JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

PAUL VINCENT

Application des ensembles électroniques à la recherche démographique

Journal de la société statistique de Paris, tome 105 (1964), p. 135-164

<http://www.numdam.org/item?id=JSFS_1964__105__135_0>

III

APPLICATION DES ENSEMBLES ÉLECTRONIQUES A LA RECHERCHE DÉMOGRAPHIQUE

Bien que les calculateurs électroniques soient en principe des calculateurs « universels », il s'en faut de beaucoup que tous les matériels soient également adaptés aux diverses tâches qu'on peut songer à leur confier. En fait, chaque matériel résulte d'un compromis entre différentes exigences plus ou moins contradictoires, et ses caractéristiques dépendent, dans une large mesure, de l'utilisation principale à laquelle le constructeur l'a destiné. C'est pourquoi on distingue souvent les *ensembles de gestion* des *ensembles scientifiques;* et, bien que certains matériels aient une vocation mixte, cette distinction résulte d'une classification assez valable des principales applications du calcul électronique.

La « gestion » comprend essentiellement des tâches administratives, telles que paie, tenue de comptes, contrôle de stocks, facturation — y compris, naturellement, les analyses statistiques correspondantes. Toutes ces tâches ont en commun certains caractères : elles se renouvellent en général périodiquement; elles intéressent un grand nombre de postes, qui peuvent être ordonnés d'après un numéro d'identification; elles conduisent à la production de nombreux documents; par contre, les calculs proprement dits qu'elles impliquent, sont généralement très simples et d'un volume relativement faible. — Les dépouillements d'enquêtes s'apparentent à ce type d'applications, parce qu'ils portent sur un grand nombre d'unités statistiques, conduisent à la production d'un volume d'états assez important, et comportent plutôt des classements et des comptages que des calculs.

Les ensembles de gestion seront donc pourvus de puissants moyens d'entrée-sortie, et de mémoires auxiliaires de grande capacité. Par contre, ils pourront se contenter d'une mémoire centrale rapide de faible capacité, et d'organes de traitement ne comportant qu'un petit nombre d'opérations câblées. La représentation interne des caractères s'effectuera de préférence par positions alphanumériques, groupées en « mots » de longueur variable.

Les travaux « scientifiques » ont rarement le même caractère répétitif que les travaux de gestion. Ceci est particulièrement vrai en matière de recherche, où les problèmes ne se présentent généralement qu'une seule fois. Il est alors impossible d'amortir, sur une succession de travaux analogues, de lourdes programmations en langage-machine; d'où la nécessité de pouvoir utiliser quelque « langage électronique » pour la programmation. D'autre part, dans les problèmes scientifiques, les données sont généralement peu nombreuses, de même que les résultats; alors qu'au contraire le volume des calculs est toujours considérable — sans quoi on n'aurait pas eu recours au calcul électronique.

Les ensembles scientifiques pourront donc ne comporter que des moyens d'entrée-sortie relativement faibles. Par contre, ils devront disposer d'une mémoire centrale à accès rapide de grande capacité. Leurs organes de traitement seront pourvus d'un grand nombre d'opérations, de préférence câblées, à tout le moins simulées. La représentation interne des nombres s'effectuera le plus souvent en binaire pur, sous la forme de « mots » de longueur fixe.

Or la vitesse coûte cher, tant pour l'accès en mémoire que pour les calculs. C'est dire que l'unité centrale d'un ensemble scientifique sera généralement d'un prix élevé. Cependant, la vitesse est d'ordinaire rentable, en ce sens qu'un matériel qui met dix fois moins de temps qu'un autre pour traiter le même problème, ne coûte pas dix fois plus cher. D'où l'apparition

d'ensembles de plus en plus puissants, qui ont en outre l'avantage de pouvoir traiter des problèmes dépassant les possibilités de matériels plus modestes. Mais la question se pose alors d'alimenter en travail des ensembles aussi puissants; car si on les laisse tant soit peu inoccupés, leur utilisation cesse d'être rentable. On ne les trouvera donc que dans de grands bureaux de calcul, qui acceptent d'ailleurs souvent d'apporter leur concours aux tiers sous forme de travail à façon.

*
* *

Compte tenu de cette situation, voyons maintenant quel bénéfice la recherche démographique peut espérer tirer du calcul électronique. Il semble qu'on puisse, à cet égard, envisager trois catégories d'applications :

1° le dépouillement d'enquêtes:

2° le calcul numérique;

3° la simulation.

Nous ne nous attarderons pas sur la première catégorie, qui n'a rien de spécifiquement démographique, et dont certains de nos collègues, de l'I. N. S. E. E. notamment, pourraient parler d'expérience, ce qui n'est pas notre cas. Nous nous contenterons de remarquer que cette catégorie d'applications se distingue des deux autres, quant au type de matériel qui lui convient le mieux.

Nous ne nous attarderons pas davantage sur la deuxième catégorie. En effet, le chercheur démographe peut bien se trouver en présence d'un problème de pur calcul. Mais alors, de deux choses l'une : ou bien ce problème aura déjà été traité électroniquement, et il figurera probablement au catalogue de quelque « bibliothèque de programmes »; ou bien il se présentera comme un problème entièrement nouveau, auquel cas notre démographe aura sans doute avantage à s'adresser à un spécialiste du calcul numérique sur machines électroniques. Car il s'agit d'une discipline particulière, le calcul électronique ayant ses méthodes propres, où interviennent des questions fort délicates, telles que la précision obtenue compte tenu de la technologie interne du matériel employé. — De toutes façons nous devrions, cette fois encore, exciper de notre inexpérience.

Reste donc la troisième catégorie, qui retiendra spécialement notre attention. Non que la simulation soit une technique propre à la démographie, ni qu'elle y revête des caractères particuliers. Mais elle offre un champ d'application particulièrement propice au calcul électronique, et il semble bien que ce soit le domaine où les ensembles électroniques puissent rendre le plus de services à la recherche démographique.

On ne saurait évidemment décrire un domaine qui peut se développer sans limite et dans les directions les plus variées, au gré de l'imagination des chercheurs. Contentons-nous donc de remarquer, à ce propos, que les perspectives démographiques relèvent de la simulation, au même titre que les modèles auxquels les chercheurs paraissent de plus en plus enclins à recourir.

Cependant, du point de vue qui nous occupe, il y aura généralement une différence assez sensible entre une perspective et ce qu'on appelle d'ordinaire un modèle. C'est que, le plus souvent, les calculs perspectifs seront relativement simples et fourniront des résultats assez volumineux, tandis que les modèles requerront des calculs beaucoup plus complexes pour aboutir à des résultats d'une grande concision. On peut donc s'attendre à ce que le matériel le mieux adapté aux calculs de modèles ne soit pas celui qui convienne le mieux aux calculs perspectifs.

Mais, qu'il s'agisse d'un calcul de modèle ou d'un calcul perspectif, un problème de simulation ne sera justiciable d'un traitement électronique, que s'il remplit simultanément les trois conditions suivantes :

1° impliquer un travail considérable par d'autres méthodes de traitement;

2° se prêter à un traitement, sinon entièrement, du moins très largement automatique;

3° permettre l'insertion de processus répétitifs dans le traitement.

C'est ce que nous nous proposons d'illustrer par un exemple, qui nous fournira en même temps quelques bases concrètes de référence pour préciser la signification et la portée des conditions que nous venons d'énoncer.

*
* *

Le problème qui nous servira d'exemple, nous paraît assez caractéristique des problèmes de simulation qui peuvent se poser en recherche démographique. En outre, sa solution complète semblait à priori devoir requérir un volume de calculs si important, que nous avions hésité à en aborder la solution avec nos moyens habituels. Aussi ce problème s'est-il, en quelque sorte, imposé à nous, lorsque nous avons voulu procéder à une expérience de calcul électronique. Avant d'indiquer en quoi a consisté cette expérience, voyons donc comment se présentait notre problème.

Dans nos *Recherches sur la fécondité biologique* (1), nous avions indiqué qu'on pourrait essayer de déduire, de la répartition des premières conceptions par mois de durée de mariage, quelques indications sur la distribution de la fécondabilité (2) dans le groupe observé. En effet, imaginons que nous puissions dénombrer les conceptions de premier rang qui surviennent, au cours de chaque mois de durée de mariage, au sein d'un ensemble de couples non contracepteurs tous fertiles et n'ayant pas conçu avant le mariage. Convenons d'appeler « groupe restant en observation » à une certaine durée de mariage, le groupe des femmes faisant partie de l'ensemble observé et n'ayant pas encore conçu à ce moment. Si l'ensemble initial est hétérogène par rapport à la fécondabilité, le groupe restant en observation s'appauvrira plus vite de ses éléments à forte fécondabilité que de ses éléments à faible fécondabilité; la fécondabilité de ce groupe (3) apparaîtra donc comme une fonction décroissante de la durée du mariage — ce qui correspond effectivement à ce qu'on observe.

Or, plaçons-nous dans l'hypothèse où l'hétérogénéité du groupe initial par rapport à la fécondabilité serait le seul facteur responsable de cette décroissance de la fécondabilité du groupe restant en observation, la fécondabilité de chaque couple demeurant constante depuis le mariage jusqu'à la première conception. La répartition par durée de mariage des conceptions de premier rang survenant dans le groupe, pourrait alors être calculée — aux fluctuations aléatoires près — si l'on connaissait la distribution de la fécondabilité dans le groupe initial.

Naturellement, cette distribution nous est inconnue, et c'est le problème inverse que nous voudrions bien pouvoir résoudre, à savoir : déterminer la distribution initiale de la fécondabilité à partir de la répartition observée des premières conceptions par durée de

(1) P. U. F., 1961, p. 215, note 3.

(2) Nous appelons *fécondabilité*, la probabilité de conception par cycle menstruel. Nous considérons cette probabilité comme une caractéristique du couple; mais, pour simplifier la terminologie, nous l'attachons à la femme.

(3) Nous définissons la *fécondabilité d'un groupe*, comme la moyenne arithmétique des fécondabilités individuelles des femmes qui le constituent.

mariage. Or les fluctuations aléatoires auxquelles nous avons fait allusion, paraissent nous interdire tout espoir d'apporter une solution incontestable à un tel problème.

Par contre, ce que nous pouvons très bien faire, c'est nous donner des distributions hypothétiques de la fécondabilité dans le groupe initial, en déduire les répartitions correspondantes des conceptions de premier rang selon la durée du mariage, et confronter ces répartitions « calculées » avec la répartition « observée ». Si nous arrivons alors à trouver des répartitions calculées assez voisines de la répartition observée, nous pourrons au moins dire : Voici un certain nombre de distributions de la fécondabilité, qui sont compatibles à la fois avec les données d'observation et avec l'hypothèse énoncée plus haut.

Si, au contraire, nous n'arrivons pas à constituer des modèles reproduisant convenablement les phénomènes observés, nous pourrons peut-être déduire des écarts apparus, quelques présomptions concernant la non-validité de l'hypothèse énoncée. En d'autres termes, nous aurons peut-être alors des raisons de soupçonner l'existence, à côté de l'hétérogénéité mise en cause, d'un autre facteur partiellement responsable de la décroissance de la fécondabilité du groupe restant en observation — une décroissance de la fécondabilité après l'instauration de rapports sexuels réguliers, par exemple (1).

Notre problème consistait donc à « essayer » des distributions hypothétiques de la fécondabilité dans le groupe initial.

* * *

Notre propos ne saurait s'étendre, aujourd'hui, à la façon dont nous sommes parvenu à évaluer la répartition des conceptions de premier rang par durée de mariage au sein de l'ensemble étudié. Nous considérerons donc comme des « données d'observation » les quantités suivantes :

E_n, effectif restant en observation au début du mois de mariage de rang n, pour $0 \leqslant n \leqslant 15$.

D'où se déduisent immédiatement les conceptions « observées » :

$G_n = E_n - E_{n+1}$, ou nombre de grossesses commencées pendant le mois de mariage de rang n $(0 \leqslant n \leqslant 14)$;

ainsi que les quotients mensuels de conception correspondants :

$f_n = \frac{G_n}{E_n}$, qui mesurent la fécondabilité du groupe restant en observation au début du mois de mariage de rang n $(0 \leqslant n \leqslant 14)$.

Remarquons d'autre part que, les couples observés étant tous fertiles, E_{15} représente les premières conceptions survenues à 15 mois de mariage au moins, de sorte que les 15 quantités G_n et la quantité E_{15}, nous donnent la répartition des E_0 femmes observées, en 16 classes de durée de mariage à la première conception.

Avant de procéder au choix des distributions à essayer, voyons comment on pourrait concevoir le calcul avec une distribution arbitraire. Le plus simple serait sans doute de considérer une distribution discrète; car, en cas de distribution continue, on obtiendrait probablement une précision satisfaisante, en découpant d'abord la distribution en trapèzes curvilignes assimilables à des rectangles de bases centrées sur les points d'abscisses $\frac{i}{100}$

(1) Nous faisons allusion à cette hypothèse, parce qu'elle a été soutenue avec conviction, et ne saurait, par conséquent, être rejetée à priori.

