

**REVUE BELGE DE STATISTIQUE
ET DE RECHERCHE OPERATIONNELLE**

**Vol. 7 - N° 3 et 4
MAI 1967**

**BELGISCH TIJDSCHRIFT VOOR STATISTIEK
EN OPERATIONEEL ONDERZOEK**

**Vol. 7 - N° 3 en 4
MEI 1967**

La « Revue Belge de Statistique et de Recherche Opérationnelle » est publiée avec l'appui du Ministère de l'Education nationale et de la Culture, par les Sociétés suivantes :

SOGESCI. — Société Belge pour l'Application des Méthodes scientifiques de Gestion.

Secrétariat : 66, rue de Neufchâtel, Bruxelles 6. Tél. 37.19.76.

S.B.S. — Société Belge de Statistique. Siège social : 44, rue de Louvain, Bruxelles.

Secrétariat : 44, rue de Louvain, Bruxelles.

Comité de Direction

E. DE GRANDE, Docteur en Sciences, Theopliel Reynlaan, 53, Mortsel.

S. MORNARD, Licencié en Sciences, rue Souveraine, 51, Bruxelles 5.

R. SNEYERS, Docteur en Sciences, Météorologiste adjoint à l'Institut Royal Météorologique de Belgique, 68, rue Copernic, Bruxelles 18.

Comité de Screening

A. HEYVAERT, Ingénieur civil, 3, Val-Fleuri, Dilbeek.

R. SNEYERS, Docteur en Sciences, Météorologiste adjoint à l'Institut Royal Météorologique de Belgique, 68, rue Copernic, Bruxelles 18.

Rédaction

R. SNEYERS, Docteur en Sciences, Météorologiste adjoint à l'Institut Royal Météorologique de Belgique, 68, rue Copernic, Bruxelles 18.

Secrétariat

J.H. LENTZEN, 66, rue de Neufchâtel, Bruxelles 6 - Tél. 37.19.76.

Het « Belgisch Tijdschrift voor Statistiek en Operationeel Onderzoek » wordt uitgegeven met de steun van het Ministerie van Nationale Opvoeding en Cultuur, door de volgende Verenigingen :

SOGESCI. — Belgische Vereniging voor Toepassing van Wetenschappelijke Methodes in het Bedrijfsbeheer.

Secretariaat : Neufchâtelstraat 66, Brussel 6. Tel. 37.19.76.

S.B.S. — Belgische Vereniging voor Statistiek.

Maatschappelijke zetel : 44, Leuvensestraat, Brussel.

Secretariaat : 44, Leuvensestraat, Brussel.

Directie Comité

E. DE GRANDE, Dr in de Wetenschappen, Theopliel Reynlaan, 53, Mortsel.

S. MORNARD, Lic. in de Wetenschappen, Souverainestraat, 51, Brussel 5.

R. SNEYERS, Dr in de Wetenschappen, Adjunct-Meteoroloog bij het Koninklijk Meteorologisch Instituut van België, Copernicusstraat, 68, Brussel 18.

Screening Comité

A. HEYVAERT, Burgerlijk Ingenieur, Bloemendal, 3, Dilbeek.

R. SNEYERS, Dr in de Wetenschappen, Adjunct-Meteoroloog bij het Koninklijk Meteorologisch Instituut van België, Copernicusstraat, 68, Brussel 18.

Redactie

R. SNEYERS, Dr in de Wetenschappen, Adjunct-Meteoroloog bij het Koninklijk Meteorologisch Instituut van België, Copernicusstraat, 68, Brussel 18.

Secretariaat

J.H. LENTZEN, 66 Neufchâtelstraat, Brussel 6 - Tel. 37.19.76.

REVUE BELGE DE STATISTIQUE ET DE RECHERCHE OPERATIONNELLE

VOL. 7 - N^{os} 3 et 4 - MAI 1967

VOL. 7 - N^{rs} 3 en 4 - MEI 1967

SOMMAIRE — INHOUD

R. DEREYMAEKER. — Voorwoord - Avant-propos	3
J. TEGHEM. — Les phénomènes d'attente	5
L.K. ENGELS. — De mogelijkheden van het statistisch onderzoek in de linguïstiek	19
R. DEREYMAEKER. — Huidige stand van de regionale statistiek .	29
E. NOLS. — Réflexions en vue d'une ventilation régionale des statistiques	45
J.F. van WATERSCHOOT. — Regionaal-economische modellen en de eisen die ze stellen aan de statistiek	53
E. MEULEPAS. — Discriminantanalyse bij het opsporen van pa- tienten met Turner-syndroom	58
J.V. JOOSSENS. — Het verkerken van kankersterftestatistieken .	65
R. COUTREZ. — Applications astronomiques et astrophysiques de la statistique	74
Publications reçues — Ontvangen publicaties	87
Demande d'emploi	88
Ouvrage reçu — Ontvangen boek	88

