REVUE DE STATISTIQUE APPLIQUÉE

G. D'HERBEMONT
J. L. GROBOILLOT

Du dialogue entre statisticiens et calculatrices électroniques

Revue de statistique appliquée, tome 11, nº 4 (1963), p. 35-48
<http://www.numdam.org/item?id=RSA_1963__11_4_35_0>

DU DIALOGUE ENTRE STATISTICIENS ET CALCULATRICES ÉLECTRONIQUES

G. D'HERBEMONT
Conseiller en Statistique, Cie des Machines BULL

J. L. GROBOILLOT
Statisticien à la Cie des Machines BULL

On parle, depuis quelque temps déjà, de la seconde génération des calculateurs électroniques, la troisième génération est en préparation et le marché des calculateurs continue à s'étendre.

Parallèlement à ce développement, la statistique a trouvé depuis la dernière guerre de nombreux utilisateurs. Cette jeune technique fait appel à des notions modernes de mathématiques ; elle est mise en application par des hommes nécessairement ouverts et réceptifs.

Cependant, contrairement à ce qu'on pourrait penser, le statisticien n'utilise pas toute l'aide que peuvent lui apporter les machines électroniques. Si l'on essaye de rechercher les motifs de réticences, on s'aperçoit que le statisticien aime voir, toucher ses chiffres, et, suivre son calcul. Or le façonnier - c'est-à-dire le centre de calcul -, préfère ne pas faire sortir de la machine trop de résultats intermédiaires. Le statisticien éprouve la nécessité de contrôler, de suivre, de vivre son calcul ; le façonnier ne souhaite pas alourdir son programme par des extractions qu'il considère inutiles.

Par ailleurs, le statisticien hésite parfois devant le temps de réponse à concéder au centre de calcul. Qu'ils fassent l'un et l'autre partie ou non de la même entreprise, il y a toujours la filière : prise de contacts, devis, préparation des données, perforation des données, prise en charge par le service d'exploitation, contrôle des résultats, remise des résultats. S'il est nécessaire d'écrire un programme spécialement pour le problème du statisticien, alors c'est en semaines, certains diront en mois, que le temps de réponse risque, le plus souvent, d'être compté.

La machine n'est pas mise en cause directement car son temps effectif de travail sera de quelques minutes ou même de quelques secondes, mais c'est le langage peu usuel qu'il faut lui tenir qui fait renoncer à bien des projets séduisants.

Ainsi, constatation est faite des difficultés du dialogue entre façonnier et statisticien, et de l'impatience de ce dernier qui désirerait des résultats de calcul dans le même temps qu'une photo-copie.

Il est bien évident que toutes les exigences ainsi exprimées ne peuvent être satisfaites pleinement. Mais si le statisticien et le centre de calcul prenaient davantage l'habitude de travailler en liaison l'un avec l'autre, il est tout aussi clair que bien des difficultés s'aplaniraient.

Le centre de calcul pourrait envisager de consacrer moins de temps à économiser des microsecondes. Il comprendrait mieux pourquoi un statisticien éclairé n'apprécie que très rarement un programme admirablement performant qui n'imprime, paradoxalement, que le OUI ou le NON de la réponse à un test. Il s'adapterait plus volontiers à la manière bien particulière qu'a le statisticien perspicace, d'observer le déroulement de son calcul et de critiquer la façon dont ses données se transforment et se plient à l'analyse qu'il en fait. Du fait des demandes plus nombreuses qui lui seraient présentées, le centre pourrait accroître encore la part consacrée à l'utilisation extensive des machines par rapport à celle qui revient à l'utilisation intensive.

Mais de son côté, le statisticien ferait mieux aussi la part des choses. Il comprendrait pourquoi l'ensemble électronique ne remplace pas la règle à calcul et dans quelle mesure il la complète. Il saurait quel petit problème est mieux résolu avec une machine de bureau, quel petit problème, devenu répétitif soit à cause du nombre de données soit à cause de sa fréquence tombe dans le domaine de la calculatrice. Il saurait enfin quelle langue il convient de parler au centre de calcul pour être aussitôt compris et servi au plus vite.

Que s'établissent des langages, c'est bien de cela qu'il s'agit et c'est cela que nous nous efforcerons de présenter dans les pages qui suivent.

