# Pour un enseignement de la statistique dans le premier cycle

*par P.L. HENNEQUIN (Professeur à l'Université de Clermont)*

## *Introduction*

La calcul des probabilités a été introduit dans notre enseignement secondaire depuis une dizaine d'années mais il jouit encore chez nous d'un préjugé défavorable : figurant jusqu'à l'année dernière parmi les allégements du programme du baccalauréat il était le plus souvent laissé de côté, du moins dans les sections les plus "nobles" des classes terminales.

Cela est dû, en partie, à son objet même : l'étude du hasard, notion qui, chez l'adolescent et souvent chez l'adulte, est plus métaphysique que mathématique (cf. les locutions "par hasard" ou "au hasard") et dont l'introduction même s'entoure de motivations subjectives (importance de la place des horoscopes dans certaine presse). L'analyse puis la résolution de problèmes sont souvent gênées, voire bloquées, par des idées à priori sur une nature "équitable" du hasard.

De fait, pour un mathématicien, les théorèmes du calcul des probabilités, lois des grands nombres et théorèmes de limite centrée, nécessitent pour être énoncés, et à plus forte raison démontrés, des connaissances de topologie et d'analyse (suites et intégrales convergentes, espaces métriques, fonctions de variable complexe), qu'on n'enseigne à l'heure actuelle qu'en deuxième année après le baccalauréat.

Bien entendu on peut, même à l'école élémentaire, aborder le calcul des probabilités par une voie purement expérimentale en faisant effectuer aux enfants des suites assez longues de répétitions de la même expérience mais une deuxième difficulté surgit due à la nature même de la convergence stochastique :
Si $f_n$ (A) désigne la fréquence d'un événement A dans une suite de n coups indépendants et pour chacun desquels A a la même probabilité p, $f_n$ (A) converge vers p comme $\frac{1}{\sqrt{n}}$ vers 0. Cette convergence est beaucoup plus lente que celles familières à nos élèves, en général en $\frac{1}{p^n}$ (par exemple celle de la suite des valeurs approchées décimales de $\pi$ ou de $\sqrt{2}$ pour lesquelles p = 10).

Nous nous proposons de montrer ici qu'il est possible de présenter la structure la plus simple de la théorie de la mesure : celle d'espace probabilisé fini, qui repose sur des considérations purement algébriques, sans aucune référence au hasard. Notre présentation repose sur l'étude détaillée d'un *recensement*. Cette méthode d'investigation qui remonte à l'antiquité s'est beaucoup répandue de nos jours où sondages et enquêtes sont utilisés aussi bien à des fins économiques que politiques et sont abondamment commentés par le moindre quotidien régional. Notre collègue G.Th. Guilbaud ne nous suggérait-il pas aux journées de Caen d'apprendre à nos élèves à lire le journal ?

Bien entendu on peut imaginer des recensements de tailles bien différentes : du questionnaire élaboré par les élèves d'une classe puis rempli par eux (ou par ceux de la classe voisine) et enfin dépouillé et analysé, à ceux dont nos élèves trouvent l'analyse dans leurs manuels de géographie, on peut en imaginer d'autres qui concernent les élèves d'un établissement ou les habitants d'une petite ville.

Un tel exercice nous semble fécond car il fera réfléchir nos élèves à la difficulté du recueil de l'information et de son résumé. Il nous semble être à la portée des élèves de fin de premier cycle (ou de seconde) et nous le suggérons aux responsables de la prochaine réforme.

Cet article a été rédigé à la suite d'un exposé au colloque inter IREM de Lyon consacré aux probabilités en Mai 1972. Il a été élaboré au cours de fructueuses discussions avec J. Badrikian, R. Berthuet et P. Farjot.

Il s'adresse, bien entendu, à des mathématiciens et sa présentation à des élèves nécessiterait une longue élaboration. Nous serons reconnaissants à tous nos collègues qui nous feront part d'expériences dans ce domaine.

Cet article comporte deux parties. La première consacrée à l'analyse du questionnaire et du recensement permet d'illustrer les notions d'anneau booléen (§ I.7), de sous-anneaux (§ I.12), d'atome (§ I.9 et I.12), de probabilité (§ II.4).

La seconde présentera, sur le même exemple, les notions de variable aléatoire (§ IV.1), de moyenne (§ IV.2), de fonction de répartition et de médiane (§ IV.3), de probabilité conditionnelle (§ V.1) et d'indépendance (§ V.2).