BELGISCH TIJDSCHRIFT VOOR STATISTIEK
EN OPERATIONEEL ONDERZOEK

VOORWOORD

Toen in 1937 de « Société Belge de Statistique » werd opgericht schreef zij in haar statuten de volgende doelstellingen (cfr. Staatsblad van 27 maart 1937) :

- 1° Contribuer au progrès scientifique de la statistique et promouvoir l'esprit d'entente et de collaboration parmi les statisticiens belges.
- 2° Orienter l'opinion publique vers une conception plus exacte de la place qui revient à la statistique dans la vie moderne.
- 3° Prendre connaissance des travaux de ses membres et les soumettre à la discussion.
- 4° Publier éventuellement des ouvrages, études et brochures se rapportant aux matières entrant dans son objet social.

Deze oprichting luidde een periode van hernieuwde belangstelling voor de statistiek in. Reeds honderd jaar vroeger evenwel genoot België dank zij A. Quetelet, de stichter van de internationale statistische congressen, een grote faam op statistisch gebied. Op het binnenlands vlak richtte Quetelet de Centrale Commissie voor de Statistiek op. In dezelfde periode werden in het buitenland statistische genootschappen opgericht : de Londense « Royal Statistical Society » in 1833, de « American Statistical Association » in 1839, de « Société Statistique de Paris » in 1860.

Genoemde doelstellingen zijn in de voorbije dertig jaren niet gewijzigd. Een groot aantal voordrachten, gaande van de theoretische statistiek tot diverse toepassingsgebieden werden georga-

AVANT-PROPOS

Lors de sa création en 1937, la « Société belge de Statistique » inclut dans ses statuts les buts suivants :

Cette création consacrait le renouveau d'intérêt qui se manifestait pour la statistique. On se souviendra cependant que cent ans plus tôt, grâce à A. Quetelet, fondateur des congrès internationaux de statistique, la Belgique jouissait déjà d'une grande renommée dans le domaine de la statistique. Dans le pays, Quetelet avait fondé la Commission Centrale de Statistique, tandis qu'à l'étranger naissaient la « Royal Statistical Society » de Londres en 1833, l'« American Statistical Association » en 1839 et la « Société Statistique de Paris » en 1860.

Les buts qu'on vient de rappeler n'ont pas changé depuis trente ans. De nombreuses conférences couvrant aussi bien la statistique théorique que le champ étendu de ses applications furent organisées. Elles furent repro-

niseerd. Van 1951 tot 1960 werden de teksten hiervan gepubliceerd in offset-druk en sedertdien in het huidige Belgische Tijdschrift voor Statistiek en Operationeel Onderzoek.

Ter verwezenlijking van hetzelfde doel werd hieraan voor de tweede maal de inrichting toegevoegd van een Statistische Dag. De referaten gehouden op de Eerste Statistische Dag, welke doorging op 1 juni 1965, vormen het januari-nummer 1966 van dit tijdschrift. Het verheugt ons thans ook de teksten te kunnen aanbieden van de voordrachten van de Tweede Statistische Dag, die plaats had op 18 mei 1966.

Namens de Belgische Vereniging voor de Statistiek danken we de eminente sprekers, welke aan deze Statistische Dagen — waaraan gemiddeld ruim 200 personen deelnamen — hun medewerking hebben willen verlenen, evenals de HH. Picard en Davin, professoren aan de Universiteit te Gent en te Luik, Dr. Sneyers, ondervoorzitter van de vereniging, en de H. Bodart van het Rijksuniversitair Centrum te Mons, voor hun bevoegde leiding van bepaalde sectievergaderingen.

duites en offset de 1952 à 1960 et dans la Revue belge de Statistique et de Recherche opérationnelle au cours des dernières années.