I - LA MACHINE ET SES LANGAGES

Il existe plusieurs moyens de donner des ordres à une calculatrice électronique. Une machine a un langage de base constitué par des codes auxquels correspondent des opérations élémentaires. Elle exécute l'instruction dont le code a été transféré dans son registre de commande. Le code est la clef qui introduite dans la serrure (registre), ouvre une porte et une seule celle de l'instruction élémentaire correspondante au code. Une suite de codes élémentaires constitue un programme.

Il existe, en plus, pour la même machine, des langages plus évolués qui comportent des codes non élémentaires. Ces codes ne seront pas transférés dans le registre de commande de la machine mais traités comme des données et analysées comme telles par un programme dit compilateur. Une suite de codes non élémentaires, ou programme symbolique, donnera naissance après traitement, par le compilateur à une suite de codes élémentaires, ou programme objet. Le programme objet est la traduction en langage de base du programme symbolique, et, en conséquence, il permettra d'exécuter, sur la machine, le programme symbolique (par simple passage du programme objet).

Une machine possède donc plusieurs langages. Il faut remarquer généralement que plus un langage est éloigné du langage de base, plus le travail du compilateur est complexe et moins le programme objet est performant. Cependant, les utilisateurs de machine ont intérêt à disposer de langages s'adaptant à leur pensée et d'un usage agréable. Ils ont même intérêt à posséder un langage universel qui leur permettra des échanges de méthodes et de programmes. C'est pourquoi ils ont crée le langage Algol.

Dans ce qui suit nous distinguerons deux catégories de langages :

1/ le langage de base et des langages voisins souvent nommés autocodes ;

2/ des langages évolués tels que Algol-Fortran-PAF.

Nous ne nous intéresserons pas aux langages Cobol, réservé aux comptables, ni aux langages de liste, Lisp, IPL, Comit, dont les possibilités et les champs d'application ne paraissent pas correspondre aux problèmes statistiques.

II - L'EXECUTION D'UN CALCUL STATISTIQUE

Lorsque le statisticien a un calcul à effectuer il doit commencer par déterminer si son problème n'a pas déjà été l'objet d'un traitement sur une machine déterminée au profit d'autres utilisateurs. Si c'est le cas, il peut bénéficier de l'effort de ses devanciers, sinon il devra programmer ou faire programmer ses calculs.

Le problème a déjà été programmé

Le calcul n'est alors qu'une simple et banale exploitation. En conséquence, il ne doit pas s'écouler plus de huit jours environ entre la prise en charge de la commande et la remise des résultats.

Pour déterminer si son problème a déjà été passé sur une machine, le statisticien devra chercher dans trois catégories de programmes.

a) Les programmes généraux, c'est-à-dire les programmes livrés avec la machine ; ils ont un caractère répétitif et sont généralement programmés en langage de base ou en auto-code. En statistique, ils permettent d'obtenir : moyenne, variance, écart-type, coefficient de corrélation, régression linéaire et non-linéaire.

b) Des programmes qui ont été écrits, par des utilisateurs de la machine, en auto-code ou en langage de base.

c) Des programmes écrits en langages Algol Fortran, ou d'autres langages symboliques (PAF...)

La documentation relative aux programmes existants sur une machine est fournie par le constructeur. Cependant signalons que :

Certaines revues contiennent de courtes notes d'une dizaines de lignes, donnant les caractéristiques des programmes dernièrement sortis sur les diverses machines américaines. C'est le cas, par exemple, de la revue "Technometrics", publiée par "The american society for quality control and the american statistical association" ;

Le laboratoire de statistique du Case Institute of technology edite périodiquement un rapport : "Abstract of statistical Computer routines" ;

La revue "Communication of the A. C. M" (Assocation for computing machinery) publie dans chaque numéro sous la direction de Wegstein un article "Algorithms" contenant essentiellement des procédures écrites en Algol ; on y trouve quelques programmes statistiques. Comme nous le verrons ces programmes écrits en Algol sont directement utilisables.

Le problème n'a jamais été programmé

Pour qu'un calcul soit effectué par une machine, il faut commencer par le définir parfaitement et dans les moindres détails. Il faut préciser ce que seront les données, les résultats, les diverses phases du calcul ; il faut prévoir tous les cas particuliers qui peuvent se présenter. Tout ceci sera consigné dans l'organigramme du problème (voir exemple § IV).