## PREMIERE PARTIE

### I — *Analyse du questionnaire*

I.1 - Pour effectuer une enquête ou un recensement, il faut d'abord préparer un *questionnaire*, c'est-à-dire une liste L de questions $\{Q_1, Q_2, \ldots, Q_n\}$ que nous appellerons *questions primaires*.

Chaque question $Q_i$ comporte une réponse choisie parmi $r_i$ : $\rho_i^0, \rho_i^1, \ldots, \rho_i^{r_i-1}$ dont l'ensemble sera noté $E_i$. Il est commode de coder $\rho_i^j$ par le nombre j. Une *réponse au questionnaire* L est alors une *suite* $s = \{q_1, q_2, \ldots, q_n\}$ de n nombres $q_i$ égaux à 0, 1, ..., $r_i-1$.

L'ensemble S de ces suites ou réponses possibles, produit cartésien des $E_i$, a pour cardinal $\prod_{i=1}^{n} r_i = \sigma$ . $\sigma$ sera appelé *richesse* du questionnaire.

I.2 - Il est d'usage de représenter les éléments de S par les branches d'un arbre ainsi construit : de la racine $M_0$ partent $r_1$ arcs conduisant aux points $M_1^0, \ldots M_1^{r_1-1}$ correspondant aux réponses $\rho_1^0, \ldots, \rho_1^{r_1-1}$ à $Q_1$ ; de chaque $M_1^1$ partent de même $r_2$ arcs et ainsi de suite jusqu'aux $\sigma$ points $M_n^s$ avec $s \in S$. L'arbre comporte donc $r_1\ (r_2\ (\ldots (r_{n-1}\ (r_n+1) + 1) \ldots) + 1)$ arcs ; ce nombre est minimum si les $r_i$ sont numérotés dans l'ordre croissant.

On peut associer à un élément s de S soit le point $M_n^s$ soit le chemin qui y conduit à partir de $M_0$ .

Cette représentation fait intervenir *l'ordre* dans lequel les questions sont posées et se justifie surtout quand les questions sont effectivement posées les unes après les autres pour préciser les réponses obtenues.

Nous n'aborderons pas ici le cas de questionnaires où l'on choisit la i^ème^ question posée en fonction des réponses obtenues aux (i—1) précédentes.

I.3 - Aux personnes qui ont répondu au questionnaire L, on pourrait poser un certain nombre d'autres questions.

Par exemple :

Répondez-vous $q_i$ à la question $Q_i$ mais pas $q_j$ à la question $Q_j$ ? La réponse à une telle question ne dépend en fait que de la suite des réponses aux questions de L. A une telle question on ne peut répondre que par OUI ou NON, nous dirons que c'est une *alternative*.

L'alternative : "répondez-vous $q_i^j$ à la question $Q_i$ ? " sera notée $A_i^j$.
Bien entendu ce qui nous importe n'est pas l'énoncé, à l'aide du langage courant, de l'alternative mais la façon dont elle se définit à partir des $Q_i$ ; nous appellerons donc $\mathcal{A}$ l'ensemble de ces alternatives où nous identifions deux d'entre elles si elles ont toujours la même réponse.

I.4 - Une telle alternative $A_\alpha$ est en fait définie par une application $\varphi_\alpha$ de S dans $\{0, 1\}$. $A_\alpha$ est l'alternative qui recevrait la réponse $\varphi_\alpha(q_1, q_2, \dots q_n)$ si elle était posée et si les réponses aux questions primaires sont $q_1, q_2, \dots, q_n$.

Nous noterons $\Phi$ l'ensemble de ces applications de S dans $\{0, 1\}$ et $I_1$ la bijection de $\Phi$ dans $\mathcal{A}$ qui à $\varphi_\alpha$ associe $A_\alpha$.

On a : card $\mathcal{A}$ = card $\Phi$ = $2^{\text{card } S} = 2^{\rho}$

$\mathcal{A}$ contient en particulier deux alternatives triviales (qui ne comportent qu'une réponse possible) :

$A_\emptyset$, dont la réponse est toujours 0, définie par $\varphi_\emptyset \equiv 0$

$A_U$, dont la réponse est toujours 1, définie par $\varphi_U \equiv 1$

L'alternative $A_i^j$ déjà rencontrée en I.3 peut se définir par l'application

$$\varphi_i^j(q_1, q_2, \dots q_{i-1}, q_i, q_{i+1}, \dots q_n) = \begin{cases} 0 \text{ si } q_i \neq j \\ 1 \text{ si } q_i = j \end{cases} \quad 1 \leqslant j \leqslant r_i$$

I.5 - A titre d'exemple, considérons le questionnaire L suivant (n = 3) :

$Q_1$ avez-vous une automobile ?