$(0 < i \leqslant 100)$, puis en supposant la masse de chacun de ces rectangles concentrée sur son axe (*cf.* le cartouche de la figure 1).

Soit donc y_i, la proportion des femmes de fécondabilité

$$x_i = \frac{i}{100} \qquad (0 < i \leqslant 100) \tag{1}$$

d'après la distribution théorique envisagée. Les quantités y_i sont telles que

$$\sum_{i=1}^{100} y_i = 1 \tag{2}$$

Pour obtenir la répartition « calculée » correspondant à la dimension de l'échantillon, il nous suffit donc de considérer les quantités

$$e_{0,i} = E_0\, y_i \tag{3}$$

obtenant ainsi une répartition de l'effectif initial E_0 en 100 classes homogènes par rapport à la fécondabilité.

Considérons l'une de ces classes. Son effectif initial est $e_{0,i}$ et sa fécondabilité x_i. Le nombre des conceptions survenant dans la classe au cours du premier mois de mariage, est

$$g_{0,i} = e_{0,i}\, x_i \tag{4}$$

et l'effectif des femmes de la classe restant en observation à la durée de mariage de 1 mois est

$$e_{1,i} = e_{0,i} - e_{0,i}\, x_i = e_{0,i}\,(1 - x_i) \tag{5}$$

On voit que, plus généralement, l'effectif des femmes de la classe restant en observation à la durée de mariage de n mois, est égal à

$$e_{n,i} = e_{n-1,i}\,(1 - x_i) = e_{0,i}\,(1 - x_i)^n \tag{6}$$

et que le nombre des conceptions survenant dans la classe au cours du mois de mariage de rang n, est donné par

$$g_{n,i} = e_{n,i}\, x_i = g_{0,i}\,(1 - x_i)^n \tag{7}$$

Considérant alors l'ensemble des classes, on obtiendra l'effectif « calculé » des conceptions survenues au cours du mois de rang n, par la sommation

$$G'_n = \sum_{i=1}^{100} g'_{n,i} \tag{8}$$

tandis que l'effectif restant en observation à $(n + 1)$ mois de durée de mariage sera calculé par récurrence (1) :

$$E'_{n+1} = E'_n - G'_n \tag{9}$$

à partir de

$$E'_0 = E_0 \tag{10}$$

Pour comparer ces résultats de calcul aux données d'observation, on pourra alors, par exemple, calculer la quantité :

$$\chi^2 = \sum_{n=0}^{14} \frac{(G'_n - G_n)^2}{G'_n} + \frac{(E'_{15} - E_{15})^2}{E'_{15}} \tag{11}$$

(1) On remarquera le parallèlisme des notations entre les quantités « observées » (E_n, G_n, f_n) et les quantités « calculées » (E'_n, G'_n, f'_n).

En cherchant dans une table de χ^2, la probabilité correspondant à la valeur ainsi obtenue pour 15 degrés de liberté, on pourra alors apprécier si la distribution considérée est compatible avec les observations.

***

Plutôt que d'aller à tâtons en essayant des distributions arbitraires, il nous a paru préférable de commencer par faire quelques tentatives avec des distributions correspondant à des formes analytiquement définies; et, en l'absence de tout renseignement sur la distribution de la fécondabilité x, nous avions évidemment intérêt à essayer un type de distribution assez général pour se prêter à la description de situations très différentes à l'intérieur de l'intervalle de variation (0 ; 1) de la variable x. C'est pourquoi notre choix s'est porté sur la distribution du type I de PEARSON correspondant à cet intervalle, distribution connue aussi sous le nom de *distribution bêta.*

Sous sa forme la plus générale, cette distribution peut s'écrire :

$$\beta\,(x;\,a,\,b) = cx^a\,(1 - x)^b \tag{12}$$

où a et b sont des constantes supérieures à — 1, la quantité c étant elle-même une constante telle que :

$$\int_0^1 \beta\,(x;\,a,\,b)\,dx = 1 \tag{13}$$

Le problème semble alors consister à chercher des couples de valeurs des paramètres a et b, tels que les distributions β correspondantes reproduisent convenablement les données d'observation. Mais en fait notre problème est à la fois plus compliqué et plus simple qu'il n'y paraît d'après cet énoncé.

La complication résulte de ce que, pour nous rapprocher des caractéristiques vraisemblables du groupe observé, nous devons essayer des distributions tronquées à gauche à partir d'une limite inférieure inconnue l de la fécondabilité. Ceci introduit un paramètre supplémentaire, puisque nous aurons à essayer des distributions telles que :

$$\begin{cases} y = 0, \text{ pour } x < l \\ y = c\,x^a\,(1 - x)^b, \text{ pour } x \geqslant l \end{cases} \tag{14}$$

où la constante c se déduit de la condition :

$$\int_l^1 y\,dx = 1 \tag{15}$$

Signalons toutefois que, si la valeur de l nous était inconnue, nous avions quelque idée de son ordre de grandeur : nous pensions qu'elle était probablement inférieure à 0,08 et vraisemblablement de l'ordre de 0,06. Comme nous envisagions — ainsi qu'on l'a vu — de faire varier x par « pas » de 0,01, nous estimions que l'introduction de ce nouveau paramètre l ne serait pas extrêmement gênante : en tout état de cause, nous n'aurions qu'un nombre très limité de valeurs de l à essayer.

D'autre part, la simplification annoncée compensait largement cette complication. En effet, nous disposions d'une estimation f_0 de la moyenne de x, ce qui nous permettait, une fois l fixé, de considérer la fonction y définie par la relation 14, comme ne dépendant que

d'un seul paramètre, a par exemple. En effet, l'autre paramètre b apparaît alors comme une fonction de a, soit b(a), définie implicitement par la condition que la moyenne de x soit toujours égale à f_0, quel que soit a.

Bien plus, si, comme nous nous y attendions, les distributions y les plus satisfaisantes admettent un mode assez voisin de la moyenne f_0 — laquelle ressort en l'espèce à 0,252 —, la troncature des distributions β correspondantes ne devrait pas altérer beaucoup leur moyenne. Or la moyenne m d'une distribution β est donnée par la relation simple :

$$m = \frac{a + 1}{a + b + 2} \tag{16}$$

En faisant $m = f_0$ dans cette formule, on peut donc en déduire une valeur approchée de b(a), à savoir :

$$b = \frac{a + 1}{f_0} - a - 2 \tag{17}$$

Mais nous savons en outre que cette valeur de b(a) est approchée par défaut. En effet, elle correspond à une distribution β complète de moyenne f_0; de sorte que la moyenne f'_0 de la distribution y correspondante, est certainement supérieure à f_0 (la distribution y résultant d'une distribution β amputée de sa partie gauche). Or, pour diminuer la moyenne d'une distribution y lorsque l et a sont fixés, il faut augmenter b. Il en résulte qu'on peut envisager de déterminer b(a) par approximations successives, en partant d'une valeur de b supérieure à celle donnée par la formule 17, et en affectant chaque valeur de b non admissible, d'une correction de même signe que la différence

$$d = f'_0 - f_0 \tag{18}$$

correspondante.

Remarquons d'autre part que cette détermination de b(a) repose uniquement sur le calcul de la moyenne f'_0 de la distribution y. Or cette moyenne s'identifie à celle de la répartition suivante :

$$\begin{cases} Y = 0, \text{ pour } x < l \\ Y = x^a (1 - x)^b, \text{ pour } x \geqslant l \end{cases} \tag{19}$$

On peut donc parfaitement, pour calculer b(a), substituer la « répartition » de la formule 19, à la « distribution » de la formule 14, ce qui nous épargnera la peine de calculer la constante c pour chaque valeur de b essayée.

En conséquence, nous concevrons l'essai d'une certaine valeur de b, de la façon suivante. Nous procéderons au calcul des deux intégrales :

$$S = \int_l^1 Y \, dx \tag{20}$$

et

$$\Gamma_0 = \int_l^1 Y x \, dx \tag{21}$$

et nous obtiendrons f'_0 par la simple division :

$$f'_0 = \frac{\Gamma_0}{S} \tag{22}$$

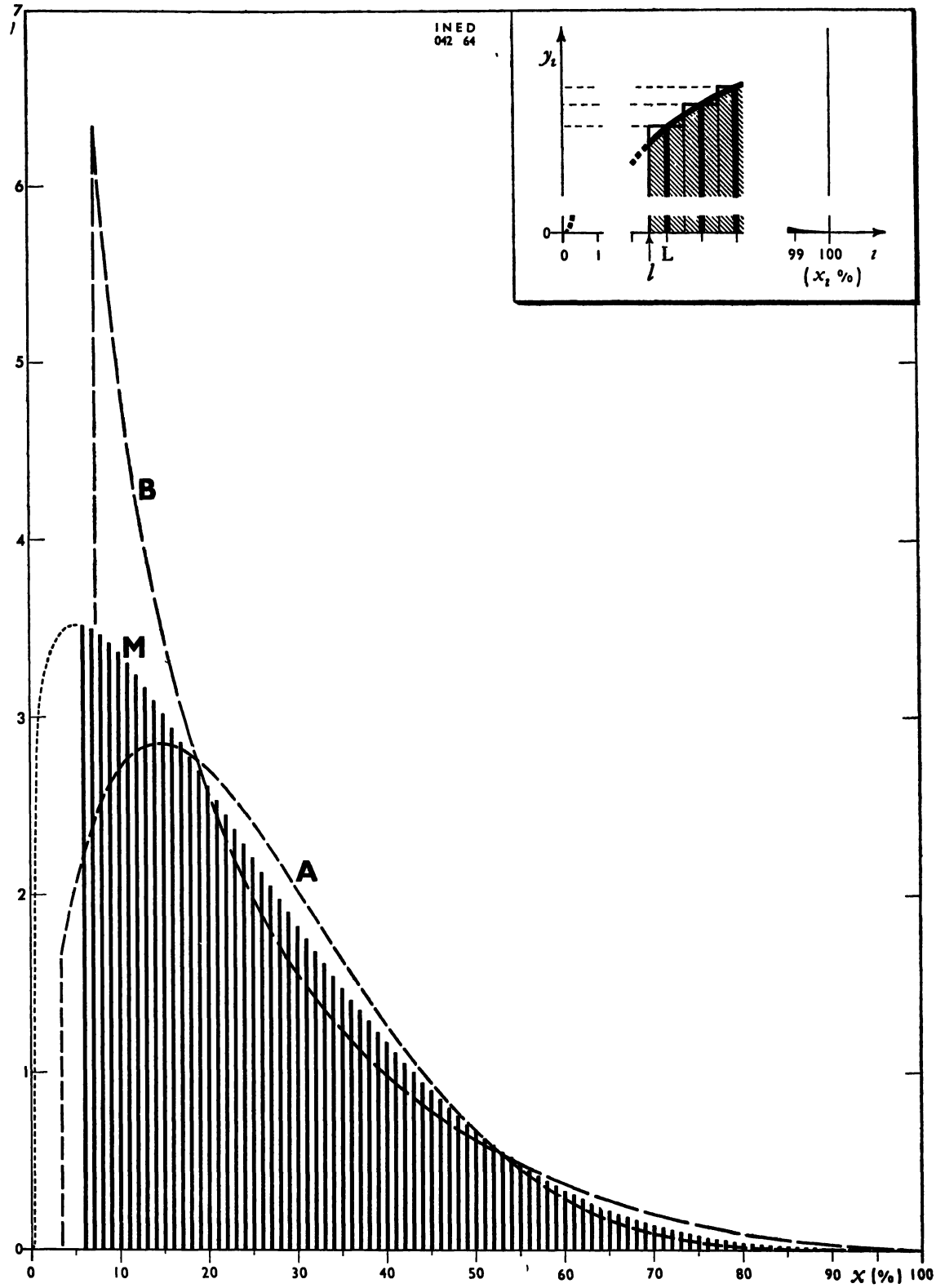


FIGURE 1. — Exemples de distributions de la fécondabilité, « possibles » pour l'échantillon étudié.

Pratiquement, nous opérerons par sommation, et non par intégration, en substituant à la répartition continue envisagée jusqu'ici, la répartition discrète suivante :

$$Y_i = x_i^a (1 - x_i)^b, \text{ avec } x_i = \frac{i}{100}, \text{ pour } i \geqslant L \tag{23}$$

où L désigne le plus petit entier qui soit supérieur ou égal à 100 l. On voit, dans le cartouche de la figure 1, qu'à un entier donné pris comme valeur de L, correspond alors en fait une limite l telle que :

$$l = \frac{L - 0{,}5}{100} \tag{24}$$

Remarquant alors que, lorsque b est positif, $Y_i = 0$ pour $i = 100$, nous écrirons les sommations correspondant aux intégrations 20 et 21, respectivement sous la forme :

$$S = \sum_{i=L}^{99} Y_i \qquad (b > 0) \tag{25}$$

$$\Gamma_0 = \sum_{i=L}^{99} Y_i x_i \qquad (b > 0) \tag{26}$$

si bien que, lorsque nous aurons trouvé la valeur b(a) telle que f'_0 résultant de la formule 22 soit égal à f_0, nous n'aurons plus qu'à faire

$$e_{0,i} = \frac{E_0}{S} Y_i \tag{27}$$

ou bien

$$y_i = \frac{1}{S} Y_i \tag{28}$$

pour obtenir à volonté, soit la répartition « calculée » correspondant à la dimension de l'échantillon, soit la distribution essayée elle-même (*cf.* le diagramme en bâtons de la figure 1, l'échelle des ordonnées indiquant alors le pourcentage des femmes comprises dans chaque classe).

