doit être faite en tenant compte de la méthode qui sera utilisée pour l'interprétation statistique des résultats. Le maximum d'informations ne sera obtenu que si ces conditions sont respectées.

La préparation minutieuse d'une série expérimentale est plus encore nécessaire en expérimentation biologique que dans toute autre science parce que les phénomènes vitaux présentent une extrême variabilité. L'importance de la variation individuelle provoque des taux d'erreurs considérables; il en résulte que la détermination de leur grandeur doit être une partie essentielle du dosage lui-même. Quand un chimiste veut déterminer la quantité de substance présente dans une solution, il lui suffira de refaire sa titration à deux ou trois reprises, les erreurs sur le dosage étant en général faibles. Par contre, la composition chimique d'un organisme vivant est essentiellement variable, les divers constituants interfèrent les uns sur les autres de façon imprévisible et de telle manière que deux organismes ne peuvent jamais être exactement semblables. Qui plus est, un organisme subit des changements : il sera rigoureusement impossible de retrouver, à deux moments différents, une valeur identique pour une constante biologique donnée. En effet, contrairement à ce qui se passe pour

les phénomènes physiques ou chimiques, une constante biologique n'est pas une valeur exactement déterminée mais est sujette à un certain nombre de fluctuations. C'est précisément pour la détermination de ces « constantes » que les biologistes ont eu tout d'abord recours aux statisticiens. Il était, en effet, nécessaire, au fur et à mesure que les tests chimiques se développaient, de préciser le taux moyen de telle ou telle substance de l'organisme, la vitesse moyenne de telle ou telle réaction et surtout leurs limites de variations pour une espèce animale déterminée. Pour cela, les méthodes classiques de statistique ont été adaptées à l'expérimentation biologique; nous en citerons, par la suite, quelques-unes des plus courantes.

Plus récemment est apparue une nouvelle branche d'expérimentation biologique : l'étude de l'effet de substances chimiques ou pharmacologiques par d'autres moyens que des tests physicochimiques. Ainsi, le développement de l'enzymologie, de l'hormonologie, de l'étude des vitamines, a nécessité l'utilisation de tests purement biologiques : c'est-à-dire l'administration d'une substance à l'animal et l'étude quantitative de la réponse donnée par cet animal. Dans ce cas, la variation biologique et la susceptibilité différente des animaux à une drogue provoquent des marges de dispersion plus grandes encore que dans l'expérimentation biochimique. Il a donc été nécessaire de mettre au point des méthodes statistiques permettant de résoudre ces nouveaux problèmes. Des tests, prévus, à l'origine, pour l'analyse de résultats dont la précision est grande, ont été adaptés à des conditions de mesure non seulement moins précises, mais aussi sujettes à des causes de variation qu'il est impossible de prévoir. Si on veut étudier l'action d'une substance et la comparer à celle d'une autre substance, il ne suffit évidemment pas d'injecter à un animal une dose de la première et à un deuxième animal une dose de la deuxième et de comparer les effets : dans ce cas, les erreurs risquent d'être énormes et il est, de plus, impossible de les calçuler. On peut remplacer l'animal unique par un groupe d'animaux, porter sur des courbes les moyennes des réponses pour chaque dose en fonction de cette dose et les comparer entre elles. Cette méthode réduit l'erreur des déterminations et permet de la calculer approximativement. En réalité, les courbes obtenues expérimentalement ne sont pas des courbes mathématiques simples, mais sont, en général, le résultat d'une infinité de fonctions mathématiques. Les statisticiens se sont attachés à les ramener à un certain nombre de courbes standard qui sont mathématiquement interprétables.

L'approximation qui en résulte est largement compensée par la possibilité nouvelle d'interprétation. Il est donc possible d'utiliser, dans chaque cas particulier, la méthode graphique la plus adaptée aux faits expérimentaux et augmenter ainsi la précision des essais. Des méthodes ont été mises au point pour standardiser les tests utilisés; ces méthodes ont pour but d'adapter la conduite du dosage biologique à la méthode d'interprétation qui sera employée ce qui a pour résultat l'utilisation d'un plus petit nombre d'animaux, associée d'ailleurs à une augmentation de la précision du dosage. Nous en verrons quelques exemples.

Les statistiques permettent donc de savoir, entre plusieurs techniques, laquelle peut conduire, dans chaque cas, à une meilleure précision des résultats. Il ne faut cependant pas oublier que le résultat donné par une série expérimentale dépend, dans une mesure que le statisticien ne peut pas apprécier, de conditions biologiques telles que l'état de santé dans lequel se trouvent les animaux mis en expérience. Le biologiste est obligé d'admettre, sans toutefois en être sûr, que chaque animal est « physiologiquement normal ». Si le contraire se produit, le test risque de ne pas apporter de conclusion ou, pire encore, des conclusions erronées.

II — *Variations biologiques*

1) *Variations entre espèces*

— La valeur des constantes biologiques, d'une part, la sensibilité de l'animal aux substances administrées, d'autre part, peuvent varier de façon considérable d'une espèce animale à l'autre. Il est connu depuis longtemps que le lapin est beaucoup moins sensible que l'homme aux effets des alcaloïdes des Solanacées mydriatiques tels que l'Atropine; de même, le rat est particulièrement sensible aux effets cardiotoxiques des hétérosides de la Digitale. Il existe donc, selon l'espèce, des sensibilités particulières aux diverses drogues pharmacologiques; il en résulte l'impossibilité d'extrapoler les résultats obtenus sur une espèce, à l'ensemble des mammifères par exemple et, en particulier, à l'homme. Si on veut effectuer des essais de toxicité d'une substance, il sera le plus souvent indispensable d'utiliser plusieurs espèces animales de manière à révéler les sensibilités particulières. Dans le cas général, les résultats obtenus lors d'une série expérimentale sur une espèce animale donnée, ne pourront être valables que pour cette espèce et ne seront généralisables à l'intérieur d'une classe, que s'ils se reproduisent pour plusieurs espèces.

2) *Variations individuelles*

A ces variations entre espèces, il faut ajouter les variations à l'intérieur d'une même espèce.

— Ainsi, la *race* peut apporter des susceptibilités différentes; chez l'homme, par exemple, les individus de race noire ou jaune sont moins sensibles à l'action toxique des

GRAPHIQUE I : Production calorique en fonction du poids pour diverses expèces (d'après Le Breton)

dérivés nitrés que ceux de race blanche (on sait que certains phénols dinitrés sont utilisés en thérapeutique comme hypermétabolisants, c'est-à-dire amaigrissants). Seuls les représentants de la race blanche caucasienne sont sensibles à l'action mydriatique de l'éphédrine.