$Q_2$ partez-vous en vacances ?

$Q_3$ habitez-vous une ville (de plus de 10.000 habitants) ?

Une réponse est un nombre de 3 chiffres binaires.

L'ensemble S des réponses possibles a $2^3 = 8$ éléments.

Considérons l'application $\varphi_\alpha$ de S dans $\{0, 1\}$ définie par la table suivante :

| s | 000 | 001 | 010 | 011 | 100 | 101 | 110 | 111 |
|---|---|---|---|---|---|---|---|---|
| $\varphi_\alpha(s)$ | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |

L'alternative $A_\alpha$ peut s'énoncer

"Avez-vous une automobile et habitez-vous une ville ? "

Il y a $2^8 = 256$ telles alternatives.

I.6 - Nous allons maintenant structurer $\mathcal{A}$ et pour cela structurer $\Phi$ et transporter la structure par $I_1$ ; pour structurer $\Phi$ ensemble des applications de S dans $\{0, 1\}$ nous allons structurer cet ensemble à deux éléments $\{0, 1\} = D_2$. Partons, par exemple, de l'ordre $\leqslant$ défini par

$a \leqslant b$ sauf si $a = 1$ et $b = 0$

(ordre induit par l'ordre naturel sur les entiers).

Tout couple $(a, b)$ de $D_2 \times D_2$ possède un plus petit majorant noté $a \vee b$ et un plus grand minorant noté $a \wedge b$ pour lesquels on a les tables :

| $\vee$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 1 | 1 |

| $\wedge$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 1 |

$\vee$ et $\wedge$ sont associatives, commutatives, distributives l'une par rapport à l'autre, mais ni $(D_2, \vee)$ ni $(D_2, \wedge)$ ne sont des groupes. Permuter 0 et 1 fait passer d'une table à l'autre. Notons $a \to \bar{a}$ cette opération de permutation. Pour munir $D_2$ d'une structure de groupe d'élément neutre 0 par l'opération + il faut choisir la table :

| + | 0 | 1 |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 1 | 0 |

On a : $a+b = (a \wedge \bar{b}) \vee (\bar{a} \wedge b)$

$= (a \vee b) \wedge (\overline{a \wedge b})$

et $\bar{a} = 1 + a$.

$(D_2, +, \wedge)$ est un corps commutatif $(Z/(2Z))$ ; de plus $\forall\ a \in D_2$, $a \wedge a = a$, ce qui implique la commutativité et $\forall\ a, a + a = 0$.

I.7 - Définissons maintenant la structure sur $\Phi$ :

A chaque opération T sur $D_2$ on peut associer l'opération T sur $\Phi$ par :

$$\forall\ s \in S, [\varphi_\alpha \mathrm{T} \varphi_\beta](s) = \varphi_\alpha(s) \mathrm{T} \varphi_\beta(s)$$

$(\Phi, +, \wedge)$ est maintenant un anneau (et non un corps si card $S > 1$) tel que

$\forall\ \varphi \in \Phi$, $\varphi \wedge \varphi = \varphi$ : on dit que $(\Phi, +, \wedge)$ est un *anneau booléen*.

$\varphi_0$ est l'élément neutre de + et V, $\varphi_U$ celui de $\wedge$ ; ($\Phi$, +, $\wedge$) est donc unitaire. V et $\wedge$ sont respectivement les opérations de plus petit majorant et de plus grand minorant pour l'ordre défini sur $\varphi$ par

$$\varphi_\alpha \leqslant \varphi_\beta \Longleftrightarrow \forall \ s \in S \quad \varphi_\alpha(s) \leqslant \varphi_\beta(s).$$

I.8 - Pour faire de $I_1$ un isomorphisme, il suffit de définir sur $\mathcal{A}$ l'implication, notée $<$ , par $A_\alpha < A_\beta \Longleftrightarrow \varphi_\alpha \leqslant \varphi_\beta$