* * *

L'organigramme de la figure 2 schématise le processus opératoire que nous venons d'indiquer, sous une forme assez voisine de celle utilisée en programmation symbolique. La lecture du schéma ne présente aucune difficulté (1), pourvu qu'on donne au signe « égal » (=) le sens qu'il a dans les langages du type FORTRAN. Cette signification est la suivante : considérer la valeur du symbole ou de l'expression figurant à droite du signe « égal », et attribuer désormais cette valeur au symbole écrit à gauche du dit signe. C'est ainsi qu'une formule telle que celle (n = n + 1) qui figure à la ligne 15 de l'organigramme, veut dire : calculer n + 1 (expression figurant à droite du signe « égal ») et attribuer la valeur ainsi obtenue à n (symbole écrit à gauche du signe « égal ») — en d'autres termes : augmenter la valeur de n d'une unité.

(1) Nous avons inscrit dans des hexagones les renseignements relatifs aux « transferts conditionnels » prévus pour le traitement du problème. La condition examinée est notée dans la partie supérieure de l'hexagone; lorsqu'il s'agit d'une comparaison entre deux quantités, celles-ci sont indiquées de part et d'autre d'un signe formé de quatre points disposés en carré (: :). Dans la partie inférieure de l'hexagone sont représentées les diverses possibilités considérées; les aiguillages qui en découlent sont matérialisés par des flèches.

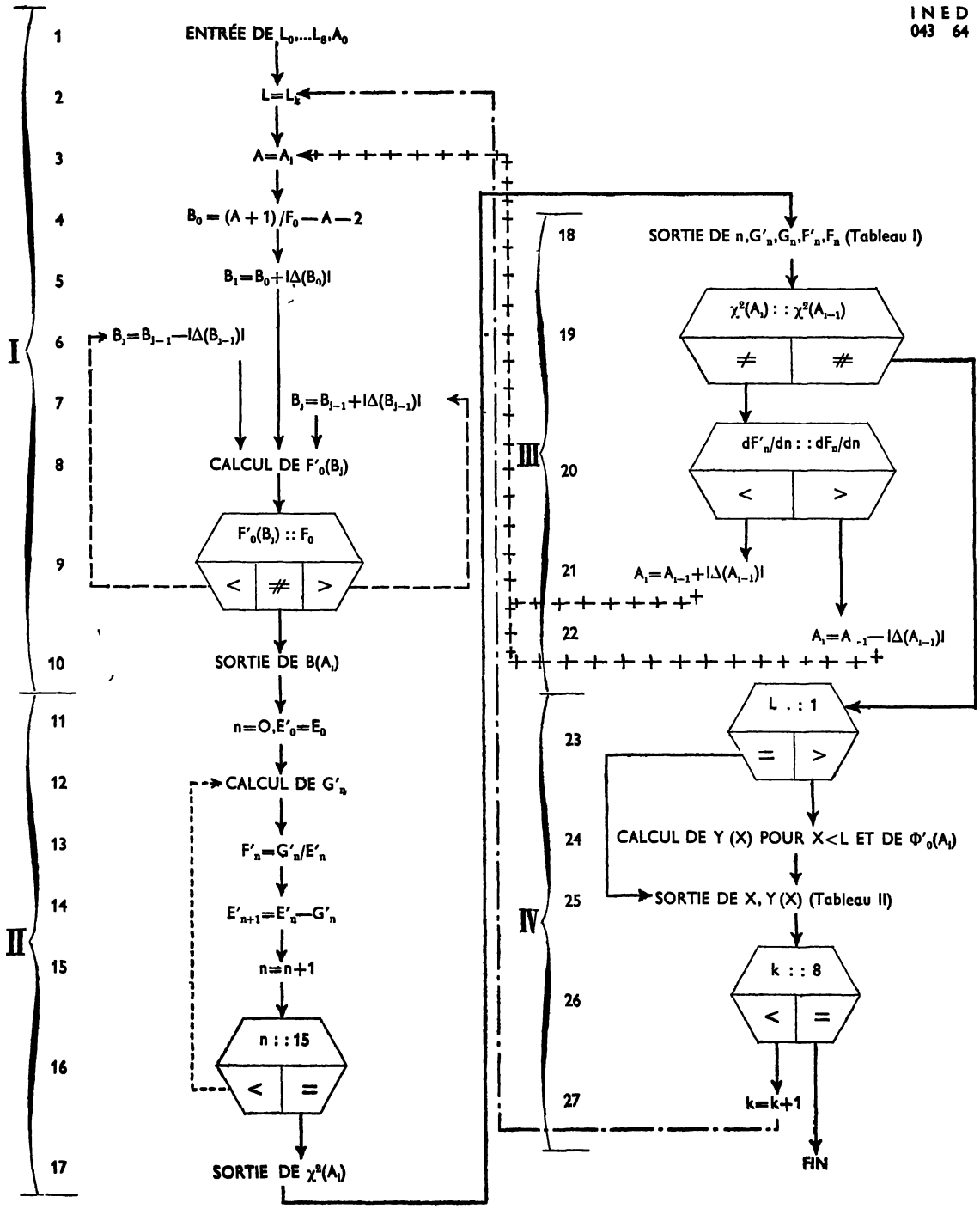


FIGURE 2. — Organigramme schématique illustrant le processus de calcul adopté.

L'intérêt de la figure 2 est de faciliter la conception d'un processus automatique de traitement de notre problème. A la première ligne, nous avons prévu l'introduction de 9 valeurs de L, qui s'échelonneront de 1 à 9 pour couvrir toutes les valeurs « possibles », compte tenu de ce que nous savons. Mais ces valeurs ne sont pas également vraisemblables, et nous placerons en tête les plus plausibles. Si donc nous ordonnons les quantités L_k d'après la valeur de leur indice ($0 \leqslant k \leqslant 8$), nous poserons, par exemple :

$L_0 = 6$; $L_1 = 7$; $L_2 = 5$; $L_3 = 4$; $L_4 = 8$; $L_5 = 3$; $L_6 = 2$; $L_7 = 9$; $L_8 = 1$;

et nous commencerons le calcul en faisant $L = L_0$ (ligne 2; $k = 0$ étant sous-entendu comme valeur initiale de k.)

Nous partirons de même d'une valeur de a arbitraire : A_0 (*cf.* ligne 3; $i = 0$ étant sous-entendu comme valeur initiale de i), cette valeur A_0 ayant été elle-même introduite au début des opérations (on prendra, par exemple, $A_0 = 2$).

La ligne 4 nous donne une valeur approximative de b calculée par la formule 17. Sachant que cette valeur B_0 est trop petite, nous la majorons aussitôt — ce que nous avons indiqué par la notation de la ligne 5 —, et nous passons au calcul de la valeur de f'_0 correspondante (ligne 8). Cette valeur est aussitôt comparée à f_0 (ligne 9), et selon le signe de la différence d (formule 18), la quantité B_1 est affectée d'une correction négative (ligne 6) ou positive (ligne 7).

Naturellement, dès qu'on dispose de deux valeurs consécutives de f'_0 — correspondant respectivement à B_1 et B_2 au début des essais —, on interpole pour évaluer la correction $\Delta\,(B_{j-1})$ à faire subir à la dernière valeur calculée B_{j-1}, pour obtenir la nouvelle valeur à essayer B_j. Lorsque la valeur f'_0 est suffisamment proche de f_0, on considère qu'on est en possession de b(a), ce qui marque la fin (ligne 10) d'une première phase du travail.

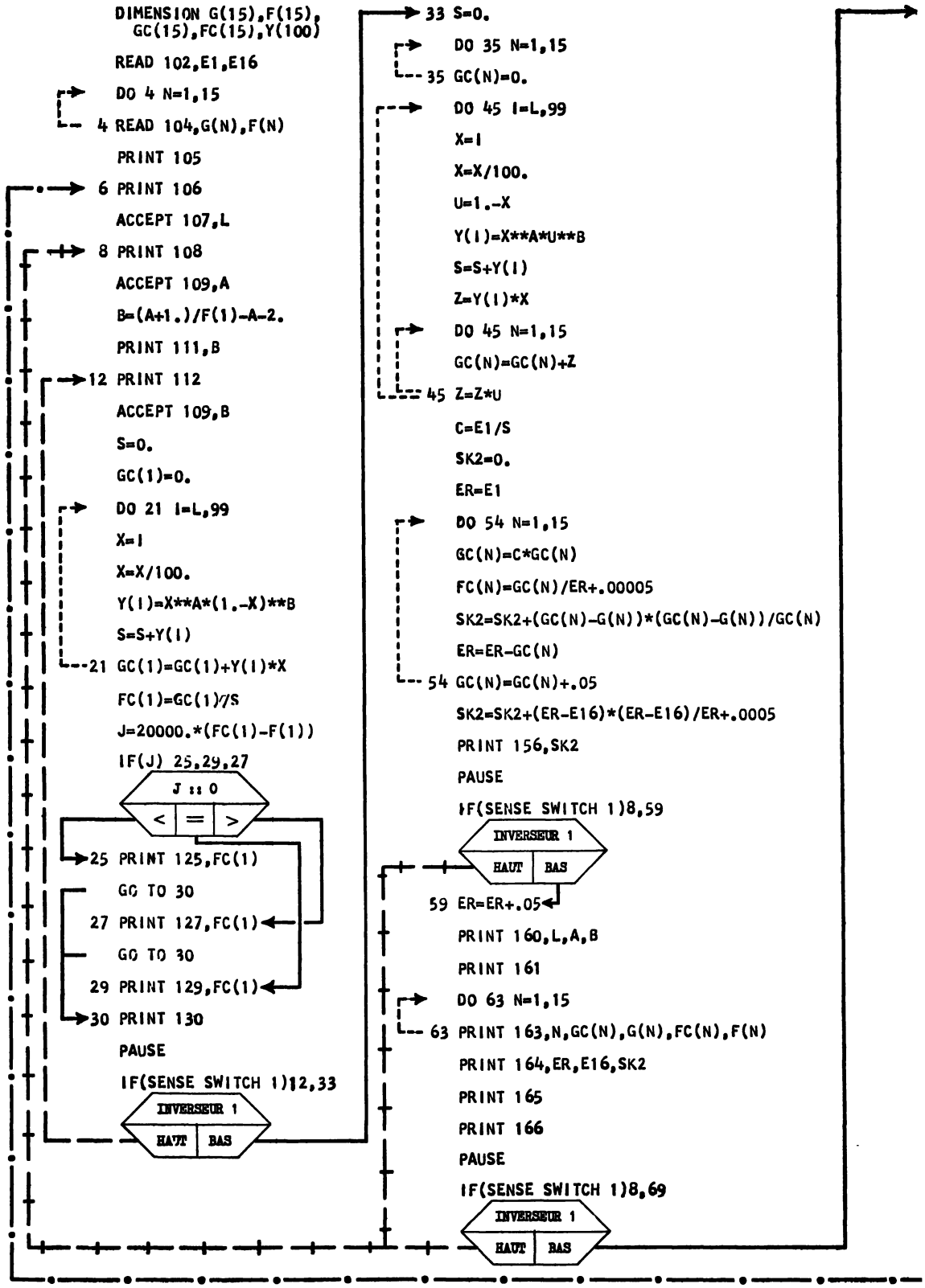
La deuxième phase (lignes 11 à 17) a déjà été décrite en détail et n'appelle aucun commentaire : elle aboutit à la valeur de χ^2 correspondant à la valeur de a essayée.