Cette journée statistique est la seconde qui a été organisée en vue d'atteindre l'objectif de notre société. Les textes présentés au cours de la première journée statistique ont paru dans le numéro de janvier 1966 de notre revue.

C'est avec une grande satisfaction que nous présentons cette fois les textes des communications qui ont été faites à la deuxième journée statistique qui eut lieu le 18 mai 1966.

Au nom de la société belge de Statistique, nous remercions les éminents orateurs qui ont accepté de collaborer à ces journées qui furent suivies par quelques deux cents personnes.

Nous remercions également MM. Picard et Davin, respectivement professeurs aux universités de Gand et de Liège, M. Sneyers, vice-président de notre Société et M. Bodart du Centre Universitaire de l'Etat à Mons pour la compétence avec laquelle ils ont bien voulu présider les réunions des sections spécialisées.

De Voorzitter - Le Président,
R. Dereymaeker.

LES PHENOMENES D'ATTENTE

par J. TEGHEM

Université Libre de Bruxelles

Résumé.

Cet exposé est destiné essentiellement à des auditeurs non initiés aux méthodes analytiques de traitement des problèmes de files d'attente. Il comprend, après des généralités sur ces problèmes et la description de modèles mathématiques simples, un aperçu de la méthode classique du bilan des probabilités et des problèmes qu'elle permet de résoudre, soit de manière précise — modèles markoviens — soit par approximation — modèles non markoviens.

1. Généralités. Modèles et problèmes.

Les phénomènes d'attente constituent l'un des chapitres les plus importants dans l'application de la théorie des probabilités, et plus précisément dans l'application de la théorie des processus aléatoires.

La bibliographie du livre de Saaty, *Elements of Queueing Theory*, qui a paru en 1961 — et elle n'est pas exhaustive — comporte déjà plus de 900 références à des mémoires et articles sur le sujet, et depuis 1961, plusieurs centaines d'articles et mémoires dignes d'être cités ont encore été publiés.

Des phénomènes d'attente méritant d'être étudiés se rencontrent dans tous les domaines de l'activité humaine. Il y a phénomène d'attente et formation éventuelle d'une file d'attente lorsque des bateaux attendent, à quai, d'être déchargés, lorsque des avions attendent l'autorisation d'atterrir, lorsque des commandes attendent d'être exécutées, du courrier d'être dactylographié, des clients d'être servis, lorsque des voitures attendent d'être contrôlées à la douane ou attendent simplement qu'un feu de signalisation devienne vert, lorsque des ouvriers se présentent au magasin d'outillage, lorsque des machines en panne sont amenées à l'atelier de réparation, lorsque des appels téléphoniques attendent qu'une ligne devienne libre, lorsque des messages attendent d'être décodés, etc.

Ces phénomènes relèvent de la théorie des processus stochastiques du fait du caractère aléatoire de plusieurs de leurs éléments, qu'il s'agisse, par exemple, des instants d'arrivée des véhicules ou des instants de survenance des pannes ou des durées de service, des durées de réparation ou de conversation, ou encore des délais de réapprovisionnement en pièces d'un stock de marchandises.

Mais on voit aussi que l'évolution de ces phénomènes et dès lors leurs conséquences peuvent être influencées partiellement par des décisions humaines, décisions portant, selon le cas, sur le nombre de pistes d'atterrissage, sur le nombre de lignes, sur le nombre de magasiniers ou d'ouvriers chargés des réparations, et sur leur qualification, ou portant sur les modes de réapprovisionnement, sur la capacité d'une salle d'attente, sur le classement des demandes de clients par ordre de priorité, sur le mode d'attribution des tâches à des serveurs différents, etc., sur la structure, en un mot, du système d'attente et de service.

Et je n'ai pas besoin, je suppose, de vous démontrer l'intérêt de l'étude des conséquences de ces diverses décisions possibles, conséquences qui s'inscrivent dans un contexte, selon le cas, social, économique ou militaire. Et il n'y a donc pas lieu de s'étonner de la profusion d'études consacrées au sujet. L'an dernier, l'O.T.A.N. a consacré spécialement aux files d'attente une grande conférence internationale d'une semaine, conférence qui a réuni les experts des pays membres de l'organisation.

Voyons maintenant quels sont les points communs à tous les systèmes d'attente et de service; voyons aussi par quoi ils peuvent se différencier, et tâchons d'en dégager une structure assez générale.