(la réponse oui à $A_\alpha$ entraîne la réponse oui à $A_\beta$)

l'opération *ou* (inclusif) notée V : $A_\alpha \vee A_\beta$ définie par $\varphi_\alpha \vee \varphi_\beta$ (question dont la réponse est oui si la réponse à $A_\alpha$ ou à $A_\beta$ est oui)

l'opération *et* notée $\wedge$ : $A_\alpha \wedge A_\beta$ définie par $\varphi_\alpha \wedge \varphi_\beta$ (question dont la réponse est oui si les réponses à $A_\alpha$ et à $A_\beta$ sont oui),

l'opération *contraire*, le contraire $\overline{A_\alpha}$ de $A_\alpha$ étant défini par $\overline{\varphi_\alpha}$ ,

l'opération *ou exclusif* notée $\Delta$ , $A_\alpha \ \Delta \ A_\beta$ défini par $\varphi_\alpha \dot{+} \varphi_\beta$.

($\mathcal{A}$ , $\Delta$, $\wedge$) est alors un anneau booléen unitaire fini.

On peut montrer que tout anneau booléen fini est isomorphe à l'anneau ($\Phi$, +, $\wedge$) des applications d'un ensemble S dans $\{0, 1\}$ L'anneau ($\Phi$, +, $\wedge$) est lui-même isomorphe à l'anneau ($\mathcal{P}(S)$,$\Delta$,$\cap$) des parties de S muni de la différence symétrique et de l'intersection, par la bijection $I_2$ qui à toute partie $S_\alpha$ de S associe l'application $\varphi_\alpha$ définie par

$$\varphi_\alpha(s) = 1_{S_\alpha}(s) = \begin{cases} 1 \text{ si } s \in S_\alpha \\ 0 \ \text{ sinon} \end{cases} \quad \text{de sorte que } S_\alpha = \varphi_\alpha^{-1} \quad (1)$$

I.9 - Dans $\mathcal{P}$ (S), jouent un rôle particulier les singletons $\{s\}$ dont l'image par $I_2$ est l'application $\varphi_s$ définie par

$$\varphi_s(s') = \begin{cases} 1 \text{ si } s' = s \\ 0 \ \text{ sinon} \end{cases}$$

et dont l'image par $I_1 \circ I_2$ est l'alternative $A_s$ dite *élémentaire*. Nous noterons $\mathcal{A}_S$ l'ensemble des alternatives élémentaires.

On vérifie que

a) si $s \neq s'$ $\quad A_s \wedge A_{s'} = A_0$

b) que $\bigvee_{s \in S} A_s = A_U$

c) qu'une alternative A de $\mathcal{A}$ est *élémentaire* si et seulement si $A \neq A_o$ et $B < A \Longrightarrow B = A$ ou $B = A_o$

d) qu'à la représentation $S_\alpha = \bigcup_{s \in S_\alpha} \{s\}$ d'une partie de S'
comme réunion des singletons qu'elle contient correspond dans la représentation

$$\forall\ A_\alpha \in \mathcal{A}\ ,\ A_\alpha = \bigvee_{A_s < A_\alpha} A_s$$

e) que S étant le produit cartésien des $E_i$, si
$s = \{q_1, q_2, \dots q_n\}$ on peut écrire le singleton $\{s\}$ comme l'intersection

$$\{s\} = \bigcap_{i=1}^{n} F_i \quad \text{où} \quad F_i = \prod_{j=1}^{i-1} E_j \times \{q_i\} \times \prod_{j=i+1}^{n} E_j$$

et l'alternative élémentaire $A_s$ correspondante à l'aide des $A_i^j$ définie en I.3 comme :

$$A_s = \bigwedge_{i=1}^{n} A_i^{q_i}$$

alternative qui peut se formuler explicitement par

"Répondez-vous $q_i$ à chacune des questions $Q_i$ , $1 \leqslant i \leqslant n$ ? "

Ainsi, dans notre exemple, il y a 8 questions élémentaires telles que $A_{101}$

"Etes-vous un citadin automobiliste qui ne part pas en vacances ? "

I.10 - Les questions $Q_i$ de L, que nous appelons questions primaires, nous ont permis de définir des alternatives à partir d'une application de S dans $\{0,\ 1\}$. Plus généralement à tout $n \in N$ et à une application $\psi_\lambda$ de S dans $\{0,\ 1, \dots m-1\} = D_m$ on peut associer une question $Q_\lambda$ dite question dérivée dont la réponse est $\psi_\lambda(s)$ si la réponse à L est s.