Alors se pose la question du choix du langage à utiliser pour donner les divers ordres à la machine. Deux cas se présentent :

a) Le problème est volumineux (calculs longs, beaucoup de données, beaucoup de résultats intermédiaires à conserver, beaucoup d'extractions). Alors il est quelquefois nécessaire de l'écrire en langage de base ou en autocode. Le statisticien ne peut alors programmer lui-même son problème ; il le présentera sous forme de l'organigramme.

b) Le problème est peu volumineux, ne doit pas être résolu souvent, alors les langages Algol, Fortran s'imposent. La plupart des problèmes statistiques qui ne sont pas des monstres sont de cette nature (exception faite des dépouillements d'enquête et de quelques autres problèmes). Dans ce cas, le statisticien peut programmer lui-même en Algol ou en Fortran, et obtenir, très rapidement des résultats, s'il n'a pas commis trop d'erreur.

Il est bien clair que la limite entre ces deux catégories de problèmes (a et b) dépend d'une part de la machine dont on dispose, d'autre part de ses compilateurs (Programmes traduisant les langages Algol, Fortran... en langage propre à la machine). Il existe diverses variantes de compilateurs selon que :

Le nombre d'instructions du programme symbolique (Algol, Fortran) est ou non limité ;

La segmentation du programme est automatique ou non ;

La gestion des données et des résultats intermédiaires (rangement sur rubans magnétiques) est organisée d'une manière ou d'une autre ;

L'insertion de procédures en langage machine dans le programme symbolique est prévue ou impossible ;

Le compilateur est lent et peut fournir un programme objet en langage machine performant, le compilateur est rapide et restitue un programme objet peu performant.

Les possibilités des langages seront donc surtout fonctions des compilateurs permettant de les utiliser. Les compilateurs sont plus ou moins évolués d'une machine à l'autre et dans une certaine mesure on pourrait aller jusqu'à dire que le langage est peu, alors que le compilateur est tout. Plus généralement, c'est l'ensemble - langage - compilateur - machine - qui permettra de dire si un problème est a écrire dans le langage considéré ou en langage de base.

Trois directions s'offrent donc au statisticien : soit l'utilisation des programmes existants (§ III), soit la présentation d'un organigramme en vue de la programmation en langage de base, qu'il devra confier à un spécialiste (§ IV), soit, enfin l'utilisation d'un langage évolué (§ V).

III - LES BIBLIOTHEQUES DE PROGRAMMES

Il n'est pas possible de donner ici une liste exhaustive des programmes statistiques existant sur les diverses machines fonctionnant en France. Cependant, on peut préciser la richesse normale d'une bibliothèque, qui sera bien garnie si la machine a de nombreux utilisateurs statisticiens.

Statistique à une variable

Calcul des moments et coefficients de Pearson ;

Test du χ^2 d'adéquation d'une variable aléatoire pour diverses lois de probabilités (Laplace gauss, Poisson, Binomiale, Hyper-géométrique, Hyper-exponentielle, Pascal, Gallhier) ;

Test de Kolmogorof (pour les mêmes lois) ;

Test de Smirnof (pour les mêmes lois) ;

Droite de Henry

Courbe de concentration.

Statistique à deux variables

Calcul des coefficients de corrélation ;

Calcul des coefficients de régression linéaire ;

Test sur la signification des coefficients de corrélation ;

Tableau de contingence avec tests divers ;

Comparaison de moyennes et de variances (test F, Sukhatrue, Welch...).

Statistique à plusieurs variables

Régression linéaire ;

Régression linéaire sélective (sur les variables originales ou sur les variables transformées, par exemple Log x) ;

Calcul des coefficients de corrélation partielle ;

Analyse de la variance (plans factoriels) ;

Test sur la moyenne vectorielle (Tester que le vecteur moyen est égal à un vecteur donné, tester l'égalité des vecteurs moyens de deux échantillons, tester l'égalité des composantes du vecteur moyen) ;

Test sur la matrice de covariance (tester l'indépendance de groupes de variables, tester qu'une matrice de covariance est égale à une matrice donnée) ;

Problème de classification (classer un vecteur échantillon dans la famille $N(M_1, \Sigma)$ ou $N(M_2, \Sigma)$) ;

Test sur le vecteur moyen et la matrice de covariance (tester simultanément que le vecteur moyen est égal à un vecteur donné et que la matrice de covariance est égale à une matrice donnée).