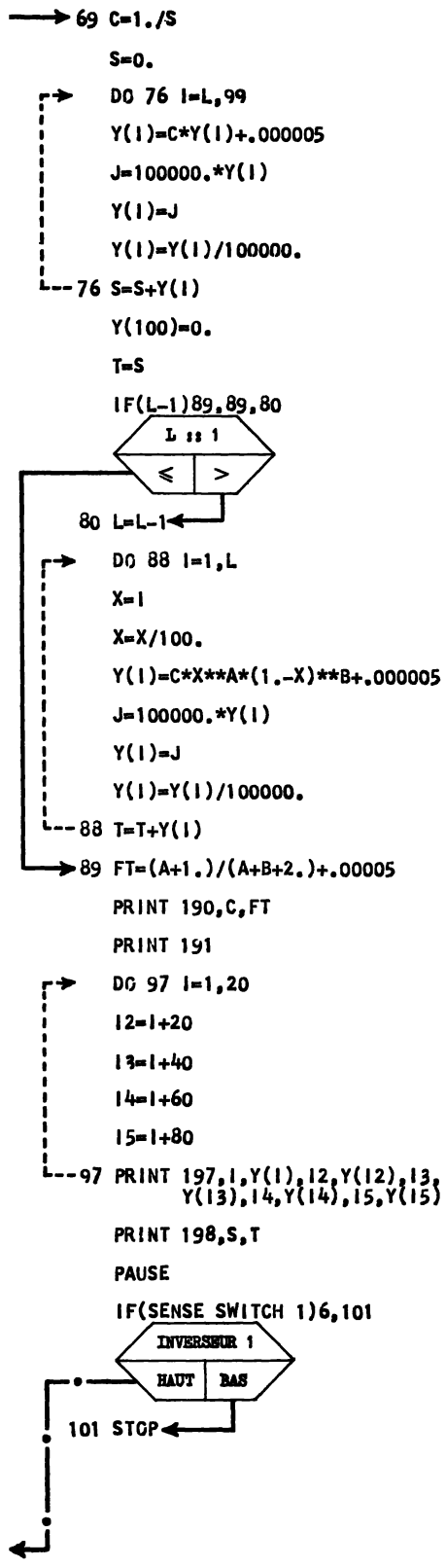
La phase suivante comprend (ligne 18) l'impression d'un tableau qui permet de comparer les quantités calculées aux données d'observation correspondantes. Dès qu'on a essayé deux valeurs de a (pour une même valeur de L), on peut établir une comparaison entre les χ^2 correspondants (ligne 19). Le sens de la correction à faire subir à a pour essayer d'améliorer le modèle, résulte de la comparaison indiquée à la ligne 20 : si f'_n décroît plus vite que f_n en fonction de la durée n du mariage, on doit augmenter a (ligne 21); si, au contraire, la fécondabilité calculée décroît moins vite que la fécondabilité observée, il faut diminuer a (ligne 22). On arrive ainsi à déterminer la valeur de a minimisant χ^2 pour la valeur de L essayée, ce qui marque la fin de la troisième phase du calcul.

La quatrième phase comporte (ligne 25) l'impression de la distribution ayant pour paramètres a, b(a) et l (formule 24). Pour toutes les valeurs de L autres que 1, les ordonnées y_i fournies par les formules 28 et 23, sont calculées pour $1 \leqslant i < L$, afin d'obtenir, en même temps que la distribution tronquée essayée, une répartition proportionnelle à la distribution β complète de mêmes paramètres a et b, et de connaître la fécondabilité moyenne φ_0' correspontant à cette distribution complète (*cf.* ligne 24).

Le processus est repris (à la ligne 2) avec une nouvelle valeur de L (ligne 27), tant qu'il reste des valeurs de L à essayer (ligne 26). A moins qu'on n'ait la possibilité d'interrompre le déroulement des opérations au vu des résultats obtenus, dès lors qu'il s'avère inutile de continuer les essais prévus.

* * *


DIMENSION G(15),F(15),
GC(15),FC(15),Y(100)
READ 102,E1,E16
DO 4 N=1,15
4 READ 104,G(N),F(N)
PRINT 105
6 PRINT 106
ACCEPT 107,L
8 PRINT 108
ACCEPT 109,A
B=(A+1.)/F(1)-A-2.
PRINT 111,B
12 PRINT 112
ACCEPT 109,B
S=0.
GC(1)=0.
DO 21 I=L,99
X=I
X=X/100.
Y(I)=X**A*(1.-X)**B
S=S+Y(I)
21 GC(1)=GC(1)+Y(I)*X
FC(1)=GC(1)/S
J=20000.*(FC(1)-F(1))
IF(J) 25,29,27
J :: 0
< = >
25 PRINT 125,FC(1)
GO TO 30
27 PRINT 127,FC(1)
GO TO 30
29 PRINT 129,FC(1)
30 PRINT 130
PAUSE
IF(SENSE SWITCH 1)12,33
INVERSEUR 1
HAUT BAS
33 S=0.
DO 35 N=1,15
35 GC(N)=0.
DO 45 I=L,99
X=I
X=X/100.
U=1.-X
Y(I)=X**A*U**B
S=S+Y(I)
Z=Y(I)*X
DO 45 N=1,15
GC(N)=GC(N)+Z
45 Z=Z*U
C=E1/S
SK2=0.
ER=E1
DO 54 N=1,15
GC(N)=C*GC(N)
FC(N)=GC(N)/ER+.00005
SK2=SK2+(GC(N)-G(N))*(GC(N)-G(N))/GC(N)
ER=ER-GC(N)
54 GC(N)=GC(N)+.05
SK2=SK2+(ER-E16)*(ER-E16)/ER+.0005
PRINT 156,SK2
PAUSE
IF(SENSE SWITCH 1)8,59
INVERSEUR 1
HAUT BAS
59 ER=ER+.05
PRINT 160,L,A,B
PRINT 161
DO 63 N=1,15
63 PRINT 163,N,GC(N),G(N),FC(N),F(N)
PRINT 164,ER,E16,SK2
PRINT 165
PRINT 166
PAUSE
IF(SENSE SWITCH 1)8,69
INVERSEUR 1
HAUT BAS




```
102 FORMAT(F5.1,F4.1)
104 FORMAT(F5.1,F4.4)
105 FORMAT(40HELC/017 DISTRIBUTION DE
                        LA FECONDABILITE/)
106 FORMAT(/8HENTRER L)
107 FORMAT(2X12)
108 FORMAT(/8HENTRER A)
109 FORMAT(2XE13.0)
111 FORMAT(/15HVALEUR CALCULEE/2HB=E14.8)
112 FORMAT(/8HENTRER B)
125 FORMAT(17HB TROP GRAND FC1=E14.8)
127 FORMAT(17HB TROP PETIT FC1=E14.8)
129 FORMAT(17HB ADMISSIBLE FC1=E14.8)
130 FORMAT(27HPOUR GARDER B BAISSER INV.1)
156 FORMAT(/5HKHI2=F11.3,27H POUR CHANGER A LEVER
                                      INV.1/)
160 FORMAT(/14HRESULTATS POUR//2HL=I3,5H   A=E14.8,
                                  4H  B=E14.8/)
161 FORMAT(6X1HN5X11HGC(N)  G(N)7X12HFC(N)   F(N)/)
163 FORMAT(I7,F10.1,F7.1,F11.4,F7.4)
164 FORMAT(/10H16 ET PLUS=7.1,F7.1,7H  KHI2=F11.3/)
165 FORMAT(36HA TROP PETIT SI FC DECROIT TROP VITE)
166 FORMAT(26HPOUR CHANGER A LEVER INV.1/)
190 FORMAT(20HREPARTITION POUR  C=E14.8//3HFT=F6.4/)
191 FORMAT(3H  X4X1HY5X1HX4X1HY5X1HX4X1HY5X1HX4X1HY
                                  5X1HX4X1HY/)
197 FORMAT(I3,F7.5,I4,F7.5,I4,F7.5,I4,F7.5,I4,F7.5)
198 FORMAT(/2HS=F8.5,4H  T=E14.8//26HPOUR CHANGER
                                  L LEVER INV.1//)
    END
```

```
715105827
180212520
116962187
082551975
061791842
046391696
035401558
026811398
022011334
019071334
015011211
012091110
011611199
010581242
009471269
006881056
```

PRESENTATION DES DONNEES
(perforées en colonnes 1 à 5 et 6 à 9 de 16 cartes)

FIGURE 3. — Organigramme détaillé du programme en FORTRAN pour ordinateur 1620.

Si l'on considère le processus schématisé par la figure 2, une question se présente immédiatement à l'esprit : Jusqu'où a-t-on intérêt à pousser l'automatisation du traitement? Cette question est très importante car, comme nous allons le voir, son examen va nous conduire à adopter un certain type de matériel pour traiter notre problème. Or si, comme nous le pensons, ce problème est assez caractéristique de ceux qui pourront se présenter aux chercheurs-démographes, nous aurons répondu par là même à la question suivante : Quel est le type de matériel le mieux adapté à la recherche démographique? Reprenons donc l'examen du traitement projeté, du point de vue des interventions éventuelles du chercheur.

Il est évident que ces interventions ne sauraient se situer qu'aux endroits où des choix doivent être effectués, et que tous les calculs qui n'impliquent aucun choix seront, en tout état de cause, entièrement automatisés. Tel est, notamment, le cas de la phase II, laquelle comprendrait sans doute le volume principal des calculs, si la phase précédente ne comportait des itérations en nombre égal aux essais nécessaires à la détermination de b(a).

Cependant, disposant d'une valeur approchée de *b*, on pourrait assez facilement automatiser la partie correspondante de la phase I (lignes 5 à 9) : par exemple en prenant cette valeur approchée comme valeur de départ, en l'augmentant d'une fraction constante de sa valeur absolue comme seconde valeur essayée, et en provoquant des interpolations ou extrapolations automatiques selon les résultats obtenus. Ceci ne serait pas difficile à programmer. Il en résulterait certainement un volume de calculs beaucoup plus important qu'en laissant au chercheur le soin d'effectuer un choix raisonné des valeurs de *b* à essayer; mais il serait probablement avantageux de consentir à un tel « gaspillage », comme rançon de l'emploi d'un ensemble puissant.

L'automatisation de l'opération 2 ne soulève, comme nous l'avons vu, aucune difficulté. Mais le « gaspillage » de calculs qui résulterait d'essais superflus serait, cette fois, beaucoup plus important que celui auquel nous étions prêt à consentir pour automatiser la détermination de b(a). En effet, il s'agirait, en l'occurrence, de reprendre entièrement le processus de calcul, de la ligne 2 à la ligne 27, pour toutes les valeurs de L dont l'essai se révélerait finalement inutile, à l'examen des résultats. D'autre part, en admettant même qu'un tel « gaspillage » s'avère admissible, c'est-à-dire rentable, ce ne pourrait être que dans le cadre d'un traitement *entièrement* automatisé. Or il est un choix dont l'automatisation soulèverait des problèmes beaucoup plus complexes, et qui s'insère dans la série des calculs prévus pour une valeur donnée de L, c'est le choix de *a*.

En cours de calcul, ce choix (ligne 3) résulte des considérations indiquées aux lignes 19 à 22 du schéma. Or l'automatisation de cette partie du traitement s'avérerait particulièrement difficile à concevoir, en l'absence de toute expérience préalable nous renseignant, par exemple, sur l'ordre de grandeur de *a*, sur celui du χ^2 correspondant, et sur la sensibilité de χ^2 aux variations de *a*. Car la règle consignée aux lignes 20 à 22, qui indique le sens de la correction à apporter à *a*, peut bien faciliter des interventions raisonnées du chercheur — surtout au début du travail —, parce qu'il saura l'interpréter intelligemment; mais il serait difficile d'en tirer parti « mécaniquement » sans risquer des déboires — notamment en raison du fait que la série des valeurs observées f_n est irrégulière.

Nous devons donc renoncer à automatiser complètement le traitement; d'abord parce que la programmation du problème s'en trouverait sensiblement alourdie et compliquée, ce qui ne manquerait pas de rendre plus coûteuse en temps-machine la mise au point du programme; ensuite parce que cela entraînerait une débauche de calculs superflus, qui risquerait fort de dépasser les limites admissibles.

Nous voyant ainsi contraint de prévoir des interventions raisonnées du maître d'œuvre,

au moins pour le choix des valeurs de a à essayer (ligne 3), et par suite également pour celui des valeurs de L (ligne 2), nous devrons évidemment avoir recours à des matériels qui permettent ces interventions en cours de travail, au vu de résultats exprimés en clair aussitôt qu'obtenus. Et nous n'aurons dès lors aucun intérêt à automatiser la détermination de b(a), puisque nous prévoyons que, là encore, l'être humain pourra affirmer sa supériorité sur la machine, et épargner au matériel non seulement du calcul, mais même du temps de fonctionnement.

Les interventions de l'opérateur se situeront donc aux lignes 2, 3 et 5 à 7. Les opérations indiquées aux lignes 1, 5 à 7, 19 à 22, 26 et 27, seront purement mentales et ne seront pas matérialisées dans les programmes : les parties correspondantes du schéma de la figure 2 s'y réduiront à des instructions d'entrée de valeurs choisies par l'opérateur, pour les quantités L, a et b.

Deux matériels actuellement disponibles sur le marché français, sont particulièrement bien adaptés aux exigences d'un tel travail; ce sont l'ordinateur 1620 de la Compagnie I. B. M., et la calculatrice CAB 500 de la Compagnie des machines Bull. Des circonstances particulières ont voulu que nous les essayons dans cet ordre, ce qui n'est pas sans importance; car il est très difficile de faire abstraction de l'expérience acquise, quand on passe d'un matériel à un autre pour une même application. Nous espérons pourtant avoir fait un effort d'objectivité suffisant, pour minimiser les conséquences inévitables du déroulement chronologique de nos essais.

* * *

La figure 3 présente, sous forme d'un organigramme, un programme en FORTRAN correspondant au traitement du problème proposé par ordinateur 1620. Pour faciliter les références à ce programme, nous avons donné aux instructions numérotées du programme proprement dit, un numéro correspondant au rang de la carte où la dite instruction se trouve perforée. D'autre part, les numéros des instructions « FORMAT »— qui indiquent à la machine sous quelle forme seront présentées les données ou devront être imprimés les résultats — rappellent le rang des instructions du programme qui s'y réfèrent.