Une telle application définit un *partage* $\omega_\lambda$ de S :
$S = \sum_{k=0}^{m-1} \psi_\lambda^{-1}(\{k\})$ c'est-à-dire une *suite* de parties disjointes $S_k = \psi_\lambda^{-1}(\{k\})$ de S, d'union S, et dont certaines peuvent être vides ; réciproquement un tel partage $\{S_k,\ 0 \leqslant k \leqslant m-1\}$ définit une application $\psi_\lambda$ par $\psi_\lambda(s) = k$ si et seulement si $s \in S_k$.

Nous noterons Q l'ensemble de toutes les questions dérivées de L. Bien entendu on peut plonger naturellement L dans Q en considérant chaque question $Q_i$ de L comme définie par l'application $\psi_i$ telle que

$$\psi_i (q_1 , q_2 , ..., q_{i-1} , q_i, q_{i+1} , ... q_n ) = q_i , 0 \leqslant q_i \leqslant r_i - 1 .$$

Comme m est quelconque, Q est infini, mais bien entendu le nombre d'éléments *non vides* du partage associé à un élément de Q est borné par $\sigma$ .

1.11 - Sur Q on peut définir un produit et une relation d'équivalence : $Q_\lambda$ et $Q_{\lambda'}$ étant deux éléments de Q définis par les applications $\psi_\lambda$ et $\psi_{\lambda'}$ de S respectivement dans $D_m$ et $D_{m'}$ , soit J l'application de $D_m \times D_{m'}$ dans $D_{mm'}$ définie par l'ordre lexicographique, alors $Q_\lambda \times Q_{\lambda'}$ est définie par $J (\psi_\lambda , \psi_{\lambda'})$.

La relation d'équivalence sera notée $\sim$ et ainsi définie : $Q_\lambda \sim Q_{\lambda'}$ s'il existe une bijection j d'une partie D de $D_m$ sur une partie D' de $D_{m'}$ telle que

$$\forall\ s \in S , \psi_\lambda (s) \in D \text{ et } \psi_{\lambda'} (s) = j \circ \psi_\lambda (s)$$

Autrement dit deux questions $Q_\lambda$ et $Q_{\lambda'}$ sont équivalentes si les deux partages $\omega_\lambda$ et $\omega_{\lambda'}$ définissent la même *partition* $\pi$ de S (rappelons qu'une *partition* est *l'ensemble* des éléments *non vides* d'un partage).

L'ensemble $\widehat{Q}$ des classes de Q pour $\sim$ , isomorphe à l'ensemble des partitions de S, est fini et son cardinal est une fonction B ( $\sigma$ ), dite fonction de Bell.

On vérifie que la classe de $Q_\lambda \times Q_{\lambda'}$ ne dépend que de $\widehat{Q}_\lambda$ et $\widehat{Q}_{\lambda'}$ et on peut donc définir le produit $\widehat{Q}_\lambda \times \widehat{Q}_{\lambda'}$ qui peut se définir aussi directement, comme associé à la partition produit des deux partitions $\pi_\lambda$ et $\pi_{\lambda'}$ , c'est-à-dire à la partition formée des intersections *non vides* d'un élément de $\pi_\lambda$ par un élément de $\pi_{\lambda'}$ .

Le questionnaire L est alors élément de la classe $\widehat{Q}_L$ associée à la partition de S en ses singletons, et égale au produit des $Q_i$ , $1 \leqslant i \leqslant n$.

1.12 - A une classe $\widehat{Q}_\lambda$ on peut associer l'ensemble des alternatives $A^j_\lambda$ définies par les éléments de la partition $\pi_\lambda$ et le sous-anneau $\mathcal{A}_\lambda$ de $\mathcal{A}$ engendré par les $A^j_\lambda$ . Ce sous-anneau est unitaire, c'est-à-dire qu'il contient $A_U$ et les $A^j_\lambda$ sont ses atomes, c'est-à-dire que $A^j_\lambda \neq A_o$ et que

$$B < A^j_\lambda , B \in \mathcal{A} \implies B = A^j_\lambda \text{ ou } B = A_o$$