Dans tous les modèles de phénomènes d'attente, il y a un centre d'attente et de service — le magasin d'outillage, par exemple, ou l'atelier de réparation, ou un central téléphonique, ou un grand magasin à rayons multiples, etc. Il y arrive des unités, provenant d'une ou de plusieurs sources — ces unités sont respectivement, dans les exemples que je viens de citer, des ouvriers ou des machines ou des appels téléphoniques ou les clients du grand magasin — lesquelles unités forment, si elles ne peuvent être servies immédiatement, une ou plusieurs files d'attente. Quand elles ont été servies, ces unités quittent le centre. Pour unités, on dit souvent « clients », employant ainsi dans le contexte général une terminologie empruntée à un type particulier d'application. Le centre de service proprement dit peut comprendre une ou des stations de service — on dit souvent « guichets » — lesquels peuvent travailler en parallèle, comme dans un bureau de poste, par exemple,

les clients se présentant à l'un ou l'autre des guichets, ou en série, chaque client passant successivement d'un guichet à un autre — songez par exemple à la production en usine de pièces devant être traitées successivement par différentes machines ou divers ateliers. On peut aussi combiner les deux modes d'association : guichets en parallèle et en séries. Il se peut enfin que les clients ne passent que par un certain nombre de guichets, à leur choix, ou que l'ordre des passages soit quelconque : songez aux divers rayons d'un grand magasin, etc. Cette diversité de dispositions quant aux guichets a pour résultat une grande diversité de possibilités pour les files. Celles-ci peuvent être disposées en parallèle et en série, en alternance alors avec les stations. Mais il faut bien d'autres renseignements pour qu'un modèle d'attente soit complètement défini; il faut connaître :

- la capacité de la ou des sources. Exemple : s'il y a m machines susceptibles de tomber en panne, cette capacité vaut m . Dans le cas d'un grand magasin ou d'un central téléphonique, la capacité sera ou pourra être infinie;
- la distribution de probabilité des arrivées. Exemple : les arrivées obéissent à un processus de Poisson homogène de paramètre λ , ce qui signifie que la probabilité de a arrivées au cours d'un intervalle de temps t est
$$e^{-\lambda t} \frac{(\lambda t)^a}{a!};$$
- la capacité de la ou des files : elle peut être finie — voire nulle en cas d'absence d'une « salle d'attente » — ou pratiquement infinie. Si elle est finie, égale à l , les clients qui se présentent lorsque la file comprend déjà l unités quittent le système sans être servis;
- la discipline de file ou ordre dans lequel sont servis les clients. Exemple : premier arrivé-premier servi ou dernier arrivé-premier servi, ou existence de priorités, et il y a bien des types de priorités possibles;
- la distribution de probabilité des durées de service. Exemple : les services ont une distribution gamma ou — cas particulier — une distribution exponentielle négative.

Toutes sortes d'autres données peuvent encore entrer en ligne de compte : des impatiences de clients, qui renoncent à se joindre à une file qu'ils jugent trop longue ou qui la quittent après avoir attendu un certain temps, des phénomènes de blocage, des pièces « bloquant », par exemple, le fonctionnement d'un atelier, c'est-à-dire y empêchant par leur présence le traitement

d'autres pièces, parce que l'atelier suivant n'a pas terminé sa tâche ou que la file éventuelle entre deux ateliers est saturée, etc.

Il se peut enfin que les clients arrivent par groupes plutôt qu'un par un, que les services se fassent par groupes, soit à effectif fixé, soit jusqu'à concurrence d'un maximum fixé — exemple : un ascenseur — et toutes sortes de variantes sont encore possibles.

Voilà pour les données constitutives du modèle. Venons-en maintenant à ce qu'on demande. Il s'agit de rendre compte d'une manière ou d'une autre de l'efficacité du système d'attente et de service. La plus ou moins grande efficacité peut s'exprimer, par exemple, dans la distribution de probabilité de la longueur de la file, ou par la distribution du temps d'attente des clients, ou par la distribution des périodes ininterrompues d'occupation des guichets, ces renseignements n'étant pas, bien entendu, sans liens entre eux, lesquels liens peuvent être exploités pour passer d'un renseignement à un autre.