La première instruction ordonne une « réservation d'indices » : 15 zones de mémoire seront réservées pour le rangement de chacune des quantités G, F, GC (correspondant à G') et FC (correspondant à f'), et 100 zones pour le rangement de Y. Noter que, le FORTRAN n'admettant pas l'indice zéro, la valeur maximale de l'indice (indiquée entre parenthèses) est parfois supérieure d'une unité à celle que nous avions envisagée dans nos formules.

Les instructions 2 à 4 provoquent l'entrée des données, lesquelles sont perforées dans des cartes sous la forme explicitée à la fin de la figure. L'instruction 2 signifie : « Entrer à partir d'une carte perforée, selon le modèle 102, les quantités E_1 et E_{16}». Elle se réfère à l'instruction 102, qui indique que les quantités en cause sont à enregistrer sous une forme normalisée dite « en virgule flottante », et qu'elles comportent respectivement 5 chiffres dont 1 à droite de la virgule, et 4 chiffres dont 1 à droite de la virgule. On voit qu'il s'agit de l'effectif initial ($E_0 = 7151{,}0$) et de l'effectif restant au début du 16e mois d'observation ($E_{15} = 582{,}7$).

L'instruction 3, qui marque le départ d'une boucle, s'interprète ainsi : « Exécuter les instructions qui suivent, jusqu'à l'instruction 4 inclusivement, pour les valeurs de N allant de 1 à 15 inclusivement ». La boucle se réduisant à la seule instruction 4, elle provoquera la lecture de 15 cartes perforées, avec enregistrement des données G_n et f_n qui y sont indiquées, les valeurs de l'indice affecté aux lettres G et F, étant supérieures d'une unité aux valeurs de l'indice n considéré dans notre exposé de principe.

L'instruction 6 provoquera l'impression d'une indication pour l'opérateur : « Entrer L ». Elle est suivie de l'instruction 7, qui provoquera la mise du calculateur en « attente de lecture » à partir du clavier de la machine à écrire. L'opérateur devra fournir la valeur de L sous la forme dite « à virgule fixe » indiquée par l'instruction 107, c'est-à-dire précédée de 2 caractères qui seront ignorés par le calculateur, et comportant elle-même deux caractères : en pratique, on tapera par exemple L=06, et le calculateur enregistrera le nombre entier 6 affecté du signe + (sous-entendu en l'absence de signe —) comme valeur de L.

Les instructions 8 et 9 correspondent de même à l'entrée de a; le matériel ne permettant que l'emploi de majuscules pour la désignation des symboles, cette quantité est appelée ici A. L'instruction 109 indique que la valeur du paramètre devra être fournie, précédée de deux caractères qui seront ignorés par le calculateur, sous une forme normalisée à 13 caractères, du type : +2.8000000+00 (pour $a = + 2{,}8$ — le point ayant valeur de virgule, et les trois derniers caractères représentant l'exposant de 10 affecté de son signe). On a prévu ce mode d'entrée de la valeur de a, faute de notions sur son ordre de grandeur et sur la précision relative nécessaire. Cette précaution permettra, si besoin est, d'entrer a en tapant, par exemple, A=+26.327500—12 au clavier, ce qui donnera au paramètre a la valeur $+ 26{,}3275 \times 10^{-12}$.

Les instructions 10 et 11 provoqueront le calcul de la valeur approchée de b (appelé B) et sa sortie sous forme normalisée. L'instruction 111 indique que la valeur de b sera imprimée à 8 chiffres significatifs, précédés du signe (1) et du point décimal (valant virgule), et suivis de la lettre E (signifiant « exposant ») et de la valeur de l'exposant de 10, signe compris.

L'entrée des valeurs essayées de b s'effectuera, sur l'indication donnée par l'instruction 12, de façon absolument comparable à l'entrée de a; aussi l'instruction 13 se réfère-t-elle à la même instruction « format » (109) que l'instruction 9.

Après les instructions 14 et 15, annulant les mémoires destinées au rangement des sommes S et GC(1) qui seront obtenues par cumul (*cf.* instructions 20 et 21), commence une boucle correspondant aux sommations des formules 25 et 26 de l'exposé de principe. On reconnaît, dans l'instruction 16, les limites de variation de l'indice i. L'instruction 17 provoquera une conversion en virgule flottante, du nombre I qui se trouve emmagasiné en virgule fixe. L'instruction 18 entraînera alors le calcul prévu par la formule 1, tandis que l'instruction suivante fournira la valeur de Y_1 donnée par la formule 23 (l'astérisque valant, en FORTRAN, signe d'exponentiation ou de multiplication, selon qu'il est répété ou non).

On reconnaît la formule 22 dans l'instruction 22, et le calcul de d par la formule 18 dans la parenthèse de l'instruction 23. La multiplication de d par le coefficient 20000 a pour but de préparer le test sur J, qui est une quantité exprimée « en virgule fixe », autrement dit un entier. Si $d < 0{,}5 \times 10^{-4}$, on aura $20000\ d < 1$, et comme l'instruction 23 entraîne conversion en virgule fixe, du nombre $20000\ d$ exprimé en virgule flottante, par suppression de sa partie décimale, on aura alors $J = 0$. Le test de l'instruction 24, explicité par les indications portées dans l'hexagone qui succède à cette instruction sur l'organigramme, provoquera alors des aiguillages en fonction du signe et de la grandeur de d : si l'écart entre f'_0 et f_0 s'avère inférieur à la précision estimée nécessaire, f'_0 comportera 4 décimales exactes et sera considéré comme suffisamment voisin de f_0 pour la détermination de b(a) — *cf.* l'instruction 29, qui entraîne l'impression de l'indication suivante : « B admissible FC1= », suivie de la valeur de la fécondabilité calculée pour le début du premier mois d'observation.

On remarquera que l'instruction 24 — qui se lit : « Si J est inférieur, égal, ou supérieur à zéro, aller respectivement en 25, 29 ou 27 » — conduit toujours finalement, après impression

(1) Remplacé par un espacement, s'il s'agit du signe +.

d'une indication destinée à renseigner l'opérateur, à la même instruction 30, laquelle est suivie d'une « pause » (instruction 31), c'est-à-dire d'une mise de la machine en attente d'une intervention de l'opérateur. C'est que nous avons voulu réserver à celui-ci la possibilité, soit de poursuivre le calcul quand bien même la différence d serait plus importante que celle considérée à priori comme « admissible », soit au contraire de reprendre le calcul en 12 pour améliorer une approximation de b(a) qui semblait à priori déjà acceptable. Le passage de la phase I à la phase II du calcul ne s'effectuera donc pas automatiquement, mais sur intervention de l'opérateur qui devra, à cette fin, placer l'inverseur n° 1 du pupitre de commande dans la position basse (*cf.* instruction 32).

Le calcul de Y_i par la formule 23 sera alors repris, de même que celui de la somme S (*cf.* instructions 39 à 41), mais en y associant cette fois celui des 15 quantités G'_n. Ceci s'effectuera grâce à la boucle supplémentaire commandée par l'instruction 43, boucle dans laquelle la quantité Z joue un rôle comparable à la quantité $g_{n,1}$ de la formule 7.

Les instructions suivantes correspondent au calcul de la quantité χ^2, appelée ici SK2. La répartition calculée est mise à l'échelle de l'échantillon (*cf.* instruction 50), grâce au coefficient C fourni par l'instruction 46. L'effectif restant en observation, ER, est calculé selon la formule 9 (*cf.* instruction 53). On profite de la boucle pour calculer f'_n avec arrondi sur la 4e décimale (*cf.* instruction 51). Des arrondis sont de même prévus sur la première décimale de G'_n (instruction 54) et sur la 3e décimale de χ^2 (instruction 55). Mais seule la valeur de χ^2 est sortie à ce moment.

En effet, l'impression du tableau I prévue à la phase III, est une opération assez lente qu'il peut être intéressant de supprimer pour certaines valeurs de a. On a donc prévu la possibilité de modifier a dès la fin de la phase II, au seul vu de la valeur de χ^2, par le moyen de l'inverseur n° 1, conformément à l'instruction 58.

Si l'opérateur veut provoquer l'impression du tableau I pour la valeur de a essayée, il n'aura qu'à disposer l'inverseur n° 1 vers le bas, et il aura encore la faculté de modifier a après impression du tableau I, conformément à l'instruction 68.

La réalisation de la phase IV n'appellerait aucun commentaire, si nous n'avions prétendu sortir le total exact des valeurs y_i *imprimées* par le calculateur. Paradoxalement, c'est un problème généralement difficile à programmer, et nous n'avons pu le faire ici en FORTRAN, que parce que les éléments à totaliser ne comportaient, en l'occurrence, qu'un nombre restreint de chiffres significatifs. Les instructions 73 à 75 et 85 à 87, qui réalisent une double conversion de virgule flottante en virgule fixe, puis de virgule fixe en virgule flottante, afin de supprimer les chiffres négligés à l'impression, permettent la sortie de totaux S et T (*cf.* instruction 98) correspondant effectivement aux valeurs de y_i imprimées par la machine. Mais nous devons souligner le fait qu'il s'agit d'une solution de fortune, qui ne saurait avoir une portée générale.

La question n'est pourtant pas sans intérêt, car la connaissance du total facilite grandement la vérification d'une série de nombres transcrits. Mais la difficulté tient au principe même du calcul en virgule flottante, et elle devient même sérieuse lorsque la représentation interne des nombres s'effectue en binaire pur.

*
* *

La figure 4 reproduit, en vraie grandeur, des résultats obtenus grâce au programme précédent pour $L = 6$ et $a = 0{,}56$. On remarquera que la première valeur de b essayée, soit 4,19, est assez différente de la valeur « calculée » de 3,63, et fort voisine de la valeur « admissible » de 4,194 qui sera finalement retenue pour b(a). Ceci tient au fait que la valeur de départ

ELC/017 DISTRIBUTION DE LA FECONDABILITE

ENTRER L
L=06RS

ENTRER A
A=+.56000000+00RS

VALEUR CALCULEE
B= .36304761E+01

ENTRER B
B=+4.1900000+00RS
B TROP PETIT FC1= .25212556E-00
POUR GARDER B BAISSER INV.1

ENTRER B
B=+4.2000000+00RS
B TROP GRAND FC1= .25180653E-00
POUR GARDER B BAISSER INV.1

ENTRER B
B=+4.1940000+00RS
B ADMISSIBLE FC1= .25199792E-00
POUR GARDER B BAISSER INV.1

KHI2= 13.526 POUR CHANGER A LEVER INV.1

RESULTATS POUR

L= 6 A= .56000000E-00 B= .41940000E+01

N	GC(N)	G(N)	FC(N)	F(N)
1	1802.0	1802.1	.2520	.2520
2	1199.9	1169.6	.2243	.2187
3	843.0	825.5	.2032	.1975
4	616.7	617.9	.1865	.1842
5	465.5	463.9	.1731	.1696
6	360.3	354.0	.1620	.1558
7	284.7	268.1	.1528	.1398
8	228.8	220.1	.1449	.1334
9	186.6	190.7	.1382	.1334
10	154.0	150.1	.1324	.1211
11	128.5	120.9	.1273	.1110
12	108.2	116.1	.1228	.1199
13	91.8	105.8	.1188	.1242
14	78.5	94.7	.1152	.1269
15	67.5	68.8	.1120	.1056
16 ET PLUS	535.1	582.7	KHI2=	13.526

```
A TROP PETIT SI FC DECROIT TROP VITE
POUR CHANGER A LEVER INV.1

REPARTITION POUR  C= .17575993E-00

FT= .2310

  X    Y       X    Y       X    Y       X    Y       X    Y
  1 .01278    21 .02729    41 .01167    61 .00257    81 .00015
  2 .01806    22 .02655    42 .01101    62 .00232    82 .00012
  3 .02171    23 .02579    43 .01037    63 .00210    83 .00009
  4 .02442    24 .02500    44 .00975    64 .00189    84 .00007
  5 .02648    25 .02420    45 .00916    65 .00169    85 .00006
  6 .02805    26 .02338    46 .00858    66 .00151    86 .00004
  7 .02924    27 .02256    47 .00803    67 .00134    87 .00003
  8 .03011    28 .02173    48 .00750    68 .00119    88 .00002
  9 .03073    29 .02089    49 .00700    69 .00105    89 .00002
 10 .03112    30 .02007    50 .00651    70 .00092    90 .00001
 11 .03132    31 .01924    51 .00605    71 .00081    91 .00001
 12 .03136    32 .01842    52 .00561    72 .00070    92 .00000
 13 .03127    33 .01761    53 .00519    73 .00061    93 .00000
 14 .03105    34 .01682    54 .00479    74 .00052    94 .00000
 15 .03073    35 .01603    55 .00442    75 .00045    95 .00000
 16 .03031    36 .01526    56 .00406    76 .00038    96 .00000
 17 .02983    37 .01451    57 .00372    77 .00032    97 .00000
 18 .02927    38 .01377    58 .00341    78 .00027    98 .00000
 19 .02866    39 .01305    59 .00311    79 .00022    99 .00000
 20 .02799    40 .01235    60 .00283    80 .00018   100 .00000

S=  .99999   T= .11034400E+01

POUR CHANGER L LEVER INV.1

                                                      I N E D
ENTRER L                                              045  64
```

FIGURE 4. — Résultats d'exécution du programme en FORTRAN par ordinateur 1620.

a été déterminée en tirant parti de résultats obtenus antérieurement pour d'autres valeurs de a. On imagine sans peine, à la lumière de cet exemple, la quantité de calculs qui peut être épargnée par un choix judicieux de la valeur de départ.