— Le *sexe* peut aussi avoir une influence; chez la femme, le nombre d'hématies par mm^3 de sang est significativement plus bas que celui des hommes; il en est de même, d'ailleurs, pour le métabolisme de base.

— Le *poids* de l'animal a aussi un effet dans la susceptibilité aux drogues. Les petites espèces, telles que les souris et les rats, ont un métabolisme beaucoup plus élevé par unité de poids que les espèces de plus grande taille (graphique I). A l'intérieur d'une même espèce d'ailleurs, le métabolisme n'est pas proportionnel au poids, mais au rapport $\frac{\text{surface}}{\text{volume}}$, la surface étant d'autant plus grande par rapport au poids que l'animal est plus petit. En fait, un petit animal sera moins sensible à une même dose/kg qu'un animal plus gros. Claude Bernard écrivait : « le kilogramme d'un petit lapin exige plus de poison pour mourir comme il exige plus d'oxygène pour vivre ». Ainsi, dans toute l'expérimentation biologique, le poids de l'animal est un caractère extrêmement important. Il est évidemment difficile d'utiliser des animaux ayant rigoureusement le même poids, mais il est toujours possible de déterminer la marge de poids à l'intérieur de laquelle les animaux répondent sensiblement de la même manière; pour cela, il faudra faire des déterminations préalables. On verra mieux, sur un exemple, l'intervention du facteur « poids ». Nous avons eu à nous occuper, dans le laboratoire de pharmacologie de la Faculté de Médecine, de l'étalonnage de solutions de d. tubocurarine (1). Le test utilisé était un test biologique : la chute de la tête du lapin, après injection I. V. d'une certaine dose de substance curarisante. Nous avons porté sur un graphique la dose moyenne active par kg d'animal en fonction du poids des animaux groupés en classes. On constate sur le graphique (graphique II) que, entre 1,6 kg et 3,3 kg, la dose active par kg d'animal diminue à peu près linéairement en fonction du poids. Ceci vérifie l'assertion de Claude Bernard : les petits animaux sont moins sensibles aux curarisants. Par contre, dans les classes extrêmes, les doses moyennes se répartissent de façon différente. Il semble que les tout petits animaux soient plus sensibles au curare peut-être parce qu'ils sont très jeunes. Les très gros lapins, par contre, sont moins sensibles et ceci est peut-être dû à un plus grand pourcentage de lipides dans l'organisme. Les graisses sont en effet un matériau de réserve qui intervient peu dans les réactions générales de l'organisme (il est vrai que ceci est actuellement discuté). Il était évidemment nécessaire d'éliminer les animaux des classes extrêmes et de ne pas en tenir compte pour l'étude statistique des résultats.

— L'*âge* de l'homme ou des animaux a aussi une importance certaine sur leur susceptibilité aux drogues pharmacologiques. On estime, par exemple, que les enfants sont plus sensibles à des substances dépressives du système nerveux central telles que la morphine; ils sont, par contre, moins sensibles aux toxiques minéraux tels que l'arsenic ou les sels de mercure. Chez les vieillards, on constate une action sédative plus faible des substances barbituriques qui, par contre, sont relativement plus toxiques.

— A ces diverses causes de variation, il faut ajouter la variabilité due à *l'état de l'animal* lui-même. Ainsi, certains animaux peuvent présenter des susceptibilités individuelles qu'il est impossible de prévoir. On connaît, par exemple, chez l'homme, des hypersensibilités à l'alcool ou à la nicotine. L'état physiologique de l'animal est lui aussi important; la sensibilité à l'adrénaline est très accentuée par la fatigue musculaire; de même, la valeur du

(1) J. Cheymol, M. Y. Bradel-Gay, Mlle M. Beauvallet et R. Portet, *Thérapie* (1961) *16* 437-457.

Graphique II : Doses actives en fonction du poids

métabolisme est augmentée au moment de la digestion des aliments. Si ces deux facteurs sont faciles à éliminer en maintenant l'animal au repos et au jeûne, il n'en est pas de même pour certaines affections pathologiques qui ne sont pas apparentes; seul le comportement des animaux permettra de les déceler. De toute manière, il est indispensable d'éliminer de l'expérimentation tous les animaux paraissant présenter un déséquilibre quelconque.

3) *Facteurs de variations extérieurs*

— Un autre facteur qui peut apporter des erreurs importantes dans les mesures ou les déterminations biologiques est le facteur « *saison* ». Ainsi, dans beaucoup d'espèces animales, certaines glandes endocrines ont leur sécrétion diminuée en hiver et nettement augmentée en été. Il en est ainsi pour la glande thyroïde et il est bien évident que les injections d'hormone thyroïdienne produiront, sur le métabolisme, un effet quantitatif différent selon la saison de leur administration. La croissance, elle-même, peut être affectée par le facteur saisonnier : ainsi, au laboratoire de pharmacodynamie biochimique, nous avons mesuré la croissance de rats par ailleurs soumis à diverses influences (1). Il ressort très nette-

GRAPHIQUE III. : Influence de la saison sur la croissance de rats maintenus à 5 °C et 30 °C

ment que la vitesse de croissance est plus grande pour les rats nés au printemps et en été que pour ceux nés en hiver (Graphique III).

(1) L. CHEVILLARD, M. C. GUNTZ, M. CADOT, R. PORTET, Communication au Congrès international de Physiologie (Leyde, 1962).

— Autre facteur de variation : *l'accoutumance de l'opérateur*. Lors des tests biologiques, utilisés fréquemment en pharmacologie et toxicologie, l'opérateur doit souvent apprécier une réponse de « tout ou rien ». Il en est ainsi pour les déterminations des doses actives ou des doses mortelles d'une substance. Dans le cas des injections de d. tubocurarine, par exemple, le test quantitatif est donné par le temps écoulé entre le début de l'injection et le moment de la chute de la tête du lapin. L'injection se faisant à vitesse constante, la quantité de substance administrée est directement proportionnelle au temps. Mais la chute de la tête n'est pas un mouvement instantané et laisse à l'opérateur une marge d'appréciation. Il est évident

GRAPHIQUE IV : Variation de la dose active en fonction du temps

que, dans ce cas, il se produit une accoutumance et l'opérateur aura tendance à déceler, de plus en plus précocement, les premiers signes du relâchement musculaire.

