Zeitschrift: Archives des sciences physiques et naturelles
Herausgeber: Société de Physique et d'Histoire Naturelle de Genève
Band: 29 (1947)

Artikel: Loi probabilitaire complètement formulée dans la théorie de l'estimation
Autor: Féraud, Lucien
DOI: https://doi.org/10.5169/seals-742269

lithologiques du Carbonifère orientés N 25°-30° E tandis que la zone carbonifère proprement dite (Carbonifère, Trias, Schistes lustrés) a une orientation moyenne N 50° E. Il suggère que la schistosité alpine doit oblitérer la stratification carbonifère primitive et que le Trias n'est pas concordant. Ces deux opinions sont contradictoires. Il ne nous appartient pas de les départager par une seule constatation. Mais dans le domaine de notre modeste étude, qui suit des niveaux lithologiques graphiteux carbonifères sur 3 km de longueur, l'orientation est nettement N 55°-65° E (donc parallèle à la zone carbonifère proprement dite N 50°-60° E) et correspond en bien des points à la schistosité. Nous examinerons prochainement s'il en est autrement ailleurs.

Université de Genève.
Laboratoire de Géophysique.

BIBLIOGRAPHIE

1. POLDINI, E. *Sur l'existence de courants électriques naturels liés aux gîtes d'anthracite valaisans.* Compte rendu des séances de la Soc. Phys. et Hist. nat. de Genève, vol. 60, nº 3, août-décembre 1943, p. 274.
2. —— *Les phénomènes de polarisation spontanée électrique du sous-sol et leur application à la recherche des gîtes métallifères.* Mémoires Société vaudoise des Sciences naturelles, nº 40, vol. 6, nº 1, 1938.
3. OULIANOFF, N. *Les anciens massifs du Mont-Blanc et de l'Aar et l'orogénèse alpine.* Eclogae Geologicae Helvetiae, vol. 37, nº 1, 1944, p. 31.

Lucien Féraud. — *Loi probabilitaire complètement formulée dans la théorie de l'estimation.*

Il arrive tous les jours qu'une mesure expérimentale d'une grandeur ayant donné une *valeur observée* x_0 on adopte celle-ci pour *valeur vraie* a de cette grandeur, bien que l'on sache que le procédé de mesure est entaché d'erreurs accidentelles. La mesure x de la grandeur est alors une variable aléatoire admettant une loi de probabilité que nous écrirons

$$\left\{ \begin{array}{c} x \\ \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2}} \end{array} \right\} \qquad (1)$$

en supposant connu l'écart quadratique moyen et en le prenant égal à l'unité, pour fixer les idées. On ignore, par contre, la moyenne a de cette loi normale. On estime ce paramètre a en adoptant la valeur observée x_0: c'est une estimation qui conduit à une seule valeur, on peut l'appeler une *estimation par un point*.

Dans un autre mode d'estimation il ne s'agit plus d'assigner une valeur déterminée au paramètre a mais seulement de définir un intervalle qui le comprend: c'est *l'estimation par un intervalle*. C'est une estimation de cette deuxième sorte que nous allons considérer sur un exemple excessivement simple: A partir de la distribution (1) ci-dessus et du résultat x_0 fourni par une mesure comment estimer, par un intervalle, le paramètre a ?

La réponse peut être donnée simplement et sa signification apparaît nettement si l'on recourt à la notion de *loi probabilitaire complètement formulée*, introduite dans des travaux précédents [1]. En modifiant légèrement les dénominations employées dans ces travaux, nous réunissons, dans le cas qui nous occupe, sous le nom de loi probabilitaire complètement formulée:

1° la distribution (1) pour une valeur déterminée a_1 du paramètre;

2° une première décision fixant un degré de probabilité α;

3° une deuxième décision désignant pour une distribution (1) et pour un α déterminés, une portion de l'axe des x à laquelle est affectée une probabilité égale à α, soit V_{a_1} cette portion de l'axe des x.

Si l'on sait que $a = a_1$ on a ainsi une loi probabilitaire complètement formulée; si l'on fait seulement l'hypothèse que $a = a_1$ on a une *hypothèse probabilitaire complètement formulée*.