Cependant, la situation de l'opérateur chargé d'une exploitation de ce genre, est assez inconfortable; s'il a conscience du coût du matériel qu'il utilise, et du fait que le rendement de ce matériel dépend essentiellement de la qualité et de la rapidité de son propre travail, il devient facilement l'esclave de la machine.

En effet, nous avons pu constater qu'en utilisant convenablement les résultats déjà acquis, on peut assez souvent obtenir une valeur de b « admissible » dès le second essai. Mais ceci exige des reports de points sur des graphiques, des tracés de courbes, des interpolations ou extrapolations susceptibles de fournir des éléments d'appréciation pour l'évaluation de la première valeur à essayer, et pour l'estimation de la correction à apporter ultérieurement à cette valeur en fonction du résultat qu'elle aura fourni. Or l'opérateur ne dispose, pour faire tout ce travail sans ralentir la machine, que du temps que celle-ci consacre au calcul ou à l'impression des tableaux : un temps qui se fractionne en périodes de 2 à 3 minutes !

Bien plus, lorsque la machine vient, par exemple, d'imprimer la valeur « FC1 » correspondant à la première valeur de b essayée, *elle attend* qu'on lui fournisse une valeur corrigée. Si l'opérateur n'est pas en mesure de calculer très rapidement la correction à effectuer, il se trouve alors devant le dilemme suivant : ou bien entrer immédiatement une nouvelle valeur de b qui ne sera sûrement pas « admissible », mais qui lui procurera dans quelque 3 minutes une bonne base d'interpolation ou d'extrapolation; ou bien tenter de *gagner du temps*, en laissant la machine en attente jusqu'à ce qu'il ait pu évaluer pertinemment la correction à partir des éléments disponibles. S'il choisit la première solution, il aura sans doute l'impression de s'être accordé un « répit » de quelques minutes pour reporter ses points, tracer ses courbes, etc. Mais il ne tardera pas à s'apercevoir qu'au bout de ce délai, la machine lui aura fourni de nouveaux éléments d'appréciation qu'il ne pourra pas ignorer... Une lutte de vitesse tend ainsi à s'instaurer entre l'opérateur et la machine, lutte qui ne tarde pas à devenir épuisante pour l'opérateur, s'il n'a pas une parfaite maîtrise du sujet traité.

Le chercheur aura donc tout intérêt à assumer personnellement la direction de l'exploitation, plutôt que de confier ce travail à un tiers qui ne saurait avoir, du problème à traiter, une connaissance aussi approfondie que la sienne. Mais alors, on devra s'interroger sur l'opportunité de prévoir, comme nous l'avons fait ici, l'impression sur l'état de directives destinées à faciliter le travail de l'opérateur. Car, pour qui connaît bien la structure d'un programme, un bref schéma rappelant les articulations essentielles de ce programme, constitue un guide opératoire très suffisant. Or l'impression de directives sur l'état alourdit la programmation et risque d'entraîner des dépassements de la capacité d'enregistrement en mémoire. En outre, cette impression ralentit l'exploitation et peut gêner la lecture des résultats.

Ce dernier inconvénient est sensible sur la partie de la figure 4 correspondant à la première phase du calcul. La deuxième phase se trouve matérialisée, sur la même figure, par une seule ligne d'impression : celle donnant la valeur de « KHI2 ». Le tableau qui suit correspond à la troisième phase. La comparaison des deux colonnes de droite de ce tableau, permet de penser qu'il est possible d'améliorer le modèle pour $L = 6$, en diminuant a . Pourtant la valeur de χ^2 montre que la distribution en cause est déjà « acceptable » : pour 14 degrés de liberté (car nous imposons ici aux répartitions essayées, une fécondabilité initiale donnée), la probabilité associée à $\chi^2 = 13{,}5$ est en effet voisine de 0,50.

Le deuxième tableau, lu à partir de $X = 6$, fournit la distribution discrète de paramètres $L = 6$ et $a = 0{,}56$. Les valeurs x_i portées dans les colonnes X, sont exprimées pour cent. On vérifie que la somme S de la distribution est, aux arrondis près, égale à 1. Le coefficient C imprimé en tête du tableau, est celui de la distribution discrète : pour obtenir le coefficient c de la distribution continue correspondant aux formules 14 et 15 (avec $l = 0{,}055$ conformément à la formule 24), il convient de multiplier la valeur de C par 100.

Les valeurs de Y données pour $X < 6$, permettent de se faire une idée de la forme qu'aurait une distribution β complète de mêmes paramètres $a = 0{,}56$ et $b = 4{,}194$. La répar-

tition ainsi complétée aurait une fécondabilité initiale FT = 0,231 (calculée par la formule 16 — *cf.* instruction 89), et un effectif total T = 1,10344 à rapprocher du total S résultant de la troncature.

*
* *

La figure 5 présente, sous forme d'un organigramme, un programme écrit en langage PAF pour le traitement de notre problème par calculatrice CAB 500. Ce programme n'est pas une simple transposition de celui de la figure 3; ceci pour plusieurs raisons, dont la principale est qu'il existe entre le 1620 et la CAB 500 d'importantes différences dont nous devions tenir compte.

Ces différences peuvent, pour l'essentiel, s'exprimer ainsi : le 1620 calcule vite, mais ne dispose que d'une capacité de mémoire limitée; la CAB 500 calcule plus lentement, mais dispose d'une grande capacité de mémoire. Il en résulte que, quand on programme pour le 1620, on consent volontiers à répéter des calculs pour économiser de la place en mémoire; tandis que, quand on programme pour la CAB 500, on cherche à tirer parti de la grande capacité de mémoire pour réduire le volume des calculs.

C'est la raison pour laquelle notre programme en PAF comporte le calcul d'une table des valeurs de :

$$X_i = \mathrm{Log_e}\, x_i, \text{ avec } x_i = \frac{i}{100}, \text{ pour } 0 < i < 100$$

et le rangement de ces logarithmes en ordre inversé pour constituer simultanément une table des valeurs de :

$$U_i = \mathrm{Log_e}\,(1 - x_i)$$

La confection de ces tables se situe aussitôt après l'entrée des données, et fait l'objet de la deuxième boucle du programme proprement dit. Celui-ci est toutefois précédé de quelques instructions non numérotées destinées, les unes à effectuer des réservations d'indices, les autres à créer des sous-programmes pour certains calculs numériques.

Dans les quatre premières instructions non numérotées, les lettres destinées à être indicées figurent à la suite de la valeur maximale prévue pour leur indice. En PAF, les indices peuvent prendre la valeur zéro, ce qui nous a permis de leur attribuer ici la même signification que dans notre exposé de principe. Quant au choix des symboles, il répond à un souci mnémonique : Q désigne le « quotient observé », F la « fécondabilité calculée »; G les « grossesses observées », C les « conceptions calculées » — G_{15} et C_{15} se rapportant à la classe « 15 mois et plus ».

Les instructions relatives aux réservations d'indices sont suivies de deux « formules ». La première correspond au calcul de Y_i par la formule 23, écrite sous la forme :

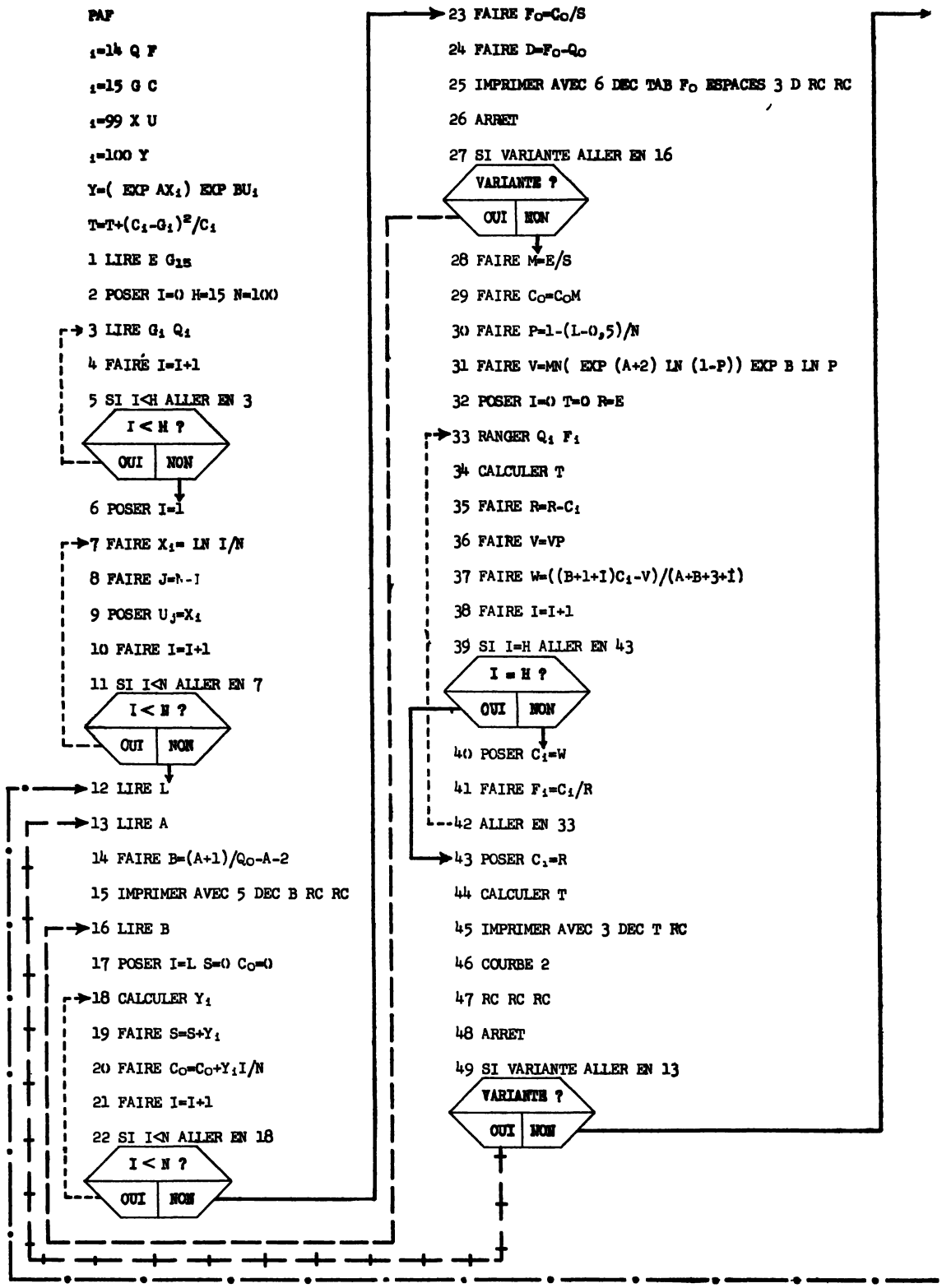
$$Y_i = e^{a\,\mathrm{Log_e}\,x_i}\, e^{b\,\mathrm{Log_e}\,(1 - x_i)}$$

La seconde permet d'obtenir χ^2 par cumul (T rappelant le mot « test »).

Le programme proprement dit commence ensuite par une instruction d'entrée des données. Celles-ci étant présentées dans l'ordre indiqué à la fin de la figure, on voit que E représente l'effectif initial de l'ensemble observé.

L'instruction 2 permet, d'une part de fixer la valeur initiale de l'indice I utilisé dans la boucle qui suit, d'autre part de déterminer deux quantités (H = 15 et N = 100) qui serviront de constantes de référence tout au long du programme.