Nous avons observé ce phénomène, lors de l'étude précédemment citée, pour laquelle les tests ont été conduits par le même opérateur, sur un intervalle de sept années. En portant, sur un graphique (Graphique IV), les valeurs moyennes des doses actives obtenues pour chaque année, on constate une diminution de la dose moyenne active qui se statilise au cours de la

troisième année seulement. Ceci montre que, dans ce genre de dosage, le facteur « opérateur » intervient de façon considérable. Pour une même série expérimentale, il faut utiliser le même opérateur qui doit être, le plus possible, familiarisé avec la technique.

Il apparaît donc que l'expérimentation biologique est sujette à un nombre considérable de causes d'erreurs. Quelques-unes sont imprévisibles et seule l'analyse statistique des résultats peut permettre d'en estimer l'importance; par contre, d'autres peuvent être évitées par une étude préalable correcte des protocoles expérimentaux et des conditions d'échantillonnage.

B — Conditions d'échantillonnage

Il est, en effet, certain qu'un échantillonnage correctement conduit doit éliminer un certain nombre de causes d'erreurs soit fortuites soit, ce qui est plus grave, systématiques.

Étant donné la variabilité biologique qui apporte une dispersion des résultats plus ou moins importante, il est souvent nécessaire de multiplier le nombre d'animaux mis en expérience. Il en résulte un allongement de la durée expérimentale ce qui peut amener l'intervention d'une variation saisonnière; en outre, la croissance des animaux, pendant cet intervalle de temps, peut provoquer des variations dans le facteur « poids ». Il faut en tenir compte lors de l'analyse des résultats et, au besoin, étudier les variations du paramètre à mesurer en fonction du temps et du poids.

L'expérimentation biologique se faisant, le plus souvent par comparaison de deux ou plusieurs échantillons, il faut que ces échantillons soient homogènes et, d'autre part, qu'ils ne diffèrent entre eux que par le caractère qu'on veut étudier. Si on veut, par exemple, comparer l'action d'une substance dont on ne connaît pas le titre à celle d'une substance étalon, il est nécessaire que les échantillons de souris à qui l'on administre ces deux substances soient, à l'origine, tirés au hasard de la même population. Pour cela, il n'est pas suffisant de plonger la main dans une boîte et d'en retirer successivement les animaux qui veulent bien se laisser attraper; ce sont, en général, les moins vigoureux et les deux échantillons risquent de ne pas être homogènes. On a d'ailleurs constaté qu'il existe une corrélation significative entre l'ordre dans lequel les animaux sont retirés de la cage et leur poids. Pour faire un échantillonnage correct, il faut numéroter les animaux et prélever les numéros au hasard selon une des méthodes classiques. Par la suite, les deux échantillons doivent être conservés exactement dans les mêmes conditions d'élevage et recevoir les mêmes interventions.

C — Analyse des résultats

Ces considérations générales sur l'expérimentation biologique ont mis en lumière la nécessité de l'analyse statistique des résultats. Nous nous proposons d'indiquer les conditions d'application de quelques-unes des méthodes générales d'analyse en nous attachant plus particulièrement à l'interprétation des essais biologiques.

I — *Expression des résultats*

a) *Groupement en classes*

Les séries expérimentales, en biologie, comportent, le plus souvent, un nombre important de résultats. Leur « groupement en classes » est indispensable. On considère, en général,

qu'une dizaine de classes est le nombre permettant la meilleure étude statistique; ces classes doivent, naturellement, être d'égale étendue.

b) *Diagrammes de fréquences*

De plus en plus, dans les publications, les chercheurs biologistes remplacent les tableaux de chiffres dont la lecture est le plus souvent malaisée par des diagrammes ou des graphiques qui ont l'avantage de faire apparaître immédiatement les caractéristiques générales d'une distribution ou l'allure d'un phénomène.

c) *Paramètres d'une distribution*

Comme *paramètre de position*, la moyenne arithmétique des résultats $\frac{\Sigma x}{n}$ est généralement utilisée. Dans certains cas, on calcule aussi la médiane et le mode lorsque la distribution de fréquences est particulièrement asymétrique.

Comme *paramètre de dispersion*, les *quantiles* (on utilisait beaucoup autrefois les *quartiles*) et l'écart moyen arithmétique $\frac{\Sigma (x - X)}{n}$ ne sont plus guère usités. Par contre, la notion de *variance* $v = \frac{\Sigma (x - X)^2}{n - 1}$ et *d'écart-type* ou écart quadratique moyen $\sigma = \sqrt{v}$ sont très généralement employés. Les valeurs moyennes des résultats expérimentaux sont en général accompagnées de l'*erreur standard :* $\frac{\sigma}{\sqrt{n}}$. La détermination du *coefficient de variation* c'est-à-dire du rapport de l'écart-type à la moyenne $V = \frac{100\,\sigma}{m}$ p. 100 peut quelquefois être utile. Il permet, en effet, de comparer la dispersion de diverses distributions étant donné qu'il est indépendant des unités employées pour mesurer les variations.

II — *Principales distributions des résultats biologiques*

a) *Distribution binomiale*

Elle est obtenue en représentant les termes successifs du développement du binôme $(p + q)^n$ avec $p + q = 1$. On rencontre des distributions de ce genre en biologie quand la fréquence d'un événement a une probabilité constante. Ce fait se produit en génétique : si on examine la fréquence d'apparition en première génération d'un caractère récessif, les lois de Mandel prévoient une probabilité de 0,25. Il en est de même pour la détermination du sexe : mâles et femelles apparaissent avec une probabilité approximativement égale; dans ce dernier cas, la distribution est symétrique.

b) *Distribution de POISSON*

C'est une limite de la distribution précédente lorsque la probabilité d'un événement devient très petite (1/100 par exemple); ce cas se rencontre assez souvent en biologie. En hématologie, certains éléments figurés se trouvent en faibles proportions dans le sang. Ainsi, la probabilité de rencontrer un polynucléaire eosinophile par rapport au nombre total de leucocytes n'est que de 1 %. L'éventualité la plus probable est donc que, pour une numération globulaire portant sur 100 éléments, il y ait un seul eosinophile. La représentation de fréquences après n déterminations de ce genre se traduit par une distribution de Poisson.