Séries chronologiques

Désaisonalisation des séries ;

Prévision par ajustement ;

Calcul de la fonction d'auto-corrélation ;

Divers modèles classiques de prévision.

Dépouillement d'enquête

Il est fortement conseillé de ne pas lancer d'enquête sans s'être préoccupé, au préalable du programme de dépouillement qui l'exploitera.

Analyse factorielle

Méthode centroïde ;

Méthode varimax.

Processus séquentiels

Détermination des fonctions OC et ASN (efficacité et effectif moyen de l'échantillon) pour diverses décisions séquentielles.

IV - L'ORGANIGRAMME D'UN PROBLEME

Nous avons vu qu'il était souhaitable lorsqu'on désire que le programme soit établi par un spécialiste en langage machine ou en autocode, de présenter au centre de calcul un organigramme.

Nous allons montrer de quoi il s'agit sur un exemple très simple.

Supposons que l'on veuille faire calculer la fonction de répartition de la loi de Gauss réduite :

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx$$

pour les valeurs de x échelonnées de 0,5 en 0,5, à partir de x = 0,5 jusqu'à x = 4,5.

Voici (fig. 1) comment se présente l'organigramme d'ensemble de ce calcul. Comme tout organigramme il commence par Début (case 1) et se termine par fin (case 7).

La case 2 indique qu'il faut introduire la valeur 1/2 (valeur initiale de x) dans une certaine mémoire réservée pour les valeurs successives de x et à laquelle on a attribué arbitrairement un nom : k. Ce nom permettra de faire appel, dans d'autres parties de l'organigramme, au contenu k ou de modifier ce contenu.

La case 3 indique qu'il faut calculer F pour la valeur contenue dans la mémoire k, puis ranger le résultat dans une seconde mémoire f.

La case 4 indique qu'on imprime aussitôt x et f, ce qui permet d'utiliser ensuite, à nouveau, k et f pour contenir les valeurs suivantes.

La case 5 donne l'ordre de mettre dans la mémoire k son ancien contenu augmentée de 0,5 (valeur de x suivante).

La case 6 contrôle si la progression de x est terminée, en comparant le contenu de k à 4,5 (dernière valeur de x à considérer) ; si $k \leqslant 4,5$ elle aiguille vers la continuation du calcul et dans le cas contraire vers la fin.

On remarquera que l'organigramme d'ensemble ainsi établi suffit si le centre de calcul dispose déjà d'un programme pour le calcul de F (x). Dans le cas contraire il faut préciser maintenant les opérations résumées par la case 3.

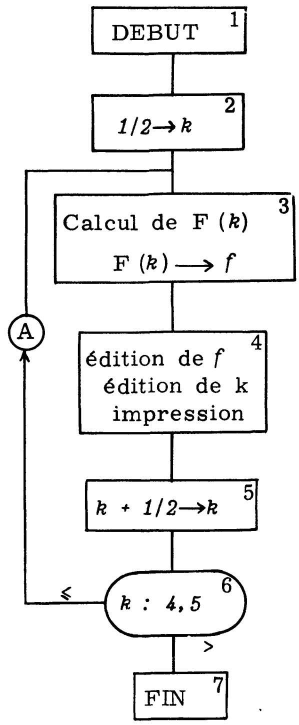

Fig. 1 - Organigramme d'ensemble.

Or, on sait (ou plutôt tout statisticien pourrait savoir) qu'une approximation(1) de F (x) est donnée par :

$$F(x) = R(x)\ \alpha(x) + 0,5 \quad \text{où} \quad \alpha(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

et :

$$R(x) = \cfrac{x}{1 - \cfrac{x^2}{3 + \cfrac{2x^2}{5 - \cfrac{3x^2}{7 + \cfrac{4x^2}{9 - \cfrac{5x^2}{11 + \cfrac{6x^2}{13}}}}}}}$$

c'est-à-dire par les premiers termes d'un développement en fraction continue, limité.

