ACADÉMIE DES SCIENCES.

SÉANCE DU LUNDI 4 FÉVRIER 1957.

PRÉSIDENCE DE M. LÉON BINET.

MÉMOIRES ET COMMUNICATIONS

DES MEMBRES ET DES CORRESPONDANTS DE L'ACADÉMIE.

M. le **Président** annonce le décès de M. **Paul Walden**, Correspondant pour la Section de Chimie, survenu à Gammertingen, Würtemberg (Allemagne), le 22 janvier, et celui de M. **Charles Killian**, Correspondant pour la Section de Botanique, survenu à Récife (Brésil), le 27 janvier. Il invite l'Académie à se recueillir en silence pendant quelques instants, en signe de deuil.

Les Notices nécrologiques d'usage seront déposées en l'une des prochaines séances, sur M. *Paul Walden* par M. **Georges Chaudron**, sur M. *Charles Killian* par M. **Henri Humbert**.

CALCUL DES PROBABILITÉS. — *Sur la distance de deux lois de probabilité.* Note de M. **Maurice Fréchet**.

Une formule explicite et simple est donnée pour représenter la distance de deux lois de probabilités quand on utilise la première des trois définitions de cette distance proposées par Paul Lévy. Une quatrième définition est proposée.

Paul Lévy a proposé (1) trois définitions de la distance de deux lois de probabilité L, L'.

Nous examinerons ici la première, qui est la plus intuitive et qui, contrairement à ce que l'on aurait pu attendre, conduit à des formules très simples.

Selon cette première définition, la distance (L, L') de ces deux lois est la borne inférieure de la « distance globale »

$$([X], [Y])$$

de deux nombres aléatoires X, Y qui ont respectivement L et L' comme lois de probabilités individuelles, quand la corrélation entre X et Y varie.

Il est clair que la distance (L, L') va dépendre de la définition adoptée pour la distance globale de X et de Y.

Nous prendrons ici, pour cette distance, l'écart quadratique moyen de X et de Y. En appelant $F(x)$, $G(y)$, $H(x, y)$ les fonctions de répartition respectives de X, de Y et du couple (X, Y), cet écart quadratique moyen D_H a pour carré

$$D_H^2 = \mathfrak{M}_H(X-Y)^2 = \iint_P (x-y)^2\, d_x d_y\, H(x, y),$$

l'intégrale étant étendue au plan P de xOy.

Un calcul immédiat montre que l'on a, avec les notations habituelles

$$(1) \qquad D_H^2 = (\overline{X} - \overline{Y})^2 + \sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y r_H.$$

Dès lors, on peut écrire pour la distance (L, L') des deux lois

$$(2) \qquad \boxed{(L, L') = \sqrt{(\overline{X} - \overline{Y})^2 + \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y},}$$

où ρ est la borne supérieure (nécessairement $\leqq 1$) du coefficient de corrélation (linéaire) r_H entre X et Y quand $H(x, y)$ varie. Nous allons calculer ρ.

Cas d'une même loi réduite. — Il y a un cas particulier important, où l'on peut donner immédiatement la valeur de ρ.

C'est celui où X et Y ont *la même loi réduite* [par exemple si X et Y ont chacune une densité de probabilité constante sur deux segments finis respectifs; si X et Y obéissent à la loi de Laplace (dite normale), etc.]. *Dans ce cas particulier*, on a $\rho = +1$, d'où

$$(3) \qquad \boxed{(L, L') = \sqrt{(\overline{X} - \overline{Y})^2 + (\sigma_X - \sigma_Y)^2}.}$$

Cas général. — Dans le cas général, non seulement r_H, mais aussi ρ seront inférieurs à 1. L'expression de (L, L') sera moins simple, mais s'obtiendra encore explicitement.

Nous pouvons, en effet, considérer $H(x, y)$ comme définissant un « tableau de corrélation » dont les « marges » sont définies par les fonctions $F(x)$, $G(y)$.

Or nous avons montré [2] que l'ensemble des fonctions $H(x, y)$ est identique à l'ensemble des fonctions de répartition dont les valeurs sont comprises entre deux d'entres elles, à savoir

$$(4) \qquad \begin{cases} H_0(x, y) = \mathrm{Max}[F(x) + G(y) - 1, 0], \\ H_1(x, y) = \mathrm{Min}[F(x), G(y)]. \end{cases}$$

Poursuivant cette étude (d'ailleurs dans un autre but), Salvemini avait conjecturé que $\mathfrak{M}_H(X-Y)^2$ atteignait sa borne inférieure pour $H \equiv H_1$. Bass a énoncé [3] le résultat correspondant pour r_H dans le cas où X et Y sont bornés (et m'en a communiqué la démonstration). Un peu plus tard, Dall'Aglio [4] a validé la conjecture de Salvemini dans un cas plus général encore.

On a donc, sous des conditions suffisantes peu restrictives (et qui ne sont peut être pas nécessaires)

$$(L, L') = \sqrt{(\bar{X} - \bar{Y})^2 + \sigma_X^2 + \sigma_Y^2 - 2 r_{H_1} G_X \sigma_Y}.$$

Dall'Aglio donne la valeur explicite de $(L, L')^2$ par une combinaison d'intégrales.