Le même genre de distribution se rencontre en génétique si on compte le nombre de mutants dans divers échantillons d'une population (la fréquence des mutations est en effet très faible). De même, les résultats d'une numération bactérienne par la méthode de dilution s'expriment aussi selon une distribution de POISSON. La formule mathématique d'une telle distribution est : $y = \frac{m}{r!} e^{-m}$.

c) *Distribution normale ou de GAUSS-LAPLACE*

On obtient une distribution de ce genre quand une grandeur subit l'influence d'un grand nombre de causes de variations petites et indépendantes les unes des autres. C'est ce phénomène qui se produit le plus souvent en biologie, la dispersion biologique étant le résultat d'un nombre considérable de causes de variation. Effectivement, dans la grande majorité des cas, les courbes de fréquences des résultats obtenus dans l'expérimentation biologique se rapprochent de la courbe normale.

Sa formule mathématique est : $y = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}}$ et ne comporte comme paramètres que la moyenne arithmétique m et l'écart-type σ. Ce sont deux valeurs faciles à estimer à partir des résultats expérimentaux. Cette courbe est symétrique (Graphique V),

GRAPHIQUE V : Distribution normale

les points d'inflexion correspondant à $x = m \pm \sigma$. L'aire comprise entre la courbe, l'axe des abscisses et les deux verticales élevées aux points $m \pm n\,\sigma$ est une fraction de l'aire totale. L'intervalle $(-\sigma, +\sigma)$ autour de la moyenne correspond à 68 % de cette aire et l'intervalle $(-2\,\sigma, +2\,\sigma)$ à 95 %. Ceci veut dire que, si on tire, au hasard, le résultat d'une mesure dans un ensemble de résultats répondant à cette loi, il y a une probabilité de 95 %

pour qu'elle ne s'écarte de la moyenne que d'une valeur inférieure à deux fois l'écart-type; 5 % seulement des résultats ne sont pas compris dans l'intervalle (-2σ, $+2\sigma$). En biologie, on admet, en général, que l'intervalle $\pm 2\sigma$ autour de la moyenne est un intervalle de sécurité suffisant : on prend un risque de 5 %. Le seuil de signification généralement admis est 95 %, plus rarement 99 % quand il s'agit de déterminations biochimiques à faible variation. Dans ces conditions, l'estimation d'une moyenne est exprimée de la manière suivante : $X = m \pm \frac{2\sigma}{\sqrt{n}}$ (m étant la moyenne, σ l'écart type et n le nombre d'expériences). Ceci n'est valable que si le nombre de déterminations est suffisamment grand; il peut se produire, cependant, que l'expérimentateur soit obligé de limiter ses épreuves étant donné le prix élevé du matériel animal (singes Rhésus, gros animaux domestiques) ou sa rareté (chien). Dans le cas où l'échantillon est inférieur à 25 ou 30 individus, il est nécessaire de tenir compte de l'erreur commise sur l'estimation de l'écart-type et d'utiliser le facteur de correction « t » de STUDENT. La valeur moyenne sera alors exprimée par $X = m \pm t_P \frac{\sigma}{\sqrt{n}}$ (la valeur de t_P est donnée par les tables de STUDENT en fonction de la probabilité d'erreur et du degré de liberté) (1).

D — INTERPRÉTATION DES RÉSULTATS

Après avoir déterminé la valeur moyenne des résultats, le biologiste se trouve souvent dans l'obligation, soit de comparer entre elles deux distributions, soit de vérifier si une distribution est conforme à une hypothèse ou à une loi déjà formulée. Les statistiques lui procurent pour cela un certain nombre de tests de signification que nous allons maintenant examiner.

1) *Tests de signification*

a) *Comparaison d'une moyenne expérimentale et d'une valeur théorique*

Il suffit de déterminer si la valeur théorique est comprise dans l'intervalle de confiance que nous avons calculé pour la variable expérimentale.

b) *Comparaison d'une distribution expérimentale et d'une distribution théorique*

Ceci se produit fréquemment en génétique quand on doit vérifier si une population obéit aux lois de MANDEL. On utilise, dans ce cas, le test du X^2 de PEARSON.

c) *Comparaison de deux moyennes expérimentales*

Ce problème se pose de façon très générale en biologie; lorsqu'on veut, par exemple, étudier l'action d'une substance ou d'une intervention sur une constante biologique, il est indispensable de comparer les résultats moyens obtenus sur deux séries d'animaux, les uns ayant reçu l'intervention, les autres étant des animaux normaux servant de témoins. On peut, aussi, avoir à comparer les actions respectives de deux doses différentes d'une même substance. Pour résoudre ce problème, on compare les deux valeurs moyennes et les deux variances par la méthode du test de « t » de STUDENT : $t = \frac{m_1 + m_2}{\sqrt{\sigma_0^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$, σ_0 étant la

(1) FISHER and YATES, *Statistical Tables of Biological, Agricultural and Medical Research* (Oliver et Boyd. Ltd, Edinburgh).

covariance des deux distributions qui se calcule par la formule suivante :

$$\sigma_0^2 = \frac{\Sigma (x - m_1)^2 + \Sigma (x - m_2)^2}{n_1 + n_2 - 2}.$$

La valeur de « t » est comparée à celle donnée par les tables de STUDENT.

d) *Comparaison de deux pourcentages*

Elle se fait de la même manière par l'utilisation des tables de STUDENT.

e) *Comparaison de deux variances.*

La comparaison des variances peut être nécessaire pour tester l'homogénéité de deux échantillons prélevés dans la même population. Il peut se produire, par exemple, qu'on ait mis en expérience, à des périodes assez éloignées, des animaux provenant du même élevage. Il est important de savoir si, dans le temps, ces animaux restent identiques à eux-mêmes pour un caractère déterminé. Il suffira, pour cela, de comparer les variances des valeurs moyennes obtenues pour ce caractère. Cette comparaison peut se faire en formant le rapport des deux variances et en comparant sa valeur à celles portées sur les tables de SNEDECOR. Ces dernières comportent, en fonction du coefficient de sécurité désiré, les valeurs limites au-dessus desquelles la divergence des variances doit être considérée comme significative.

f) *Analyse de la variance*

Il peut se produire aussi qu'on ait à étudier la constance d'un caractère quantitatif en fonction d'une variable, le temps par exemple. Ceci s'est produit lors du dosage des solutions de d-tubocurarine. Nous avons constaté que la dose moyenne active diminuait en fonction du temps; il fallait montrer que cette variation était significative. La méthode d'analyse de la variance proposée par FISHER (1) permet de résoudre ce problème. Elle consiste à estimer la variance de l'ensemble des échantillons de deux façons différentes :

— la première a pour but d'éliminer l'influence du facteur qui agit différemment selon les divers lots.