On voit immédiatement apparaître deux suites :

suite A : *-1 ; 2 ; -3 ; 4 ; -5 ; 6 ;*

suite B : *1 ; 3 ; 5 ; 7 ; 9 ; 11 ; 13 ;*

qui correspondent à deux progressions très simples.

----

(1) En fait il serait préférable d'utiliser une "meilleure approximation" polynomiale. La méthode des fractions continues est présentée dans la réf. [6].

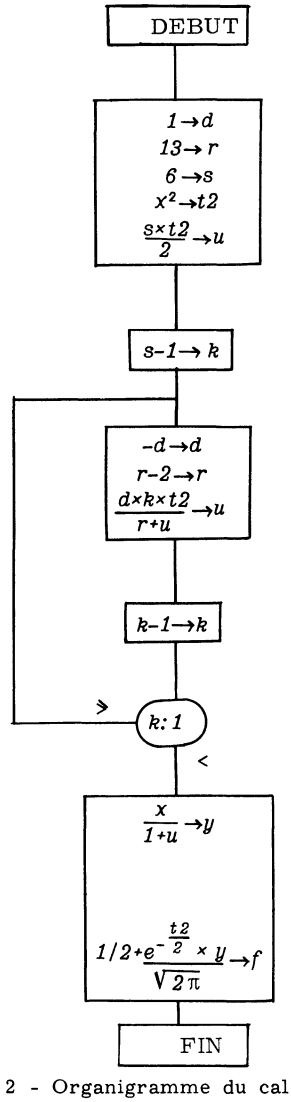

Fig. 2 - Organigramme du calcul de F (x).

En utilisant ce fait, on peut dresser le nouvel organigramme (fig. 2) correspondant au calcul de F (x), pour x donné. Les seules opérations utilisées ici sont la multiplication et l'addition algébrique ainsi que la fonction exponentielle, notées de façon habituelle.

Nous ne décrirons pas en détail cet organigramme que le lecteur pourrait certainement établir lui-même, compte tenu des notations choisies ici :

d : signe du terme courant de la suite A
r : terme courant de la suite B
s : 1er terme de la suite A
$t2$: résultat du calcul du carré de x
u : second terme des dénominateurs successifs
k : progression des termes de la suite A au signe près
y : résultat du calcul de R (x)
f : résultat du calcul de F (x)

Pour grouper maintenant les deux organigrammes il faut se préoccuper d'éviter les confusions de notations. Ainsi, le symbole *k* devra être modifié. Il n'aurait pas été difficile ici de choisir d'emblée une autre lettre, mais ce genre de difficulté se présente automatiquement lorsque l'on désire grouper des organigrammes déjà établis par des voies différentes, extraits d'un autre problème par exemple.

Sur la figure 2, on remarquera le rôle de la mémoire *t2* : elle évite d'avoir à calculer 6 fois, au lieu d'une, le carré de x. Pour la simplicité de l'exposé, ce point mis à part, nous n'avons pas autrement cherché à réduire le nombre d'opérations ni à minimiser le nombre de mémoires. Mais il faut évidemment avoir en vue ces préoccupations qui seront éminemment celle du centre de calcul.

Bien que l'établissement d'un organigramme ne soit pas toujours aussi simple que dans le cas présenté, le travail de réflexion nécessaire pour le tracer, en s'appuyant sur les quelques indications que nous venons de donner, est, on en conviendra certainement, à la portée de tout statisticien. En possession de ce document, le centre de calcul peut alors confier la programmation à son personnel spécialisé.

Mais nous allons voir maintenant, sur le même exemple, qu'il n'est guère plus difficile de décrire les opérations en un langage directement utilisable par un compilateur.

V - UTILISATION DES LANGAGES EN STATISTIQUE - L'ALGOL

On peut dire qu'il existe à l'heure actuelle deux langages mathématiques, Algol et Fortran, qui permettent très facilement d'exprimer rationnellement les calculs statistiques qu'on se propose d'effectuer.

Ainsi, un statisticien qui désire résoudre un problème ne mettra pas plus de temps à écrire en Algol le calcul à exécuter qu'à expliquer à un calculateur humain ce qu'il doit faire à la machine de bureau.