Réciproquement, comme $\mathcal{A}$ est fini, on peut montrer que tout sous-anneau unitaire de $\mathcal{A}$ possède des atomes et qu'il est engendré par ces atomes (c'est-à-dire qu'il coïncide avec le plus petit sous-anneau qui les contient). Ces atomes définissent une partition de S donc une classe $\widehat{Q}_\lambda$ : on établit ainsi une bijection entre l'ensemble des sous-anneaux unitaires de $\mathcal{A}$ et $\widehat{Q}$.
Au produit $\widehat{Q}_\lambda \times \widehat{Q}_{\lambda'}$ de deux éléments de Q correspond par cette bijection le sous-anneau $\mathcal{A}_\lambda \vee \mathcal{A}_{\lambda'}$ engendré par les deux sous-anneaux $\mathcal{A}_\lambda$ et $\mathcal{A}_{\lambda'}$ (qui est plus grand que l'union de ces deux sous-anneaux).

I.12 - Il est alors intéressant de se poser la question suivante : $\sigma$ étant fixé, construire une famille de questions $\{Q'_i , 1 \leqslant i \leqslant k\}$ telle que $Q'_i$ ait $m_i$ réponses possibles,

que $Q'_1 \times Q'_2 \ldots \times Q'_k \in \widehat{Q}_L$ et que $\sum_{i=1}^{k} m_i$ soit *minimum*.

Pour qu'une telle famille existe, il est nécessaire et suffisant que $\prod_{i=1}^{k} m_i \geqslant \sigma$.

En effet, le produit des $Q'_i$ comporte au plus $\prod_{i=1}^{k} m_i$ réponses possibles et la condition est nécessaire ; si elle est satisfaite on peut établir une bijection entre S et une partie du produit cartésien des $D_{m_i}$ et il en résulte que la partition de S définie par le produit des $Q'_i$ est formée des singletons de S.

Pour minimiser $\sum_{i=1}^{k} m_i$ sous la condition $\prod_{i=1}^{k} m_i \geqslant \sigma$ il suffit de choisir

- si $3^q < \sigma \leqslant 3^{q-1} \times 4$, $m_i = 3$ pour $1 \leqslant i \leqslant q-1$ et $m_q = m_{q+1} = 2$ de sorte que $k = q+1$ et $\sum_{i=1}^{k} m_i = 3q+1$

- si $3^{q-1} \times 4 < \sigma \leqslant 3^q \times 2$, $m_i = 3$ pour $1 \leqslant i \leqslant q$ et $m_{q+1} = 2$ de sorte que $k = q+1$ et $\sum_{i=1}^{k} m_i = 3q+2$

- si $3^q \times 2 < \sigma \leqslant 3^{q+1}$, $m_i = 3$ pour $1 \leqslant i \leqslant q+1$ de sorte que $k = q+1$ et $\sum_{i=1}^{k} m_i = 3q+3$.

k est donc le plus petit entier supérieur à $\log_3 \sigma$ et on a intérêt à employer le plus grand nombre de questions ternaires (c'est ce que l'on fait quand on offre le choix entre les réponses "OUI", "NON" et "ABSTENTION").

Minimiser $\sum_{i=1}^{k} m_i$ est un critère raisonnable si le coût d'une question est proportionnel au nombre des réponses qu'elle comporte. On peut en choisir d'autres.

## II — *Analyse du recensement*

II.1 - Effectuer un recensement, c'est distribuer les questionnaires vierges, les faire remplir (ou les remplir d'après les indications du questionné) et recueillir les questionnaires remplis.

Notons $\omega$ un questionnaire rempli, $\Omega$ l'ensemble des $\omega$ ou population recensée (qui ne coïncide pas en général avec la population qu'on *se proposait* de recenser).

A chaque $\omega \in \Omega$ nous pouvons associer la suite $\{ q_i(\omega), 1 \leqslant i \leqslant n \}$ des n réponses aux questions $Q_i$. Nous définissons ainsi, sur $\Omega$, n applications $q_i$ respectivement dans $E_i$ ou encore une application R appelée *recensement*, de $\Omega$ dans S.