On peut en donner une expression différente et peut être plus simple et plus commode sous la forme suivante.

Nous avions fait observer que si $F(x)$ et $G(y)$ sont continus et croissantes, $H_1(x, y)$ est la fonction de répartition qui correspond au cas où il y a une relation fonctionnelle croissante entre X et Y, soit

$$F(x) = G(y) \qquad \text{ou} \quad y = \lambda(x)$$

d'où

$$\boxed{(L, L') = \sqrt{\int_{-\infty}^{+\infty} [x - \lambda(x)]^2 \, dF(x)}.}$$

Une quatrième définition. — La première définition de la distance (L, L') due à Paul Lévy répond bien à notre intuition. Mais elle exige la détermination préalable d'une borne inférieure.

Nous avons vu que cette détermination préalable a été faite dans le cas où

$$([X], [Y])_H = \sqrt{\mathfrak{M}_H (X - Y)^2}.$$

Mais elle reste à faire pour d'autres choix de la distance globale. Et non seulement ceci reste à faire, mais ce sera, en général, un problème un peu délicat, comme le montre le fait que Bass et Dall' Aglio ont dû imposer des conditions restrictives à $F(x)$ et à $G(y)$.

Pour esquiver ces deux difficultés, nous allons proposer une quatrième définition. Si celle-ci les supprime, en effet, il faut reconnaître qu'elle est moins intuitive que celle de Lévy.

Nous poserons, *a priori*, sans explication

$$(L, L') = ([X], [Y])_{H_1}.$$

On peut en donner deux justifications. D'une part, elle coïncide avec celle de Lévy, au moins dans le cas, examiné plus haut, où la distance globale de X et Y est égale à leur écart quadratique moyen. D'autre part, on peut prouver que cette valeur de (L, L') vérifie bien même dans le cas général les trois conditions imposées à la notion de distance.

Remarque. — Salvemini avait déjà observé [1] que « l'indice quadratique de dissemblance » de Gini est égal au second membre de (3) quand X et Y sont *laplaciennes*. D'où il résulte que *dans ce cas*, cet indice est égal à la « distance » des deux lois de fréquence de X et de Y. On peut même prouver qu'il en est ainsi pour deux lois de fréquence quelconques de X et de Y.

Cette Note sera développée dans le numéro spécial dédié à Paul Lévy du Journal de l'École Polytechnique.

(1) *Voir* p. 331 de la 2e édition du Premier Livre de nos *Recherches théoriques modernes sur le Calcul des Probabilités*, chez Gauthier-Villars, Paris.

(2) *Ann. Univ. Lyon*, Section I, 1954, p. 53-77. *Voir* aussi un complément dans notre Note sous le même titre dans ces *Comptes rendus*, **242**, 1956, p. 2426.

(3) *Comptes rendus*, **240**, 1955, p. 839.

(4) *Ann. Sc. Norm. Sup.*, Pisa, série III, **10**, 1956, p. 35-74.

(5) *L'indice quadratico di dissomiglianza tra distribuzioni gaussiane*, *Societa italiana di Statistica*, Atti della XIV Riunine Scientifica, Roma, 1954, p. 1-7.

HYDRAULIQUE. — *Construction graphique et grandeurs relatives pour l'étude des surpressions dans les conduites.* Note (*) de M. **Léopold Escande.**

L'emploi de grandeurs relatives permet d'appliquer directement la méthode graphique classique de Bergeron à l'étude de problèmes généraux de surpressions.

Soit q_0 le débit de régime permanent, a la célérité des ondes, S la section de la conduite, en M, à l'instant t, posons

$$\xi_* = \frac{aq_0}{gS}, \qquad \mu = \frac{2L}{a}.$$

La première quantité représente la surpression consécutive à une fermeture instantanée du débit q_0, la seconde, la durée de l'aller-retour d'une onde d'un bout à l'autre de la conduite.

F et f désignant les ondes de surpression, de célérité $\pm a$, qui se croisent en M à l'instant t, où la surpression et le débit sont ξ et q, considérons les grandeurs relatives suivantes :

$$\xi' = \frac{\xi}{\xi_*}, \qquad F' = \frac{F}{\xi_*}, \qquad f' = \frac{f}{\xi_*}, \qquad q' = \frac{q}{q_0}, \qquad t' = \frac{t}{\mu}.$$

On déduit alors immédiatement des équations fondamentales d'Allievi, les relations suivantes :

$$\xi' = q' - 1 + 2F'$$
$$\xi' = -(q' - 1) + 2f'$$

D'après ces relations, le point $M'(q', \xi')$, point de fonctionnement à l'instant relatif t', dans le plan des axes $Oq'\xi'$, satisfait aux conditions suivantes : les points M' correspondant à une même valeur des ondes F' ou f' sont situés sur une même droite Φ' ou φ', inclinée de $+45°$ pour Φ' ou $-45°$, pour φ' sur l'axe des abscisses.

De même, les points de fonctionnement, au droit du distributeur, sont situés