C'est ce que nous allons chercher maintenant à mettre en évidence, en espérant que le lecteur sera tenté, par notre propos, de se référer à l'un des cours d'Algol cité en fin de cet article(1) : il constatera alors par lui-même qu'il ne faut pas plus d'une dizaine d'heures pour apprendre ce langage.

Dans le cas de l'exemple donné au cours du paragraphe précédent, le texte écrit en Algol, à donner au centre de calcul, pourrait être celui de la figure 3. Toutefois les nombres écrits en marge ne servent qu'à repérer les lignes et ne font pas partie du programme.

En fait, les lignes 1, 2 et 22 à 30 correspondent au processus décrit dans l'organigramme général (fig. 1), tandis que les lignes 3 à 21 correspondent au processus de la figure 2. On peut estimer d'ailleurs que, l'habitude aidant, le tracé préalable de l'organigramme n'est pas nécessaire et l'écriture en Algol est alors obtenue directement.

Avant de commenter le programme proposé, nous présenterons quelques remarques sur les symboles de l'Algol. Ceux-ci sont strictement définis. Ainsi, pour l'addition et pour la multiplication on utilise les signes usuels (+ et ×) ; pour la division on utilise le signe : / ; pour l'élévation à une puissance (x^2 par exemple) on utilise une flèche verti-

----

(1) réf. [7] à [11]

cale : *x*↑*2*. La fonction exponentielle est notée ici par *EXP(x)* pour e^x et la racine carrée de x par *RAC2(x)*. On notera que la virgule, le point et virgule, les parenthèses sont nécessaires aux endroits indiqués. Quant au signe *:=*, il signifie que la quantité désignée à sa droite doit être introduite dans la mémoire nommée à sa gauche (*k:=3×s* signifie donc que le triple du contenu de la mémoire *s* doit être introduit dans la mémoire *k*).

A - Organisation générale du programme

Comme tout programme Algol, celui-ci commence par **début**(ligne 1) et se termine par **fin** ; (ligne 30).

Le premier **début** est suivi des déclarations d'identificateurs ; ici il n'y a qu'une telle déclaration :

```
2    Réel k, f ;
```

qui indique à la machine la réservation de mémoires à effectuer pour les deux variables réelles *k* et *f*.

Ensuite, viennent les déclarations de procédures utilisées dans le programme. Ces procédures sont des parties de programme auxquelles on peut se renvoyer. Ici, il n'y a qu'une procédure, celle qui se rapporte à la méthode de calcul de f = F (x) pour une valeur x quelconque :

```
3  | Procédure Probabilité (x, f) ;
4  | Début
5  |     Entier r, s, k ;
:  | .................
21 | Fin Probabilité ;
```

La déclaration de procédure comporte :

a) L'identification de la procédure

(On a choisi ici le mot *Probabilité*) suivi des paramètres de la procédure (soit ici *(x, f)*). Ces paramètres sont des identificateurs non locaux, c'est-à-dire qu'ils sont à préciser, comme nous le verrons plus loin, lors de l'appel de la procédure dans le programme.

b) Le corps de procédure

```
4  | Début
5  |     Entier r, s, k ;
:  | .................
21 | Fin Probabilité
```

où on constate que *k*, notamment, est déclaré comme identificateur local, entier ; il est donc sans rapport avec le *k* de la ligne 2. Le contenu détaillé du corps de la procédure sera commenté en B.

Les déclarations précédentes sont suivies du programme à proprement parler :

```
22 | Pour k : = 1/2 Pas 1/2 Jusqu'à 4,5
23 | Faire
24 | Début
25 |      Probabilité (k, f) ;
26 |      Editer ("F 19.9", f) ;
27 |      Editer ("F 19.9", k) ;
28 |      Imprimer (1) ;
29 | Fin ;
```

Celui-ci se compose d'une instruction **pour** portant sur quatre instructions (lignes 25 à 28) précédées de **début** et suivies de **fin** ; (cet encadrement par **début**, **fin** ; précise la portée de l'instruction **pour**).

L'instruction **pour** dit d'exécuter plusieurs fois les quatre instructions "pour" des valeurs de k successives qui progressent à partir de $1/2$ jusqu'à $4,5$, par saut (**pas**) de $1/2$. Ceci est donc une façon conventionnelle de dire comment on désire faire progresser le contenu de la mémoire k.