Toute valeur de a donne une hypothèse admissible puisque nous ignorons complètement la valeur de ce paramètre. Nous

[1] Cf. *Paramètre ignorable dans une loi de probabilité*. Compte rendu des séances de la Soc. de Phys. et d'Hist. nat. de Genève, vol. 62, n° 2, avril-juillet 1945, pp. 58-59, et *Les instruments mathématiques de la statistique*. Rouge, Lausanne, et Gauthier-Villars, Paris, 1946. Note p. IV.

allons mettre en évidence une classe d'hypothèses qui pourront être regardées comme plus intéressantes que les autres. Supposons admise une hypothèse déterminée $a = a_1$. Qu'en tirerons-nous ? Nous sommes alors en présence d'une loi probabilitaire complètement formulée. Le degré de probabilité α est choisi, nous le supposons inférieur à $1/2$. Nous choisissons V_{a_1} de telle sorte que son complémentaire $V^*_{a_1}$, par rapport à l'axe des x tout entier, soit un intervalle d'un seul tenant. En vertu de $\alpha < \frac{1}{2}$ les extrémités de $V^*_{a_1}$ seront de part et d'autre de a_1: nous les notons $a_1 - l$ et $a_1 + \lambda$. Nous arrivons donc à la conclusion suivante: « au degré de probabilité α, la mesure donnera un résultat compris dans $V^*_{a_1}$. »

Ceci posé, il est clair que toute hypothèse $a = a_1$ pour laquelle $V^*_{a_1}$ ne contiendra pas x_0 sera en contradiction avec le seul fait expérimental dont nous disposons. Par contre il n'y aura pas contradiction si x_0 appartient à $V^*_{a_1}$. Nous avons donc une raison pour distinguer, parmi toutes les hypothèses admissibles celles que n'infirme pas le résultat x_0 et, par exemple, de commencer par poursuivre la vérification de celles-ci. Pour ces hypothèses, x_0 appartient à $V^*_{a_1}$, c'est-à-dire

$$a_1 - l < x_0 < a_1 + \lambda$$

d'où

$$x_0 - \lambda < a_1 < x_0 + l$$

L'intervalle $x_0 - \lambda$, $x_0 + l$ définit les valeurs du paramètre a qui donnent des hypothèses plus intéressantes que les autres (au sens qui vient d'être indiqué): c'est *l'intervalle de confiance* (confidence interval) dans la terminologie de Neyman et Pearson. Il va de soi que cet intervalle dépend, pour un α donné, du choix de V, ce dernier n'étant pas nécessairement un intervalle, mais pouvant être un ensemble quelconque pourvu que P (V) $= \alpha$. On prend souvent V_{a_1} tel que son complémentaire $V^*_{a_1}$ soit un intervalle symétrique par rapport à a_1, c'est-à-dire tel que $l = \lambda$, ce qui fait que l'intervalle de confiance ne dépend plus que de α, pour un x_0 donné, mais il n'y a aucune nécessité théorique qui impose le choix d'un intervalle $V^*_{a_1}$ symétrique par rapport à a_1.

Le raisonnement est essentiellement le même dans le problème dit « des probabilités inverses » dont voici un cas particulier.

Soit une urne contenant 10^8 boules parmi lesquelles B blanches et $10^8 - B$ noires. On a fait un tirage dans les conditions dites « au hasard ». Il a donné une boule blanche. Que peut-on dire de B ?

On peut dire d'abord $1 \leqslant B \leqslant 10^8$ et à première vue il semble que l'on ne puisse rien dire de plus. Ce qui est vrai en un certain sens.

Toutefois, si l'on considère toutes les hypothèses admissibles

$$B = 1, 2, \ldots \qquad 10^8$$

ne peut-on accorder à certaines d'entre elles une préférence ? Nous allons montrer dans quelles conditions et ce que l'on peut entendre par là.

Il reste à poursuivre la vérification de chacune des hypothèses admissibles. Supposons que l'une d'elles soit admise $B = B_1$. Qu'en fera-t-on ? On est alors devant un tirage au hasard dans une urne de composition connue. Décidons d'adopter le degré de probabilité $\alpha = 10^{-6}$, c'est-à-dire de considérer que tout ensemble auquel est affectée une probabilité inférieure ou égale à 10^{-6} sera *vide de résultats.*

Ainsi, au degré de probabilité 10^{-6} et pour un seul tirage:

si $1 \leqslant B_1 \leqslant 10^2$ on conclura qu'il ne sortira pas une boule blanche,

si $10^2 < B_1 < 10^8 - 10^2$ on ne pourra tirer aucune conclusion,

si $10^8 - 10^2 \leqslant B_1$ on conclura qu'il ne sortira pas une boule noire.