— la deuxième permet, au contraire, de mettre ce facteur en évidence.

Il est alors possible de comparer ces deux estimations par le test de SNEDECOR. Dans le cas qui nous préoccupe, les variances étaient significativement différentes, nous en avons déduit la non-homogénéité de la population, c'est-à-dire une influence du facteur « temps », vraisemblablement en augmentant l'accoutumance de l'opérateur.

g) *Signification d'une moyenne de différences*

En pharmacologie, un autre problème assez fréquent est la comparaison de l'action, sur la même série d'animaux, de deux drogues différentes provoquant une action semblable; ou bien, ce qui revient au même, la comparaison de l'action d'une même substance sur deux organes différents des animaux. Il serait possible de comparer les effets moyens des deux drogues par la méthode de comparaison des moyennes mais on fait intervenir, dans ce cas, un facteur de dispersion qu'il est préférable de supprimer. STUDENT a proposé d'effectuer

(1) FISHER R. A., *Statistical méthods for research Workers* (Oliver and Boyd, Edinburgh).

la différence des valeurs individuelles et d'étudier la signification de l'ensemble de ces différences après calcul du rapport « t » : $t = \frac{m}{\sqrt{\frac{S^2}{n}}}$.

2) *Corrélation existant entre deux variables*

Une autre application importante de l'analyse statistique dans le domaine de la biologie est l'étude de la corrélation existant entre deux caractères rencontrés chez un organisme vivant. Il peut être nécessaire, par exemple, d'étudier, pendant certaines périodes de la croissance d'un animal, les poids relatifs de deux organes. Ou bien, lors d'une expérimentation qui provoque des perturbations organiques, il peut se produire que deux caractères varient et il est intéressant de savoir si ces variations sont liées. Par exemple, en plaçant un animal à des températures basses, on peut comparer les variations de sa consommation d'oxygène et de son quotient respiratoire. Nous nous préoccupons actuellement d'étudier l'effet de certaines substances sur la croissance de larves d'insectes; il est nécessaire de rechercher si ces substances agissent sur le développement pondéral, mais aussi si elles

GRAPHIQUE VI : Diagramme de corrélation entre le poids des larves et la durée de la période larvaire (Blattella Germanica)

accélèrent les métamorphoses et dans quelle mesure ces deux phénomènes sont liés (graphique VI) (1).

La construction d'un *diagramme de dispersion* permet d'avoir une idée de l'allure

(1) A paraître dans les *Annales de la Nutrition et de l'Alimentation* (BOURNIQUE C., CADOT M., PORTET R., GHEVILLARD L.).

de la corrélation entre deux variables. De plus, il est toujours possible de définir un paramètre qui indique dans quelle mesure les variables sont liées : c'est le *coefficient de corrélation.* Le coefficient de BRAVAIS-PEARSON $r = \frac{p}{\sigma_x \cdot \sigma_y}$ s'obtient en fonction de la covariance des deux variables $p = \frac{\Sigma (x - \overline{X})(y - \overline{Y})}{n}$.

Sa valeur est toujours comprise entre -1 et $+1$, la corrélation étant d'autant plus grande que r est plus voisin de $+1$ et d'autant plus faible qu'il est plus voisin de zéro. Il est possible aussi de connaître la sécurité de ce coefficient : on peut en déterminer l'erreur standard et, pour des petits échantillons, FISHER a proposé une transformation « valeur z » permettant d'en faire une estimation plus rigoureuse.

Dans le cas où la corrélation obtenue est assez proche de zéro, il est nécessaire d'en établir la valeur réelle, c'est-à-dire de savoir si la valeur obtenue est significative. Il suffit, pour cela, de la comparer à la *corrélation zéro* dont on peut fixer les limites en fonction du nombre de déterminations et du seuil de signification désiré.

La comparaison de deux coefficients de corrélation est quelquefois nécessaire : on estime leur covariance et on forme le rapport $t = \frac{r_1 - r_2}{S^2_{r_1 \cdot r_2}}$. La signification de leur valeur ainsi obtenue est déterminée sur les tables de STUDENT.

Le calcul du coefficient de corrélation nécessite la connaissance des dispersions marginales σ_x et σ_y qui permettent d'établir la formule mathématique de la *ligne de régression* des valeurs d'une variable sur celles de l'autre. Si on appelle Y la droite de régression de y sur x, sa formule est la suivante :

$$Y = m_y + r \frac{\sigma_y}{\sigma_x} (x - m_x).$$

L'établissement de la formule de la ligne de régression est indispensable quand on veut obtenir, chez un animal, la valeur d'une réponse en fonction de l'intensité d'une action. C'est d'ailleurs ce que nous allons maintenant examiner, en étudiant les relations dose-réponse, problème qui se pose très souvent au pharmacologiste et au physiologiste.

E — UTILISATION DE L'ANALYSE STATISTIQUE DANS LES ESSAIS BIOLOGIQUES

Dans le domaine des essais biologiques, c'est-à-dire des essais de substances pharmacologiques sur l'organisme animal pris en général, les méthodes d'expérimentation et d'interprétation des résultats sont différentes. Les effets de ces substances intéressent un nombre considérable de fonctions physiologiques qui toutes sont en relation plus ou moins directe. Ainsi, si on administre de l'atropine à un animal, des effets seront constatés sur le rythme cardiaque, la pression artérielle, la respiration, le tonus musculaire, la fonction rénale, la conduction nerveuse, la température corporelle, le métabolisme, etc... Ces divers changements physiologiques vont avoir des répercussions l'un sur l'autre; il se produira des phénomènes de compensation ou de régulation de telle manière que l'effet mesuré ne sera en général pas directement proportionnel à la dose administrée. De plus, parmi tous les effets possibles, l'expérimentateur doit choisir celui qui présente le plus de sécurité par sa constance et la simplicité de sa mesure.