La boucle constituée par les instructions 3 à 5 permet de compléter l'entrée des données. Signalons à cette occasion qu'en PAF, les boucles sont construites en détail par le program-


PAF
i=14 Q F
i=15 G C
i=99 X U
i=100 Y
Y=(EXP AXi) EXP BUi
T=T+(Ci-Gi)²/Ci
1 LIRE E G15
2 POSER I=0 H=15 N=100
3 LIRE Gi Qi
4 FAIRE I=I+1
5 SI I<H ALLER EN 3
I < H ?
OUI NON
6 POSER I=1
7 FAIRE Xi= LN I/N
8 FAIRE J=N-I
9 POSER Uj=Xi
10 FAIRE I=I+1
11 SI I<N ALLER EN 7
I < N ?
OUI NON
12 LIRE L
13 LIRE A
14 FAIRE B=(A+1)/Q0-A-2
15 IMPRIMER AVEC 5 DEC B RC RC
16 LIRE B
17 POSER I=L S=0 C0=0
18 CALCULER Yi
19 FAIRE S=S+Yi
20 FAIRE C0=C0+YiI/N
21 FAIRE I=I+1
22 SI I<N ALLER EN 18
I < N ?
OUI NON
23 FAIRE F0=C0/S
24 FAIRE D=F0-Q0
25 IMPRIMER AVEC 6 DEC TAB F0 ESPACES 3 D RC RC
26 ARRET
27 SI VARIANTE ALLER EN 16
VARIANTE ?
OUI NON
28 FAIRE M=E/S
29 FAIRE C0=C0M
30 FAIRE P=1-(L-0,5)/N
31 FAIRE V=MN(EXP (A+2) LN (1-P)) EXP B LN P
32 POSER I=0 T=0 R=E
33 RANGER Qi Fi
34 CALCULER T
35 FAIRE R=R-Ci
36 FAIRE V=VP
37 FAIRE W=((B+1+I)Ci-V)/(A+B+3+I)
38 FAIRE I=I+1
39 SI I=H ALLER EN 43
I = H ?
OUI NON
40 POSER Ci=W
41 FAIRE Fi=Ci/R
42 ALLER EN 33
43 POSER Ci=R
44 CALCULER T
45 IMPRIMER AVEC 3 DEC T RC
46 COURBE 2
47 RC RC RC
48 ARRET
49 SI VARIANTE ALLER EN 13
VARIANTE ?
OUI NON

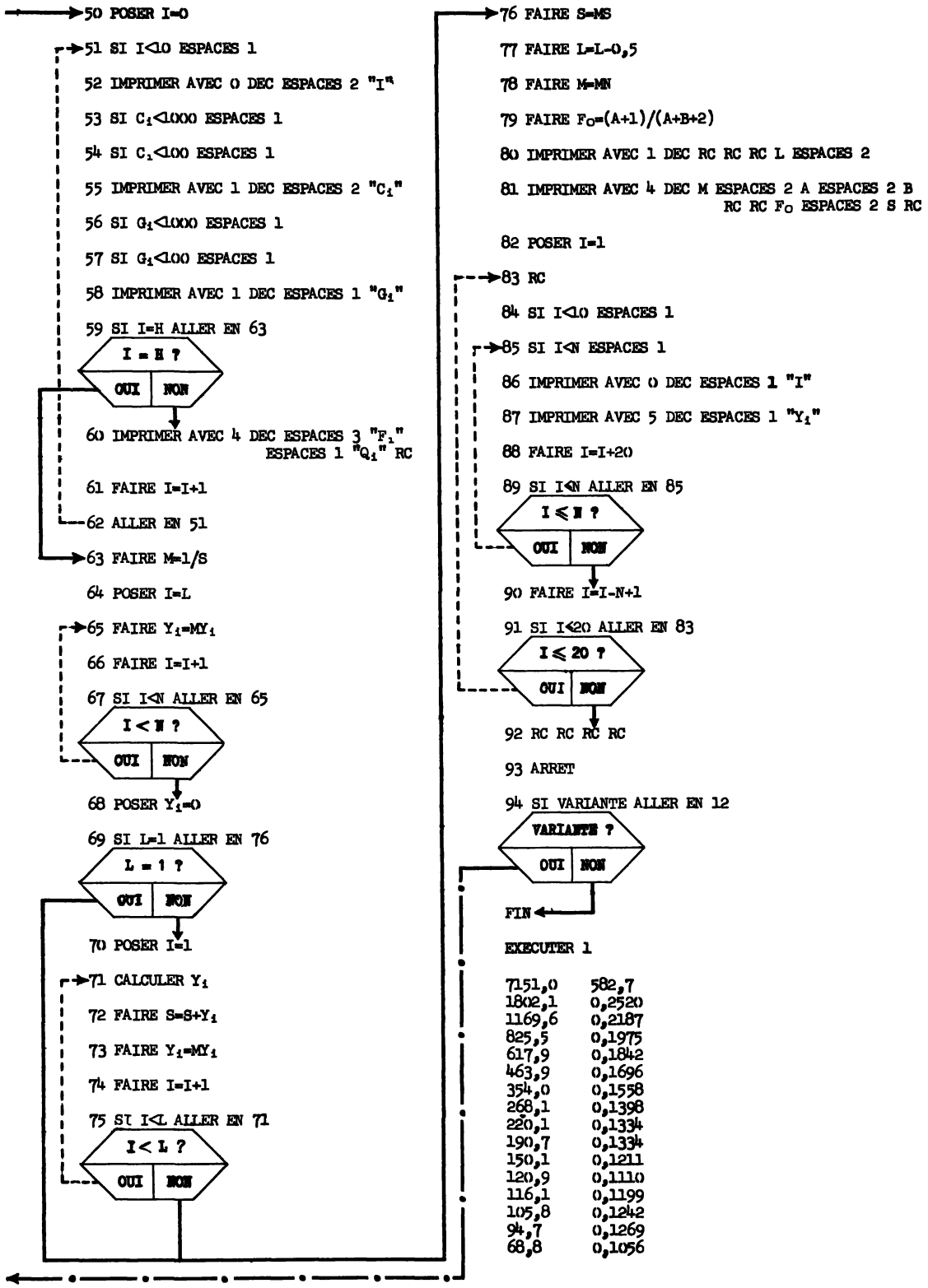


FIGURE 5. — Organigramme détaillé du programme en PAF pour CAB 500.

meur. Cette solution, qui paraît à première vue plus compliquée que celle du FORTRAN, s'avère à l'usage avantageuse par sa souplesse.

La boucle destinée à la confection des tables est préparée par l'instruction 6. L'instruction 7 provoque le calcul du logarithme népérien (noté par le groupe LN encadré d'espaces) des quantités I/100, et le rangement de ces logarithmes dans l'ordre des I croissants. Les instructions 8 et 9 provoquent le rangement des mêmes logarithmes dans l'ordre inverse.

On reprend ensuite quelque temps le processus de traitement déjà décrit. Entrée de L, puis de *a*, au clavier (*cf.* instructions 12 et 13). Calcul et impression de la « valeur approchée » de *b* (*cf.* instructions 14 et 15). Entrée au clavier de la valeur de *b* à essayer (*cf.* instruction 16). Calcul de Y_1, effectué ici par référence à la « formule » insérée en début de programme, grâce à l'instruction 18. Calcul, par cumul, des sommes correspondant aux formules 25 et 26 (*cf.* instructions 19 et 20). Détermination de f'_0 par la formule 22 (*cf.* instruction 23), puis de *d* par la formule 18 (*cf.* instruction 24), et impression avec 6 décimales des valeurs de ces deux quantités (*cf.* instruction 25).

Un arrêt de la machine est alors prévu, pour permettre à l'opérateur de décider s'il doit retourner à l'instruction 16 pour essayer une nouvelle valeur de *b*, ou conserver la valeur actuelle comme approximation de b(a). L'aiguillage correspondant est commandé par la touche « Variante » du pupitre (*cf.* instruction 27). Si le voyant de cette touche est éteint lorsque la machine est remise en marche, l'exécution du programme se poursuit par l'instruction 28. Celle-ci provoque alors le calcul d'un coefficient M (« multiplicateur »), qui va servir à mettre la répartition calculée « à l'échelle » de l'échantillon (*cf.* instruction 29).

La suite du programme comporte une simplification qui nous a été suggérée par un essai de traitement « manuel » de notre problème; essai au cours duquel nous nous sommes aperçu qu'on pouvait tirer parti de la forme analytique particulière revêtue, en l'occurrence, par l'expression des conceptions calculées pour le mois de rang *n*.

Avec les notations de notre exposé de principe, ces conceptions peuvent en effet s'écrire :

$$G'_n = \int_l^1 k\, x^{a+1} (1 - x)^{b+n}\, dx \tag{29}$$

où *k* désigne une constante ayant pour valeur :

$$k = E_0\, c \tag{30}$$

Or, considérons la fonction

$$k\, x^{a+2} (1 - x)^{b+n} \tag{31}$$

Sa dérivée peut être mise sous la forme

$$(a + 2 + b + n)\, k\, x^{a+1} (1 - x)^{b+n} - (b + n)\, k\, x^{a+1} (1 - x)^{b+n-1} \tag{32}$$

expression dont l'intégration entre *l* et 1 donne

$$(a + 2 + b + n)\, G'_n - (b + n)\, G'_{n-1} \tag{33}$$

On peut donc calculer G'_n à partir de G'_{n-1} par la formule :

$$G'_n = \frac{(b + n)\, G'_{n-1} - k\, l^{a+2} (1 - l)^{b+n}}{a + b + n + 2} \tag{34}$$

C'est ce qui a été fait dans le programme de la figure 5, comme on peut le voir facilement à l'aide de quelques changements de notation. Tout d'abord, faisons $n = i + 1$ dans

la formule 34, et désignons les conceptions calculées par le symbole C, comme dans le programme. Nous obtenons :

$$C_{i+1} = \frac{(b + i + 1)\, C_i - k\, l^{a+2} (1 - l)^{b+i+1}}{a + b + i + 3} \tag{35}$$

Remarquons ensuite qu'on a intérêt à calculer le deuxième terme du numérateur de l'expression précédente, par la formule

$$V_{i+1} = V_i (1 - l) \tag{36}$$

à partir de

$$V_0 = k\, l^{a+2} (1 - l)^b \tag{37}$$

ce qui nous conduit à écrire la formule 35 de la façon suivante :

$$C_{i+1} = \frac{(b + i + 1)\, C_i - V_{i+1}}{a + b + i + 3} \tag{38}$$

Il suffit alors de signaler l'emploi de la notation $P = 1 - l$ (*cf.* instruction 30, compte tenu de la formule 24), pour qu'on reconnaisse dans l'instruction 31, le calcul de V_0 par la formule 37 (k ayant, en l'occurrence, la valeur MN).

Le symbole R, qui apparaît dans l'instruction 32, est utilisé pour désigner l'effectif « restant » en observation (*cf.* instruction 35). L'instruction 33 provoque le rangement des valeurs de Q_i et de F_i en vue de l'impression ultérieure de courbes représentant les variation de ces deux quantités en fonction de i. L'instruction 34 entraîne le calcul de T par cumul, selon la « formule » insérée en début de programme. L'instruction 36 correspond à l'application de la formule 36. Il en résulte que la valeur conférée à W par l'instruction 37, doit être attribuée à C_{i+1} (*cf.* formule 38). Cette attribution s'effectue grâce à l'instruction 40, l'indice i affectant la lettre C ayant été, entre temps, augmenté d'une unité par le jeu de l'instruction 38. L'instruction 41 détermine la valeur de F_i, avant le renvoi en début de boucle commandé par l'instruction 42.

La sortie de boucle s'effectue grâce à l'instruction 39. L'instruction 43 fournit la valeur de C_{15}, ce qui permet de compléter le calcul de T (*cf.* instruction 44). On imprime alors la valeur de χ^2 ainsi obtenue (*cf.* instruction 45). Après quoi, l'instruction 46 provoque l'impression des deux courbes préparées par l'instruction 33.

Ces courbes permettent d'apprécier visuellement si la fécondabilité calculée décroît plus vite ou moins vite que la fécondabilité observée, en fonction de la durée du mariage, et par suite de déterminer le signe de la correction à apporter à a pour améliorer le modèle. A cet égard, l'impression des courbes remplace avantageusement celle du tableau I en début d'exploitation: mais elle risquerait d'être gênante par la suite, si on ne pouvait pas la supprimer.

En fait, on dispose d'un moyen très simple pour supprimer l'impression des courbes quand elle devient inutile, bien que rien ne signale cette possibilité dans le programme. En effet, la grande capacité de la mémoire de la CAB 500, permet d'y conserver le traducteur PAF pendant l'exécution du programme, et par suite de modifier celui-ci en cours d'exploitation. Pour ce faire, l'opérateur n'a qu'à taper l'indication PM (programme modifié) suivie des instructions modifiées du programme. En l'occurrence, il lui suffira donc de taper :

```
PM
33 ALLER EN 34
46 ALLER EN 47
```

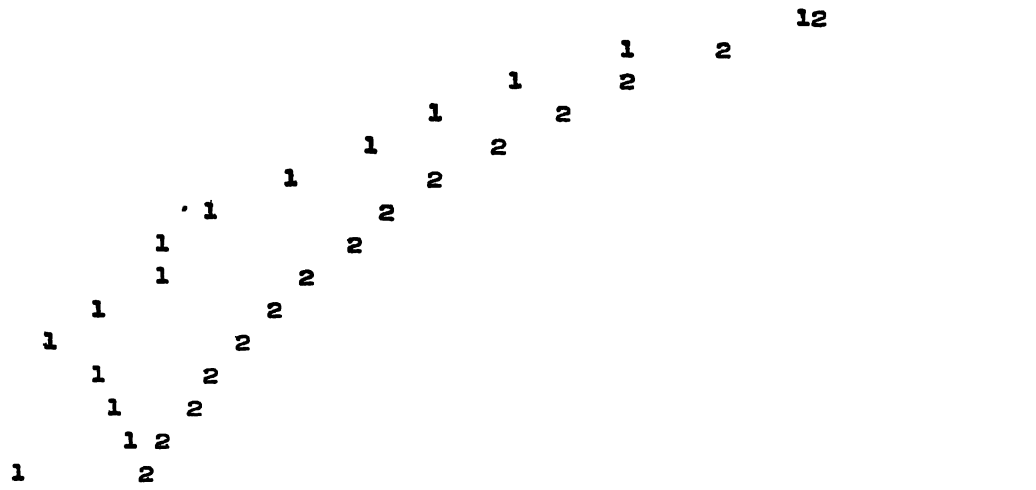
pour supprimer l'impression des courbes.

