CALCUL DES PROBABILITÉS. — *Les problèmes de diagnostic séquentiel.* Note (*) de MM. **Jean Ville** et **Marcel Paul Schützenberger**, présentée par M. Émile Borel.

Faute d'un meilleur terme, nous appellerons *problèmes de diagnostic séquentiel*, les problèmes du type suivant qui ne semblent pas avoir encore été envisagés systématiquement :

Soit une situation expérimentale. A chaque instant les hypothèses incompatibles encore admissibles, $H_1, H_2, \ldots, H_n$, constituent un ensemble $\mathfrak{H}$ contenant l'unique hypothèse H^* qui est vraie. Pour déterminer celle-ci, on décompose $\mathfrak{H}$ en ν sous ensemble disjoints $\mathfrak{H}_1, \mathfrak{H}_2, \ldots, \mathfrak{H}_\nu$ et l'on effectue une observation grâce à laquelle est connu celui d'entre eux $\mathfrak{H}^*$ qui contient H^*. Nous admettrons qu'il existe toujours au moins une succession d'observations menant à la détermination rigoureuse de H^*. Supposant maintenant que le coût de chaque expérience est fixé, on cherche à définir la succession de celles-ci [la *stratégie* au sens de von Neumann ([1])] qui conduit au résultat avec un coût total moyen minimum pour des valeurs données des probabilités *a priori* $\mathrm{Pr}(H_i)$ de chacune des hypothèses initiales.

Comme exemples d'un tel problème citons :

Un circuit électrique est en dérangement, comment procéder le plus économiquement à la série des vérifications partielles qui permettra d'en retrouver la cause ?

On cherche le seuil de sensibilité d'un sujet à une drogue. Quelle sera la combinaison la plus rapide de tests élémentaires indiquant chacun seulement si telle dose employée est supra ou infra-liminaire ?

Quelle structure donner à un système de clefs dichotomiques pour une flore, un ensemble nosologique, etc., tel que l'identification soit assurée après la recherche d'un nombre minimum de signes ?

Sans prétendre ici à plus qu'à attirer l'attention sur cette question, nous

---

(*) Séance du 27 novembre 1950.

([1]) von Neumann et O. Morgenstern, *Theory of games and economic behaviour*, 1946, p. 79.

énoncerons divers théorèmes relatifs au cas où le coût de toutes les observations est le même et où c'est donc leur nombre moyen qu'on cherche à minimiser. Ces théorèmes sont la généralisation très immédiate et l'application à ce domaine de résultats dus à C. E. Shannon [2] dans le problème du codage optimal d'un message aléatoire et reposent sur la théorie de l'information de N. Wiener [3].

Supposons que toutes les observations soient possibles; soit $\bar{l}$ le nombre moyen d'observations et I l'information *a priori* définie par

$$\mathrm{I} = \Sigma \mathrm{Pr}(\mathrm{H}_i)\,\mathrm{Log}_2[\mathrm{Pr}(\mathrm{H}_i)].$$

THÉORÈME 1. — *Il existe au moins une stratégie telle que*

$$\bar{l} \leq 1 - \frac{\mathrm{I}}{\mathrm{Log}_2 \nu}.$$

THÉORÈME 2. — *Pour toute stratégie*

$$\bar{l} \geq - \frac{\mathrm{I}}{\mathrm{Log}_2 \nu}$$

et l'égalité est atteinte quand et seulement quand les $\mathrm{Pr}(\mathrm{H}_i)$ *sont telles que chaque observation divise l'ensemble des hypothèses encore admissibles en* ν *sous-ensembles ayant même somme de probabilité.*

Dans le cas particulièrement important de $\nu = 2$ et de $\mathrm{Pr}(\mathrm{H}_i) = (1/n)$ pour toutes les $n\mathrm{H}_i$ on a :

THÉORÈME 3. — *Il existe au moins une stratégie telle que*

$$\bar{l} = [\mathrm{Log}_2 n] + 2\left(1 - \frac{2^{[\mathrm{Log}_2 n]}}{n}\right).$$

($[\mathrm{Log}_2 n]$ représente le plus grand entier contenu dans le logarithme de base 2 de n).

---

[2] *A mathematical theory of communication. The Bell system technical journal*, 27, 1948, p. 379-423.

[3] *Cybernetics*, 1948, p. 47.

(Extrait des *Comptes rendus des séances de l'Académie des Sciences*, t. 232, p. 206-207, séance du 15 janvier 1951.)

GAUTHIER-VILLARS, IMPRIMEUR-LIBRAIRE DES COMPTES RENDUS DES SÉANCES DE L'ACADÉMIE DES SCIENCES.
138775-51 Paris. — Quai des Grands-Augustins, 55.