On peut chercher à obtenir ces distributions en régime transitoire, c'est-à-dire en fonction du temps, ou se contenter — c'est plus simple — de les obtenir en régime-limite, pour t , le temps, $\rightarrow \infty$. Ceci suffit, en pratique, lorsqu'il s'agit de systèmes qui évoluent rapidement vers un régime d'équilibre statistique (ou régime permanent). On peut aussi, plus modestement, en présence de trop grandes difficultés, ne chercher qu'à déterminer certains paramètres de ces distributions, principalement la moyenne.

Quelquefois, on considère une fonction économique, dépendant de coûts, tels que le coût d'attente d'un client, le coût d'inactivité des serveurs, le coût de l'installation éventuelle d'un serveur supplémentaire ou d'une organisation tendant à accroître la vitesse des services, etc. Et il convient alors de minimiser cette fonction économique, ce qui conduit à des résultats tels que nombre optimal de serveurs ou vitesse optimale d'arrivée des unités, ou autres éléments optimaux encore.

D'autres fois, il n'est pas question de coûts, mais simplement de voir comment on peut, avec les moyens dont on dispose, par un régime de priorité adéquate, répondre le mieux à un volume donné de demandes de service. Imaginez, par exemple, une machine qui ait à satisfaire à deux types de demandes, et que le passage d'un type de service à un autre nécessite un temps de réglage, un « switch-over time ». Vous comprendrez que s'il y a beaucoup d'alternance dans les arrivées des deux types de clients, la discipline premier arrivé-premier servi — c'est-à-dire services dans l'ordre des

arrivées — risque, en cas de réglages répétés, d'être très mauvaise. Et qu'on peut espérer améliorer le rendement par l'introduction d'une priorité convenablement choisie et adaptée aux données du problème.

J'ai dit ainsi successivement de quoi dépendent les modèles d'attente et quelle est la nature générale des problèmes qu'on peut se poser. Ces problèmes sont nombreux et divers, et la plupart sont difficiles, et vous pensez bien qu'ils ont suscité la création de nombreuses méthodes. Comme vous pensez bien aussi qu'il n'est pas possible de passer ici toutes ces méthodes en revue.

2. Méthode du bilan des probabilités et modèles poissonniens.

Je me propose essentiellement, dans la suite de mon exposé, de dire quelques mots d'une méthode appelée méthode du bilan des probabilités, par laquelle on peut traiter d'une manière précise les problèmes associés aux modèles les plus simples, et d'une manière approximative ceux relatifs à des modèles plus généraux. Mais d'abord, quels sont les modèles les plus simples, quelles sont les hypothèses qui rendent les modèles simples ? Elles concernent essentiellement les distributions des arrivées et des durées de service, et sont donc de nature statistique. Elles apparaissent en b et c dans le modèle $M/M/1$ qui est traité ci-dessous.

Modèle poissonnien $M/M/1$.

Hypothèses.

a. Une station et une file, composée de clients d'un même type (pas de priorités), source et file à capacité infinie, arrivées individuelles, services individuels.

b. Arrivées poissonniennes. Soit $\Delta t > 0$ un intervalle de temps quelconque et soit $A(\Delta t)$ le nombre correspondant d'arrivées de clients. $A(\Delta t)$ est stochastiquement indépendant du nombre d'arrivées dans tout intervalle antérieur à Δt et il existe un nombre $\lambda > 0$ tel que

$$\text{Prob. } [A(\Delta t) = 1] = \lambda \Delta t + o(\Delta t),$$

$$\text{Prob. } [A(\Delta t) > 1] = o(\Delta t),$$

$$\text{avec } \lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0.$$

Il résulte de ces hypothèses que

$$\text{Prob. } [A(\Delta t) = 0] = 1 - \lambda \Delta t + o(\Delta t),$$

et on peut établir que cet ensemble d'hypothèses est équivalent à

$$\text{Prob. } [A(t) = a] = e^{-\lambda t} \frac{(\lambda t)^a}{a!} \quad (a = 0, 1, \dots),$$

où $A(t)$ représente le nombre d'arrivées de clients pendant un intervalle de temps quelconque $t > 0$.

On dit que le processus des arrivées est poissonnien homogène, de paramètre λ . Celui-ci, appelé taux des arrivées, est le nombre moyen d'arrivées par unité de temps.