A priori l'ensemble $\Omega$ ne possède aucune structure et il est sans intérêt pour le statisticien d'analyser R elle-même (il n'en serait pas de même pour le percepteur ou pour le Ministère de l'Intérieur). Par contre l'ensemble $\mathcal{P}(\Omega)$ des parties de $\Omega$ possède une structure d'anneau booléen unitaire fini et il est naturel d'associer à R l'application $R^{-1}$ de $\mathcal{P}(S)$ dans $\mathcal{P}(\Omega)$ défini par

$$R^{-1}(S_\alpha) = \{ \omega ; R(\omega) \in S_\alpha \} = \Omega_\alpha$$

Rappelons que $\mathcal{P}(S)$ est lui-même isomorphe à $\Phi$ et à $\mathcal{A}$ par les applications $I_2$ (définies au § I.8) et $I_1$ (définies au § I.4) suivant la figure

$$\mathcal{P}(\Omega) \xleftarrow[R^{-1}]{} \mathcal{P}(S) \xrightarrow[I_2]{} \Phi \xrightarrow[I_1]{} \mathcal{A}$$

de sorte que si nous considérons au lieu de $S_\alpha$ l'alternative

$A_\alpha = I_1 \circ I_2 (S_\alpha)$ définie par l'application $\varphi_\alpha = I_2(S_\alpha) = 1_{S_\alpha}$, on a aussi

$$\Omega_\alpha = R^{-1}\left\{ \varphi_\alpha^{-1} ( \{1\} ) \right\} = (\varphi_\alpha \circ R)^{-1} \{1\} = R^{-1} \circ I_2^{-1} \circ I_1^{-1} (A_\alpha)$$

$\Omega_\alpha$ est l'ensemble des questionnés qui répondraient oui à l'alternative $A_\alpha$ si elle leur était posée ($\Omega_o = \phi$, $\Omega_U = \Omega$).

$R^{-1}$ est un homomorphisme d'anneaux de $\mathcal{P}(S)$ dans $\mathcal{P}(\Omega)$ et il en est de même de $R^{-1} \circ I_2^{-1} \circ I_1^{-1}$ de $\mathcal{A}$ dans $\mathcal{P}(\Omega)$ ; il revient au même de se donner R ou $R^{-1}$.

Soit $R(\Omega) = S_R$ le sous-ensemble de S des réponses *effectivement obtenues* dans le recensement.

Les $R^{-1}(\{s\}) = \Omega_s$, $s \in S_R$ forment une partition de $\Omega$ qui engendre le sous-anneau unitaire $\mathcal{B}_R$ de $\mathcal{P}(\Omega)$ image par $R^{-1}$ de $\mathcal{P}(S)$. $R^{-1}$ est un isomorphisme de $\mathcal{P}(S_R)$ sur $\mathcal{B}_R$ . $\mathcal{P}(S_R)$ est lui-même isomorphe à un sous-anneau $\mathcal{A}_R$ de $\mathcal{A}$ qui ne contient pas $A_U$ si $S_R$ est strictement contenu dans S.

II.2 - Comme $\Omega$ n'est pas structuré et que $\mathcal{P}(\Omega)$ est fini, il est naturel d'introduire sur $\mathcal{P}(\Omega)$ la fonction cardinal notée card qui possède les deux propriétés de *positivité* et *d'additivité* :

$$\text{card } \Omega_1 \geqslant 0 \text{ et, si } \Omega_1 \cap \Omega_2 = \phi, \ \text{card}(\Omega_1 \cup \Omega_2) = \text{card } \Omega_1 + \text{card } \Omega_2$$

Il en sera de même de toute fonction $\lambda$ card avec $\lambda \geqslant 0$. Pour pouvoir ultérieurement comparer deux recensements relatifs à $\Omega$ et $\Omega'$, on introduit la fonction $f_\Omega$ dite *fréquence* dans $\Omega$ définie par

$$f_\Omega(\Omega_1) = \frac{\text{card } \Omega_1}{\text{card } \Omega} \text{ et telle que } f_\Omega(\Omega) = 1.$$

En composant f avec $R^{-1} \circ I_2^{-1} \circ I_1^{-1}$ on obtient une application de $\mathcal{A}$ dans $[0, 1]$ que nous noterons $P_\Omega$ et que nous appellerons *probabilité par rapport à* $\Omega$ :

$$P_\Omega(A_\alpha) = f_\Omega(\Omega_\alpha) = \frac{\text{card } \Omega_\alpha}{\text{card } \Omega}.$$

$P_\Omega$ satisfait sur $\mathcal{A}$ les trois propriétés

$$\forall A \in \mathcal{A} \quad P_\Omega(A) \geqslant 0 \tag{1}$$

$$P_\Omega(A_U) = 1 \tag{2}$$

$$\text{Si } A \wedge B = A_0, P_\Omega(A \vee B) = P_\Omega(A) + P_\Omega(B) \tag{3}$$