Par suite, dans le premier cas et dans le premier cas seulement, la sortie d'une blanche, qui a été constatée, est en contradiction avec une conséquence de l'hypothèse. L'adoption de toute hypothèse définie par un $B_1 \leqslant 10^2$ est en contradiction avec le seul fait expérimental que l'on possède. Aucune des hypothèses $B_1 > 10^2$ ne conduit à une contradiction.

C'est pourquoi on accordera une préférence aux hypothèses qui n'ont pas déjà rencontré un démenti, c'est-à-dire à la classe $B_1 > 10^2$. En particulier on a une raison pour commencer par poursuivre la vérification des seules hypothèses $B_1 > 10^2$. L'ensemble $10^2 < B \leqslant 10^8$ est *l'ensemble de confiance* (confidence set) dans la terminologie de Neyman et Pearson. Il va sans dire qu'il peut être beaucoup plus restreint et par conséquent beaucoup plus intéressant, beaucoup plus efficace que dans l'exemple schématique que nous venons de considérer.

Bien entendu, nous ne décidons pas entre les hypothèses admissibles d'une manière absolue mais nous décidons, et encore provisoirement, seulement en vue de ce que nous voulons faire: se servir de l'hypothèse qui sera admise pour en tirer une conclusion au degré de probabilité adopté 10^{-6}. Mais cela détermine l'action à entreprendre, au moins dans une certaine mesure.

Remarques.

1. Dans le second problème que nous venons de traiter, le choix de V ne comporte plus d'arbitraire comme dans le premier. Il n'y a qu'un seul ensemble V qui permet d'arriver à une conclusion. En conséquence l'ensemble de confiance ne dépend que de α. Par contre, dans le second problème apparaissent moins nettement l'essentiel du raisonnement qui est à la base de la théorie de l'estimation et le rôle tenu dans ce raisonnement par la loi probabilitaire complètement formulée.

2. Le second problème peut être généralisé:

1) On peut considérer le cas où le paramètre inconnu est un coefficient p représentant la probabilité d'arrivée d'un événement E; le coefficient p n'est plus une fraction de dénominateur donné;
2) On peut encore considérer n tirages au hasard qui ont amené r fois l'événement E. Ce problème a été traité par Clopper et Pearson [2] et les graphiques qu'ils ont

[1] *The use of confidence or fiducial limits illustrated in the case of the binomial.* Biometrika, XXVI, 1934, p. 404.

établis permettent de déterminer aisément l'ensemble de confiance lorsque l'on prend pour degré de probabilité 0,05 ou 0,01 et un sensemble V* « central ».

RAPPORT DE M. LÉON-W. COLLET

En fin de séance, M. Léon-W. Collet fait un bref rapport sur une expédition océanographique dirigée par le géologue suédois Petersen qui s'occupera de l'étude de la sédimentation dans le fonds des mers. Cette étude sera facilitée grâce à un appareil nouveau dû à Petersen permettant de prélever des échantillons de sédiments marins sur une épaisseur de 20 mètres, tandis que les anciens dispositifs n'opéraient que sur 2 à 3 mètres.

Séance du 6 mars 1947.

Emile Cherbuliez et **Pierre Baudet.** — *Recherches sur la caséine et sa transformation en paracaséine.*

On ignore encore ce que représente au juste, au point de vue chimique, la transformation irréversible que subit la caséine du lait lors de l'action du labferment, transformation qui aboutit à la paracaséine différenciable de la caséine jusqu'à présent uniquement par l'insolubilité de son sel de calcium.

On doit à Warner [1] une méthode de séparation de la caséine, par précipitation fractionnée à des pH déterminés, à $+ 2^o$, en deux fractions, dites α et β, qui se sont montrées homogènes à l'électrophorèse.

Nous avons constaté que α donne en présence de l'ion $Ca^{\cdot\cdot}$, à des pH de 5,3 à 7,0, des solutions d'apparence laiteuse analogues à celles de la caséine native ou du mélange α et β reconstituant la caséine primitive. La fraction β, par contre, est précipitée dans le même domaine de pH lors de l'addition de chlorure de calcium aux solutions de son sel sodique. La solu-

[1] R. C. WARNER, *J. Am. Chem. Soc.*, vol. 66, p. 1725 (1944).