Considérons maintenant l'instruction de la ligne 25. Celle-ci fait appel à la procédure *probabilité* définie parmi les déclarations et tout se passe pratiquement comme si on remplaçait dans le programme, la ligne 25 par le corps de la procédure correspondante dans laquelle on aurait préalablement remplacé k par $k1$ et x par k.

Quant aux trois instructions 26, 27, 28, elles permettent d'extraire les résultats de calcul soit :

26 *Editer ("F19.9", f) ;*

Edition (préparation de l'impression), sur dix neuf caractères avec neuf chiffres après la virgule, du contenu de la mémoire f (qui contient alors la valeur de F (k)).

27 *Editer ("19.9", k) ;*

Edition de l'indice k ;

28 *Imprimer (1) ;*

Saut de papier d'une ligne et impression de la ligne éditée. Ces procédures d'édition ne sont pas normalisées et ne font pas partie du langage Algol. Ce langage, en effet, ne définit pas les entrées sorties.

B - Le corps de procédure

Le corps de procédure est ici constitué en "Bloc" : il commence par **début** (ligne 4), suivi des déclarations d'identificateurs locaux à la procédure :

5 **Entier** *r, s, k ;*
6 **Réel** *d, t2, u, y ;*

et il est terminé par **fin** ;

On peut commenter ainsi les instructions de ce corps de procédure :

Commentaire	Programme
on affecte à d la valeur 1	$d := 1$;
" " " r " " 13	$r := 13$;
" " " s " " 6	$s := 6$;
" " " $t2$ " " x^2	$t2 := x \uparrow 2$;
" " " u " " $\frac{s \times t2}{r}$	$u := s \times t2/r$;

On a remisé dans u la valeur $\frac{6x^2}{13}$, dernier terme du développement tronqué en fraction continue $\left(\text{le prochain terme à calculer est donc } \frac{-5x^2}{11+u}\right)$.

k régresse à partir de 5 jusqu'à la valeur 1 par pas de -1.	**Pour** $k := s-1$ **Pas** -1 **Jusqu'à** 1 **Faire**

On voit que k va générer la suite A en valeur absolue.

L'instruction **pour** porte sur les trois instructions suivantes.	**Début**
L'alternance du signe de la suite A sera obtenu par *d*.	*d:=-d ;*
La suite B sera générée par *r* qui régresse à partir de *13* par pas de *-2* jusqu'à *3*.	*r:=r-2 ;*
Cette instruction permet de calculer pour $k=5$ $\frac{-5x^2}{11+u}$ remisé en *a*; pour $k=4$ $\frac{4x^2}{9+4}$ remisé en *a*; etc.; jusque $k=1$ $\frac{-x^2}{3+u}$ remisé en *a*	*u:=d×k×t2/(r+u) ;* **Fin ;**
Il reste pour le calcul de $\overline{R}(x)$ à effectuer $\frac{x}{1+u}$ remisé en *y*	*y:=x/(1+u) ;*
La valeur de $F(x) = \overline{R}(x)\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} + 0,5$ est remisée dans *f*	*f:=0,5+EXP(-t2/2)×y/RAC2 (2×3.1415) ;*
	Fin *Probabilité ;*

Si la machine dont on dispose, possède une entrée par cartes perforées, le programme (fig. 3) devra être mis sur cartes. A chaque symbole correspond un code de perforation variant souvent d'une machine à l'autre. En général on perfore un symbole par colonne et une carte par ligne de programme.

C - Remarques générales sur la programmation en Algol

Le temps d'écriture du programme en langage machine est au moins dix fois supérieur au temps d'écriture en Algol. Lorsque le programme est écrit, il contient souvent des erreurs, il faut donc faire des essais de mise au point du programme sur la machine. Le nombre d'essais de mise au point d'un programme écrit en langage machine est au moins dix fois supérieur à celui du programme écrit en Algol.

En Algol, la détection des erreurs est alors neuf cas sur dix immédiate, la machine spécifie elle-même la nature de l'erreur faite dans l'instruction dont elle indique le numéro. Plusieurs erreurs peuvent être détectées par un même essai.