Dans tous les cas, cependant, l'intensité de la réponse est une fonction plus ou moins

complexe de la dose administrée liée aux vitesses d'adsorption et d'excrétion de la substance, aux propriétés physicochimiques de cette dernière et à la susceptibilité du sujet. LEAKE (1948) a proposé une formule :

$$I = f\left(D \cdot \frac{\tau A}{\tau E} \cdot p \cdot S\right)$$

I = intensité de la réponse,
D = dose administrée,
τ A et τ E = vitesses d'adsorption et d'excrétion,
p = propriétés physicochimiques de la substance,
S = susceptibilité individuelle.

1 — *Étude de la ligne Dose-Réponse*

Il est possible de porter sur un graphique (graphique VII) les diverses valeurs de la variable « réponse » en fonction de la variance « dose ». La dose X est la variable *indépendante*, la réponse Y est la variable *liée*. Il est aussi possible de tracer la ligne de régression de

GRAPHIQUE VII : Lignes dose-réponse

Y et sur X qui peut être linéaire comme c'est le cas pour l'action de l'éthyluréthane sur les réflexes. En général, la ligne de régression a une forme incurvée; cependant, le plus souvent, la relation entre un effet pharmacologique et la dose administrée se traduit approximativement par une hyperbole équilatère. C'est le cas par exemple pour l'effet de l'adrénaline sur la pression artérielle ou l'action dépressive de la morphine. Il est alors possible de transformer la ligne de régression en une droite en portant en abscisses les valeurs du logarithme de la dose. Cette droite peut être construite par la méthode des moindres carrés ou, plus valablement, après calcul de sa formule qui est de la forme $E = a + b X$, b étant la tangente de

l'angle fait par la droite avec l'axe des X, c'est-à-dire son coefficient angulaire, est défini par $b = \frac{S\,xy}{S\,x^2}$; x et y représentent les diverses valeurs des deux variables. Le calcul des valeurs moyennes de x et de y permet d'obtenir la formule de la droite E.

La détermination de la ligne de régression sera simplifiée si on tient compte du fait suivant : l'intensité de la réponse augmente en progression arithmétique quand la dose croît en progression géométrique. On peut donc choisir les doses administrées de telle manière qu'elles soient en progression de base 2 (par exemple : 1, 2, 4, 8, 16...); dans ce cas, on peut porter en abscisses les logarithmes de base 2 des différentes doses.

La transformation en une droite, de la ligne dose-réponse est très avantageuse car il est théoriquement possible de la construire en connaissant deux de ses points c'est-à-dire en administrant deux doses différentes seulement. Dans la pratique, cependant, il ne faut pas oublier que ces points ne sont pas exactement connus et que nous n'en faisons qu'une estimation. Il sera nécessaire, en général, d'utiliser un minimum de trois ou quatre doses différentes.

2 — *Comparaison de l'effet relatif de deux substances*

Le problème de la comparaison de l'effet d'une substance à essayer à celui d'une substance étalon se pose souvent en pharmacologie. Les effets sont, en général, différents mais il est possible de déduire l'activité de la substance inconnue en la comparant à celle de l'étalon.

a) *Méthode des trois points*

On se contente quelquefois d'utiliser la méthode dite des « trois-points » : pour cela, il est nécessaire de déterminer, par des essais préliminaires, les doses des deux substances ayant un effet voisin. On administre ensuite, à deux séries d'animaux deux doses a et b

GRAPHIQUE VIII : Méthode des trois points

GRAPHIQUE X : Détermination de l'indice thérapeutique

de la substance étalon et, à une autre série, une dose c de la substance inconnue de telle manière que l'effet de c soit compris entre ceux de a et de b. On porte sur un graphique (graphique VIII) les effets de a et de b en fonction du logarithme des doses et on joint ces deux points par une droite. Il suffit de lire l'abscisse correspondant à l'effet de c pour obtenir la dose de substance étalon donnant le même effet que la dose c de substance inconnue. Le rapport de c à cette dose indique l'*activité relative* de la substance inconnue.

Cette méthode n'est valable que si la réponse est sensiblement proportionnelle au logarithme de la dose et, de plus, sa précision est assez faible.

b) *Méthode de BLISS* (1)

On construit, pour la substance à étudier et la substance étalon, deux lignes dose-réponse séparées. Un test préliminaire permet d'établir les doses à administrer de manière que ces deux lignes soient approximativement parallèles. Dans ces conditions, les pentes de ces deux droites ne diffèrent pas de façon significative. L'effet relatif de ces deux substances dépend alors de la différence entre les deux doses moyennes et de la différence entre les deux réponses moyennes à l'ensemble des doses.

Si on appelle M l'effet relatif c'est-à-dire, le rapport de l'effet de la substance inconnue à la substance étalon on peut écrire :

$$M = X_E - X_I + \frac{Y_E - Y_I}{b},$$

X_E et X_I étant les logarithmes des doses moyennes des deux substances; Y_E et Y_1, les réponses moyennes pour l'ensemble des doses. Bliss et Marks (1939) ont donné un exemple d'utilisation de cette méthode pour le dosage de préparations d'insuline. Le test a été conduit en injectant trois doses de chaque substance espacées logarithmiquement (rapport 2), à trois groupes de lapins comportant le même nombre d'animaux (8 lapins). Le calcul est fait par analyse de la variance des diverses distributions en déterminant successivement :

— la somme des carrés des écarts entre la réponse moyenne pour chaque substance et réponse moyenne générale,

— la somme des carrés des écarts entre la réponse moyenne pour chaque dose et la réponse moyenne générale,

— la somme des carrés des écarts entre tous les résultats partiels et la moyenne générale.

On en tire la valeur de l'effet relatif M accompagnée de son intervalle de confiance avec la sécurité désirée. Cette méthode est suffisamment précise en utilisant trois doses pour chaque substance; il est rarement nécessaire d'en utiliser quatre.

3 — *Réponses par « tout ou rien ». Détermination des doses léthales.*

Les réponses par « tout ou rien » sont très fréquentes en biologie; dans ce cas, lorsque l'effet recherché se produit, c'est toujours avec la même intensité. Il en est ainsi, par exemple, pour l'excitation électrique du muscle cardiaque : la contraction d'amplitude maximum a lieu lorsque l'intensité du courant atteint un certain seuil; pour des intensités inférieures, aucune réponse ne se produit. Un phénomène de ce genre se rencontre lors des dosages de

(1) Bliss and Marks, Quart. *J. Pharm. Pharmacol.*, *12*, 182 (1939).