L=6
A=2 B= 7,90476

B=8 F_o= 0,254189 D= 0,002189

T= 144,102

\+ 0,105599 + 0,254189



A=0,19 B= 2,53222

B=3,26 F_o= 0,251993 D=-0,000007

T= 6,303

\+ 0,105599 + 0,252000

I N E D
047 64

0	1802,0	1802,1	0,2520	0,2520
1	1179,0	1169,6	0,2204	0,2187
2	823,3	825,5	0,1974	0,1975
3	602,3	617,9	0,1800	0,1842
4	456,3	463,9	0,1663	0,1696
5	355,1	354,0	0,1552	0,1558
6	282,5	268,1	0,1461	0,1398
7	228,7	220,1	0,1385	0,1334
8	187,8	190,7	0,1321	0,1334
9	156,2	150,1	0,1266	0,1211
10	131,2	120,9	0,1218	0,1110
11	111,3	116,1	0,1176	0,1199
12	95,1	105,8	0,1138	0,1242
13	81,8	94,7	0,1105	0,1269
14	70,8	68,8	0,1076	0,1056
15	587,6	582,7		

L= 5,5 M= 7,3579 A= 0,1900 B= 3,2600

F_o= 0,2183 S= 1,1668

1	0,02968	21	0,02537	41	0,01112	61	0,00311	81	0,00031
2	0,03276	22	0,02455	42	0,01057	62	0,00287	82	0,00026
3	0,03422	23	0,02374	43	0,01003	63	0,00264	83	0,00022
4	0,03494	24	0,02293	44	0,00951	64	0,00242	84	0,00018
5	0,03523	25	0,02213	45	0,00900	65	0,00221	85	0,00015
6	0,03524	26	0,02134	46	0,00852	66	0,00202	86	0,00012
7	0,03504	27	0,02057	47	0,00805	67	0,00184	87	0,00009
8	0,03470	28	0,01980	48	0,00759	68	0,00167	88	0,00007
9	0,03424	29	0,01904	49	0,00715	69	0,00151	89	0,00005
10	0,03370	30	0,01830	50	0,00673	70	0,00136	90	0,00004
11	0,03308	31	0,01757	51	0,00633	71	0,00122	91	0,00003
12	0,03242	32	0,01685	52	0,00594	72	0,00109	92	0,00002
13	0,03171	33	0,01615	53	0,00556	73	0,00097	93	0,00001
14	0,03097	34	0,01547	54	0,00521	74	0,00086	94	0,00001
15	0,03021	35	0,01480	55	0,00486	75	0,00076	95	0,00000
16	0,02942	36	0,01414	56	0,00453	76	0,00067	96	0,00000
17	0,02862	37	0,01351	57	0,00422	77	0,00058	97	0,00000
18	0,02782	38	0,01289	58	0,00392	78	0,00050	98	0,00000
19	0,02700	39	0,01228	59	0,00364	79	0,00043	99	0,00000
20	0,02618	40	0,01169	60	0,00337	80	0,00037	100	0,00000

L=7

FIGURE 6 — Résultats d'exécution du programme en PAF par CAB 500.

L'impression du tableau I résulte des instructions 50 à 62. L'alignement « à droite » des trois premières colonnes du tableau a nécessité plusieurs instructions (cf. instruction 51 pour la première colonne). En revanche, nous n'avons pas eu à nous soucier des arrondis, les instructions d'impression entraînant l'arrondi automatique de la dernière décimale demandée.

La préparation et l'impression du tableau II font l'objet des instructions 63 à 91. On remarquera que les valeurs de L et de M imprimées en tête du tableau, sont celles des paramètres l et c de la distribution *continue* (*cf.* instructions 77 et 78), l étant exprimé pour cent. On notera d'autre part que la somme de la répartition complète, appelée en l'occurrence S, est impropre au contrôle par totalisation des valeurs de y_i *imprimées* dans le tableau. C'est d'ailleurs pour garantir l'utilisateur contre une telle méprise, que la quantité S est imprimée avec une décimale de moins que les effectifs dont elle représente la somme.

* * *

La figure 6 reproduit, en vraie grandeur, un document établi à l'aide du programme précédent. On y trouve d'abord des résultats obtenus pour L = 6, $a = 2$ et $b = 8$. Il s'agit d'une reprise, sur CAB 500, de l'essai par lequel nous avions prévu de commencer le travail d'exploitation. On remarquera que les valeurs des paramètres a et b avaient été choisies de façon à conférer à la distribution complète une fécondabilité de 0,25 (*cf.* formule 16). La fécondabilité de la distribution tronquée correspondante ressort à 0,254. Elle ne diffère guère de la valeur requise (0,252), et il a paru inutile de chercher une meilleure approximation de b(a) pour un premier essai destiné à orienter le travail. On a donc poursuivi immédiatement l'exécution du programme, ce qui a provoqué l'impression, sous la forme de deux courbes, des résultats attendus.

Les courbes sont pointées à l'aide de petits chiffres : 1 pour la première (Q_1), observée; 2 pour la seconde (F_1), calculée. Elles se lisent par rapport à des axes qu'on disposera mentalement de façon que celui des abscisses, vertical, soit orienté vers le bas, et que celui des ordonnées, horizontal, soit orienté vers la droite. L'échelle des ordonnées est fournie par les valeurs des ordonnées extrêmes, valeurs qui s'impriment automatiquement sur la ligne précédant les courbes. L'imprécision dans la détermination de b(a) se traduit par le fait que, sur ce premier graphique, les deux courbes n'ont pas exactement la même ordonnée à l'origine. Ceci n'empêche pas de constater que la courbe 2 décroît trop lentement, et qu'il convient donc de diminuer a pour améliorer le modèle.

La figure 6 nous offre ensuite, l'ensemble des résultats obtenus en reprenant, sur CAB 500, un essai auquel nous avait conduit l'exploitation par 1620. Cet essai présente l'intérêt de correspondre à la plus faible valeur de χ^2 qu'il soit possible d'obtenir avec les modèles considérés. Cette valeur ressort à 6,3. Le nombre de degrés de liberté se trouve ici réduit à 13, par le fait qu'on a imposé à la répartition de minimiser χ^2. La probabilité correspondante ne s'en situe pas moins entre 0,90 et 0,95.

Pour dégager la signification d'un tel résultat, il convient toutefois de l'insérer dans un contexte. Considérons donc les trois distributions de la figure 1. La distribution M, représentée par des bâtons, est celle qui minimise χ^2. Mais les deux autres, représentées par des courbes, sont tout à fait « possibles ». On obtient en effet $\chi^2 = 9{,}9$ avec la distribution A ($l = 0{,}035$; $a = 0{,}76$; $b = 4{,}446$), et $\chi^2 = 10{,}6$ avec la distribution B ($l = 0{,}075$; $a = -0{,}65$; $b = 1{,}783$), valeurs qui correspondent, avec 14 degrés de liberté, à des probabilités supérieures à 0,70.

C'est dire qu'on peut constituer un large éventail de modèles compatibles avec nos

observations. Et la figure 1 montre clairement que, s'il en est ainsi, c'est avant tout parce que la méthode mise en œuvre, ne permet pas de situer avec précision la troncature d'où résulte l'échantillon.

L'expérience décrite ne nous en a pas moins apporté quelques renseignements précieux, car les distributions « possibles » sont beaucoup plus étroitement apparentées qu'elles ne le paraissent à première vue. C'est ainsi, notamment, que la proportion des éléments à forte ou très forte fécondabilité, est importante dans toutes ces distributions. On peut donc tenir pour acquis qu'il existe de très grandes différences de fécondabilité entre couples, et pour probable que la distribution de la fécondabilité est largement étalée dans la population.

*
* *

Nous conclurons ce compte rendu d'expérience par quelques réflexions sur la rentabilité du calcul électronique. Quiconque s'est penché sur cette question, sait qu'elle n'a de sens que pour une application donnée, ou pour un ensemble d'applications bien définies, la réponse pouvant varier énormément d'une application à l'autre. C'est dire qu'on ne saurait parler de la rentabilité du calcul électronique en matière de recherche, où les problèmes sont perpétuellement changeants, sans disposer d'informations précises sur les problèmes qui se sont effectivement présentés au traitement pendant une longue période, et que toute expérience isolée ne peut apporter, à cet égard, que des éléments d'appréciation tout à fait fragmentaires.

De tels éléments n'en sont pas moins d'autant plus précieux qu'ils sont plus rares, leur collecte nécessitant des mesures spéciales et un travail supplémentaire. C'est pourquoi nous croyons utile d'indiquer brièvement ici, les renseignements essentiels que nous a procurés notre expérience à ce propos.

Comme nous l'avons vu, l'exploitation a été effectuée essentiellement par 1620. C'est donc à ce matériel que se rapportent les résultats dont nous allons faire état. Signalons toutefois que les essais effectués par la suite sur CAB 500, nous ont conduit à la conclusion que le temps de travail global sur machine, nécessité par le traitement complet du problème en cause, aurait sans doute été assez voisin pour les deux matériels.

Voici d'abord quelques temps partiels relevés en cours d'exploitation. On doit compter que chaque valeur de b essayée réclame en moyenne 4 minutes de travail. Pour chaque valeur de a essayée, il faut compter en outre : 2,5 minutes environ pour le calcul de χ^2 par le programme de la figure 3, et 2 minutes supplémentaires pour l'impression (éventuelle) du tableau I. Le calcul et l'impression du tableau II nécessitent environ 4 minutes.

Une expérience de calcul « manuel » nous a donné les temps comparatifs suivants, pour obtenir des résultats aussi précis que ceux fournis par la machine : une douzaine d'heures pour chaque valeur de b essayée; environ 22 heures de plus par valeur de a essayée; 40 minutes pour le calcul du tableau II.

Si coûteuse que puisse être l'heure-machine (on peut se référer, à cet égard, au tarif d'utilisation en « service bureau »), de tels chiffres plaident manifestement en faveur du calcul électronique, même à ne s'en tenir qu'au plan strictement budgétaire, du moment que le volume des calculs est assez important pour que les travaux préparatoires de programmation et de mise au point du programme puissent être amortis par l'exploitation.

Cependant, la comparaison n'est pas aussi décisive qu'il n'y paraît, parce qu'il est hors de doute que, si nous avions dû traiter « manuellement » notre problème, nous aurions cherché, et trouvé, des moyens d'alléger les calculs. N'avons-nous pas, d'ailleurs, établi la formule 34

à l'occasion d'une expérience de calcul manuel? Or la modification correspondante du programme FORTRAN, permet bien de gagner 2 minutes pour chaque valeur de *a* essayée; mais l'usage de la dite formule en calcul manuel, ramène de 22 heures à 70 minutes le temps de travail correspondant...

D'autre part, on n'avait pas besoin de toute la précision fournie par la machine. En outre, la détermination de b(a) a exigé, en moyenne, l'essai de 3 valeurs de *b* en exploitation électronique; alors que, pour des raisons déjà indiquées, le nombre d'essais nécessaires eût certainement été nettement moins élevé en calcul manuel. Enfin, nous sommes convaincu que le nombre des valeurs de *a* essayées, pouvait être lui-même sensiblement réduit, sans que l'intérêt des résultats obtenus en fût considérablement affecté.

Nous touchons là une des raisons essentielles qui risquent de compromettre la rentabilité du calcul électronique : c'est qu'on est toujours tenté d'abuser des facilités qu'il procure. Ce genre d'abus est particulièrement flagrant dans la mise au point des programmes, toujours fort coûteuse en temps-machine. L'expérience montre qu'il est parfaitement possible d'écrire un programme tel que celui de la figure 5, sans commettre aucune erreur; le calcul des prix de revient prouve qu'on a toujours intérêt à s'efforcer d'y parvenir.

Paul VINCENT

DISCUSSION

M. G. THÉODORE. — Le problème posé revient à ajuster à une courbe « expérimentale », la représentation d'une fonction théorique dont par itération successive on recherche les paramètres ou coefficients. La méthode de programmation présentée résulte vraisemblablement de l'utilisation d'ordinateurs ou de calculateurs à programme par cartes perforées.

Dans le cas présent, le volume du travail de programmation serait sensiblement réduit si on pouvait disposer d'une machine analogique du type de celle qui fut construite pour les besoins du Ministère de l'Air, en vue d'ajustement de fonctions théoriques à des trajectoires ou des distributions de pression sur des profils (répartition de portance).