Il est recommandé de découper l'ensemble du calcul en de nombreuses procédures élémentaires, qui pourront être mises au point indépendemment les unes des autres et simultanément. L'assemblage de l'ensemble des procédures élémentaires ne sera plus alors qu'une formalité.

Des revues publient d'ailleurs régulièrement des procédures Algol permettant de résoudre certains problèmes. Il suffit de les recopier en respectant les contraintes d'écriture de la machine dont on dispose (par exemple sur le $\Gamma 60$, le point virgule (;) est codifié ◊) avant de les donner à la perforation. Ces procédures sont alors directement utilisables.

VI - L'AVENIR ET LE PRESENT

Les langages naissent et se développent (Algol, Cobol, LISP, COMIT, IPL...). Ce sont des langages généraux pouvant être utiles à la statistique. Dans les années à venir, cependant, à l'intérieur même de la technique statistique, doivent voir le jour de nombreux langages tels que :

Langage de simulation ;

Langage de traitement des plans d'expériences ;

Langage de traitement des enquêtes et recensement.

Il serait souhaitable que dans ces domaines comme cela s'est produit pour l'Algol, la nécessité se fasse sentir de coordonner et de normaliser les actions.

On peut estimer aussi qu'une utilisation plus fréquente des machines, aura des répercussions profondes sur les méthodes de travail. Ainsi, pourquoi s'encombrer de volumes de tables plus ou moins complètes alors qu'il est beaucoup plus simple de connaître les procédures d'approximation des fonctions rencontrées en statistique. Une grande partie du temps consacrée à décrire comment utiliser les tables et à disposer les calculs d'un χ^2 ou d'un coefficient de corrélation, ne serait-elle pas avantageusement concédée à l'exposé des méthodes d'approximation et des méthodes de programmation ?

Peut-être verrait-on alors, un peu plus souvent, pris en compte le risque de deuxième espèce dans les raisonnements statistiques et ne serait-on plus aussi hésitant lorsqu'interviennent les probabilités attachées aux variables χ^2, t, ou F décentrées.

Mais, dès à présent, un certain nombre de problèmes se posent, parmi lesquels on peut noter les suivants :

Faire connaître les divers programmes statistiques existants sur les différentes machines fonctionnant en France ;

Elaborer des langages statistiques ;

Promouvoir l'usage de méthodes nouvelles ou de méthodes anciennes peu utilisées de calculs statistiques ;

Coordonner et stimuler la publication en France des calculs statistiques écrits en Algol.

Poser les problèmes statistiques rencontrés lors de l'utilisation des calculateurs digitaux dans le domaine de l'automatisme industriel (relevé d'information par sondage, décisions séquentielles...).

Tout ceci laisse penser qu'il reste beaucoup à faire dans la mise au point et la vulgarisation de méthodes de calcul statistiques.

BIBLIOGRAPHIE

[1] F. YATES - (1962) "Computers in Research - promise and performance". The Computer journal. Vol. 4, page 273.

[2] F. YATES et M.R. SIMPSON - (1960) "A general program for the analysis of surveys". The Computer journal. Vol. 3, p. 136.

[3] F. YATES, J.C. GOWER et M.R. SIMPSON - "A specialized autocode for the analysis of replicated experiments". The Computer journal. Vol. 5, p. 313.

[4] F.C. LEONE - "Statistical programs for high speed Computers" dans la plupart des numéros de Technometrics.

[5] K.D. TOCHER et D.G. OWER - "The automatic programming of simulations", compte rendu du congrès internationnal de recherche opérationnelle 1960.

[6] M.G. KENDALL - "The Advanced theory of statistics" (cf. Griffin).

[7] M. BOTTEN BUCH - "Structure and use of Algol 60", journal of the A.C.M. Vol. 9, n° 2, avril 1967.

[8] M. CRACHEN - "A guide to Algol programming". J. Wiley and sons, N.V. 1962.

[9] M.R. SCHWARZ - "An introduction to Algol 60". Communication of the A.C.M., n°5 (fev. 62), page 89 à 95.

[10] E.W. DIGKSTRA - "Primer of Algol 60 programming". Academic Press 1962.

[11] N. GASTINEL et L. BOLLIET - "Cours d'Algol de l'université de Grenoble". (à paraître chez Hermann).