On déduit de 3 et de la relation I.8 d , $A = \bigvee_{A_s < A} A_s$, que

$$P_\Omega(A) = \sum_{A_s < A} P_\Omega(A_s) \tag{4}$$

$P_\Omega$ est donc entièrement définie par sa restriction à $\mathcal{A}_S$ ; celle-ci satisfait

$$\forall A_s \in \mathcal{A}_S, P_\Omega(A_s) \geqslant 0 \text{ et } \sum_{s \in S} P_\Omega(A_s) = 1 \tag{5}$$

Réciproquement toute fonction P sur $\mathcal{A}_S$ satisfaisant (4) peut, de façon unique, se prolonger par (4) en une fonction sur $\mathcal{A}$ qui satisfait (1), (2), (3).

Si de plus elle est à valeurs rationnelles, on peut lui associer un ensemble $\Omega$ et une partition $\Omega = \sum_{s \in S} \Omega_s$ telle que

$$\frac{\text{card } \Omega_s}{\text{card } \Omega} = P(A_s),$$

puis l'application R de $\Omega$ dans S définie par $R(\omega) = s$ si $\omega \in \Omega_s$ ; alors le prolongement de P sur $\mathcal{A}$ coïncide avec $P_\Omega$.

II.3 - Soit $N_\Omega = S \setminus S_R$ le sous-ensemble de S des réponses qui n'ont pas été obtenues dans le recensement de $\Omega$. Posons $A_N = \bigvee_{s \in N} A_s = I_1 \circ I_2 (N_\Omega)$.

On a $P_\Omega(A_N) = 0$, et $P_\Omega(A_\alpha) = 0$ si et seulement si $A_\alpha < A_N$. Les alternatives $A_\alpha$ qui impliquent $A_N$ seront dites $P_\Omega$-négligeables ou $P_\Omega$ presque impossibles.

Toute alternative de $\mathcal{A}$ diffère d'une alternative de $\mathcal{A}_R$ par une alternative $P_\Omega$-négligeable. Pour la population $\Omega$ considérée seules ont un intérêt les alternatives de $\mathcal{A}_R$. Un questionnaire sera adapté à $\Omega$ si $A_N = A_0$ et $\mathcal{A}_R = \mathcal{A}$.

II.4 - Pour déterminer $P_\Omega$ il suffit de connaître $S_R$ et la restriction de $P_\Omega$ à $\mathcal{A}_R \cap \mathcal{A}_S$ ensemble des alternatives élémentaires auxquelles il a été répondu oui par au moins un questionné.

On a $\text{card}(\mathcal{A}_R \cap \mathcal{A}_S) = \text{card } S_R = \sigma - \text{card } N_\Omega$

et $\text{card}(\mathcal{A}_R \cap \mathcal{A}_S) \leqslant \text{card } \Omega$ d'où la majoration

$\text{card}(\mathcal{A}_R \cap \mathcal{A}_S) \leqslant \inf(\text{card } S, \text{card } \Omega)$

II.5 - Nous avons vu aux § I.10 et I.11 comment associer à une partition $\pi_\lambda$ de S une classe $\widehat{Q}_\lambda$ de questions. Nous appellerons distribution de probabilité associée à $\widehat{Q}_\lambda$ par $P_\Omega$ l'application $A_\lambda^j \to P_\Omega(A_\lambda^j)$ où les $A_\lambda^j$ sont les alternatives définies par les éléments de la partition $\pi_\lambda$.

On a $P_\Omega(A_\lambda^j) \geqslant 0$ et $\sum_j P_\Omega(A_\lambda^j) = 1$ ; étant donnée une distribution elle se prolonge de façon unique en une probabilité sur $\mathcal{A}_\lambda$.

On se contente souvent de publier comme résultat du recensement la restriction de $P_\Omega$ aux alternatives primaires $A_i^j$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant r_i$ c'est-à-dire les n distributions de probabilité associées aux $\widehat{Q}_i$. Cela permet seulement de calculer $P_\Omega$ sur la famille des unions d'éléments de $\mathcal{A}_i$, $1 \leqslant i \leqslant n$, qui est plus petite que $\mathcal{A}$ dès que $n > 1$.