GRAPHIQUE IX : Courbes de mortalité en fonction de la dose

solutions curarisantes par la méthode de la chute de la tête du lapin dont nous avons déjà parlé. L'effet se produit quand la dose administrée est suffisante. Ce sont là des réponses par « tout ou rien » ou réponses « quantales ».

Il en est ainsi pour les essais de toxicité, la réponse étant, dans ce cas, la mort de l'animal. Il est bien certain que les animaux ne répondront pas tous pour la même dose, la sensibilité individuelle intervenant dans ce domaine. Il est donc nécessaire de répartir le lot d'animaux en plusieurs groupes qui recevront chacun une dose différente. Si les sujets sont comparables entre eux, on peut ainsi déterminer, pour chaque dose, un pourcentage de mortalité.

Il est possible de représenter graphiquement (graphique IX) le pourcentage de mortalité en fonction de la dose. La courbe représentative est d'allure sigmoïdale asymétrique. C'est TREVAN qui a établi, le premier, ce genre de courbe en 1927. La dose pour laquelle il n'apparaît aucun décès est la *dose maximum tolérée*; celle qui donne 100 % de décès est appelée *dose léthale 100*. On a coutume, en toxicologie, de déterminer une toxicité par la dose donnant 50 % de mortalité : c'est la *dose léthale 50*, qui correspond, approximativement, au point d'inflexion de la courbe. Cette courbe est asymétrique, mais, lorsque l'échantillonnage est correct, il est possible de la rendre symétrique en exprimant le pourcentage de réponses en fonction du logarithme de la dose (1). On démontre en effet que les logarithmes des doses minimum actives individuelles se distribuent selon une distribution normale. Il existe entre le pourcentage P d'animaux qui réagissent pour une dose donnée et cette dose X une relation linéaire donnée par la formule

$$P = \frac{1}{2\pi}\int_{-\alpha}^{Y} e^{-\frac{1}{2}Y^2}\cdot dY \qquad \text{avec} \qquad Y = \frac{X - \overline{X}}{\sigma_x}$$

Des tables ont été construites par BLISS qui permettent d'obtenir la valeur de Y appelée « *probit* » en fonction du pourcentage de réponses P. Il est donc possible de transformer la sigmoïde en une ligne dose-réponse linéaire qu'on peut construire grossièrement par la méthode des moindres carrés (la D. L. 50 correspond au probit 5). Cette transformation présente l'avantage de permettre une réduction sensible du nombre d'animaux utilisés. En effet, d'après TREVAN, la sigmoïde ne peut être correctement construite qu'en utilisant 6 à 7 doses différentes et un minimum de 30 animaux par dose soit, en tout, près de 200 animaux. Si la régression est linéaire, deux doses différentes suffisent mais, pour plus de sécurité, il est préférable d'en utiliser trois ou quatre.

En réalité, la construction de la ligne dose-réponse par la méthode des moindres carrés donne une précision insuffisante; aussi a-t-on proposé des méthodes plus correctes.

a) *Méthode de MILLER et TAINTER*

Elle permet d'utiliser les deux doses extrêmes (D. L. O et D. L. 100) pour lesquelles la valeur du probit correspondant tend vers l'infini. BARTLETT a proposé de remplacer les pourcentages 0 % et 100 % par $0 + \frac{25}{n}$ % et $100 - \frac{25}{n}$ %, n étant le nombre d'animaux utilisés pour chaque dose. La ligne de régression est construite graphiquement en utilisant les réponses moyennes obtenues pour chacune des doses. La D. L. 50 est lue sur le graphique, elle correspond à l'ordonnée 50 % c'est-à-dire au probit 5. Son écart-type est estimé à partir

(1) BLISS, *J. Amer. Pharm. Assoc.*, *33*, 225 (1944).

des doses correspondant aux probits 4 et 6. L'écart standard de la D. L. 50 est égal à

$$S = \frac{2\,s}{\sqrt{2\,N}},$$

s est la demi-différence de ces deux doses,
N le nombre total d'animaux des groupes compris entre les probits 3,50 et 6,50.

Cette méthode est rapide mais ne donne que des valeurs approchées pour la D. L. 50 et pour la pente de la droite de régression.

b) *Méthode de BLISS*

On construit d'abord une *ligne provisoire* à partir des réponses moyennes obtenues pour chaque dose et exprimées en probits qu'on appelle *probits empiriques*. On détermine ensuite les *probits attendus* qui correspondent aux ordonnées lues sur la *ligne provisoire* pour les diverses doses. Bliss a donné des formules permettant d'obtenir, à partir des valeurs précédemment déterminées et d'un facteur de pondération, un *probit corrigé* pour chacune des doses. Ces *probits corrigés* permettent de calculer la formule d'une ligne de régression de première approximation qui sera en général suffisante. Pour plus de précision, on peut construire une droite de deuxième approximation en utilisant la droite de première approximation comme ligne provsoire et ainsi de suite.

Cette méthode implique des calculs assez longs mais permet d'obtenir la pente de la ligne de régression qui correspond à l'étalement des doses toxiques. Pour une même drogue, cet étalement peut, en effet, varier d'une espèce animale à l'autre. Connaissant sa valeur et la marge de sécurité de la détermination, il est possible de comparer l'effet toxique sur deux espèces ou deux races différentes ou de tester l'homogénéité d'une population. La D. L. 50 peut aussi être obtenue, par cette méthode, de façon assez rigoureuse; le calcul de ses limites de sécurité permet la comparaison de la toxicité de deux substances différentes par les tests statistiques classiques.

4 — *Détermination de l'Indice Thérapeutique*

Pour certaines substances, le pharmacologue doit déterminer *l'indice thérapeutique* c'est-à-dire le rapport existant entre la dose efficace et la dose toxique. La dose efficace se mesure comme la dose léthale par la méthode de Bliss en établissant la quantité de substance susceptible de provoquer l'effet désiré chez 50 % des sujets. Elle s'exprime par le symbole E. D. 50 (effective dosis 50). L'indice thérapeutique qu'on appelle aussi *facteur relatif de sécurité* est le rapport E. D. 50/D. L. 50 — (Graphique X). On estime que la valeur de cet indice doit être au moins égale à 10, pour que la drogue soit sans danger aux doses actives. On peut aussi établir un *facteur absolu de sécurité* (F. A. S.) qui est le rapport ED. 100/DL. 0.

F — Conclusion

Nous avons passé en revue un certain nombre de méthodes statistiques actuellement utilisées dans l'expérimentation biologique et pharmacologique. Ces méthodes présentent généralement un double intérêt. Elles permettent, tout d'abord, d'utiliser, dans les séries expérimentales, la plus petite quantité de substance active ainsi que le plus petit nombre

possible d'animaux. Le biologiste travaille, en effet, dans tous les cas, sur de la matière vivante qu'il lui est quelquefois difficile de se procurer et toujours impossible de conserver identique à elle-même. L'établissement des protocoles expérimentaux et de l'échantillonnage est conditionné par ces deux impératifs : économie du matériel vivant et homogénéité de ce matériel. Les tests statistiques correctement adaptés à la catégorie d'expérience permettent de limiter l'effectif total utilisé et de vérifier si les divers groupes sont homogènes.

Le grand intérêt de l'utilisation des méthodes statistiques reste, cependant, la possibilité qu'elles offrent d'exprimer et d'interpréter correctement les résultats obtenus expérimentalement. Dans une science, où la variabilité individuelle implique une dispersion considérable de ces résultats, il est absolument nécessaire d'exprimer les valeurs moyennes accompagnées de leur intervalle de sécurité. De plus, mise à part la détermination des constantes biologiques et encore ne sont-elles valables que dans des conditions bien définies, l'expérimentation implique presque toujours la comparaison de deux ou plusieurs groupes d'animaux. Cette comparaison ne peut être correctement faite qu'en utilisant des tests de signification quantitatifs. On voit, fort heureusement, disparaître, dans les publications, des termes tels que « différence hautement significative » et apparaître, à la place, des estimations de la probabilité de cette différence.

Enfin, pour les tests biologiques plus spécifiquement utilisés par le pharmacologue, l'utilisation de méthodes d'analyse strictes s'avère plus encore indispensable. Les essais biologiques de drogues ou substances médicamenteuses sur l'animal ne sont, en effet, qu'une étape avant les essais cliniques pratiqués sur l'homme. Si ces derniers sont indispensables parce qu'ils peuvent seuls donner la certitude de l'efficacité et de l'absence de toxicité des médicaments, ils ne peuvent être entrepris avec le minimum de risques que dans la mesure où des essais sur animaux, bien conduits et bien interprétés, ont parfaitement déterminé les doses limites acceptables.

René Portet

Directeur-adjoint de laboratoire

École Pratique des Hautes-Études (1).

DISCUSSION

M. F. Milhaud. — Dans les applications de la Statistique en biologie et en médecine, je pense que l'on abuse de l'affirmation que 95 % des résultats que l'on peut observer dans un ensemble parent sont compris dans un intervalle de $\pm 2\sigma$.

L'affirmation est légitime si on considère la fréquence d'un phénomène susceptible de tout ou rien observé dans des conditions données, par exemple sous l'influence d'une dose déterminée de toxiques. En effet, nous sommes là dans le schéma de Bernoulli, si toutefois il s'agit d'une fréquence suffisante pour appliquer celui-ci et non pas la loi des petites probabilités de Poisson. L'application de la loi de Laplace et des valeurs de $\theta(x)$ qui lui correspondent est justifiée.

Mais il n'en serait pas de même si on examinait la courbe de fréquence des doses nécessaires pour obtenir une réaction biologique donnée.

---

(1) Laboratoire de Pharmacodynamie Biochimique (Collège de France).

D'une façon générale, il n'y a aucune raison d'appliquer les tables de θ (λ) quand on ne se trouve pas dans le schéma de Bernouille. Il ne suffit pas qu'une courbe de fréquence construite expérimentalement paraisse unimodale et que son intégrale ressemble à une ogive Galton.

Un cas particulièrement intéressant est celui où on détermine une « Constante » morphologique ou biochimique. Ce serait une erreur grave d'affirmer systématiquement qu'un écart de 2σ à la moyenne exclut de l'ensemble parent caractérisé par celle-ci. Il est peu vraisemblable que la courbe de fréquence de cette « Constante » soit une courbe de Gauss. Considérons en effet 2 caractères mesurables. Si la courbe de fréquence de 1e est une courbe de Gauss obéissant à la loi de Laplace, il est mathématiquement invraisemblable qu'il en soit de même de celle de 2e, à moins que les 2 caractères soient totalement indépendents, ou qu'ils soient fonctions linéaires l'un de l'autre, ce qui est très rare.

Remarquons que si les résultats d'une mensuration ou d'un dosage ne se répartissent pas selon une courbe normale, leurs moyennes se répartissent cependant de même, et le σ de student peut encore servir à prouver l'hétérogénéité de 2 populations.

M. Larrieu. — J'ai fait observer que si la loi normale n'est effectivement pas toujours vérifiée dans les cas pratiques étudiés, cette remarque a surtout une valeur théorique, car dans la majorité des cas et sauf dissymétrie accusée des répartitions de fréquence, les *résultats numériques*, fournis par la loi normale sont très voisins des résultats numériques fournis par les vraies lois pourvu qu'on ne s'attache pas aux probabilités très grandes ou très petites. En tout état de cause dans ces cas-là, les écarts observés entre loi « vraie » (pour autant qu'elle existe !) et loi normale sont inférieurs aux erreurs provenant de l'imprécision des mesures et des perturbations diverses qui se rencontrent inévitablement dans tout phénomène physique observé.

En réponse à une autre intervention, M. Larrieu a ajouté qu'il y avait un danger supplémentaire à rechercher une loi de probabilité à plus de deux paramètres car une telle loi nécessairement plus souple s'adaptait mieux à l'histogramme résultant d'un phénomène observé donc *passé*, mais risquait de représenter très mal le comportement *futur* du même phénomène. Autrement dit une loi plus souple est plus précise, mais elle risque de se trouver beaucoup plus « à côté » du phénomène « vrai »; elle provoque donc une fausse impression de sûreté.

J. Dufrenoy. — Je suis pleinement d'accord avec notre collègue qui vient d'attirer l'attention sur l'usage abusif de la « Distribution normale » comme Distribution théorique à laquelle on s'efforce de comparer une Distribution observée.

C'est pourquoi on ne peut que féliciter M. Portet d'avoir fait mention de la Distribution des Séries de Poisson, ainsi que de la Distribution log. normale, que l'emploi d'échelle de probabilité normale (dite Gaussienne) en ordonnée, et d'échelle log., en abscisse, permet de représenter par une droite.