On peut encore établir qu'il résulte des hypothèses *b* que

$$\text{Prob. (intervalle entre deux arrivées consécutives} \leq t) = 1 - e^{-\lambda t}$$

(distribution exponentielle négative, de paramètre λ).

c. Durées de service de distribution exponentielle négative. On suppose que les durées de service sont indépendantes entre elles et des intervalles entre arrivées de clients et qu'il existe un nombre $\mu > 0$ tel que

$$\text{Prob. (durée} \leq t) = 1 - e^{-\mu t}$$

(distribution exponentielle négative, de paramètre μ ; celui-ci est l'inverse de la durée moyenne).

On peut voir que cette hypothèse implique que les instants de fins de service constituent, pour une station constamment occupée, un processus de Poisson de paramètre μ .

C'est ce double caractère poissonnien des arrivées et des fins de service qui rend ce modèle plus simple que tout autre modèle aléatoire, plus simple même que le modèle analogue dans lequel les arrivées seraient équidistantes ou les durées de service constantes. Grâce aux hypothèses *b* et *c*, le processus $\{N(t)\}$, où $N(t)$ représente le nombre de clients présents dans le système à l'instant t , est, en effet, un processus de Markov, c'est-à-dire un processus dans lequel le futur ne dépend en probabilité que du présent, non du passé, circonstance qui facilite grandement le calcul des probabilités de transition d'un état à un autre. Ce caractère markovien du processus est une conséquence de la propriété d'« oubli », propre à la distribution exponentielle négative :

la probabilité pour qu'un service en cours — ou un intervalle entre deux arrivées consécutives — prenne fin avant l'expiration d'un intervalle de temps t quelconque, ne dépend pas de la durée déjà atteinte de ce service — ou de la durée déjà atteinte de cet intervalle entre arrivées — et la distribution de la « survie » du service — ou de ... — est en fait la même que celle de la durée tout entière. Que cette propriété d'oubli disparaisse dans le cas de durées de service constantes ou d'arrivées équidistantes est évident, puisqu'une diminution future — ou une augmentation — de $N(t)$ est alors d'autant plus rapprochée que le service est déjà plus avancé — ou que l'arrivée précédente appartient déjà plus au passé.

On a constaté, dans bien des cas, une concordance statistique fort satisfaisante de distributions observées de durées avec une distribution exponentielle négative. Il en est notamment ainsi pour les durées des conversations téléphoniques, du moins... lorsque la tarification n'en dépend pas.

C'est l'auteur anglais D.G. Kendall — Ann. Math. Stat., v. 24, 1953, pp. 338-354 — qui a introduit pour le modèle décrit en a, b, c , la notation, devenue classique, $M/M/1$ (M et M respectivement pour les arrivées poissonniennes et la distribution exponentielle négative des durées de service, 1 pour une seule station).

J'ai dit plus haut que le processus $\{N(t)\}$ correspondant est markovien. C'est plus précisément un processus de Markov discret dans l'espace et continu dans le temps (puisque $N(t)$ ne peut prendre d'autres valeurs que les valeurs entières 0, 1, ..., alors que t peut prendre toutes les valeurs de 0 à ∞). C'est plus précisément encore, par l'assimilation d'une arrivée à une naissance et d'une fin de service à un décès, un processus du type « de vie ou de mort » (ou « de naissance et de décès » : « birth and death »), tel qu'on en utilise dans l'étude de l'évolution de certaines populations d'êtres vivants ou de particules.

Proposons-nous maintenant de voir comment on peut rechercher la distribution de probabilité de $N(t)$, étant donné la présence de n_0 clients dans le système à l'instant $t = 0$. Soit $P_n(t) = \text{Prob. } [N(t) = n]$. La méthode du bilan des probabilités transforme les hypothèses a, b, c , par le truchement des équations

$$\begin{cases} P_n(t + \Delta t) = P_n(t) [1 - (\lambda + \mu) \Delta t] + P_{n-1}(t) \lambda \Delta t + P_{n+1}(t) \mu \Delta t + o(\Delta t), \\ P_0(t + \Delta t) = P_0(t) (1 - \lambda \Delta t) + P_1(t) \mu \Delta t + o(t), \end{cases} \quad n \geq 1$$

en les équations, dites de Kolmogorov,

$$\begin{cases} \frac{dP_n(t)}{dt} = -(\lambda + \mu) P_n(t) + \lambda P_{n-1}(t) + \mu P_{n+1}(t), & n \geq 1 \\ \frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t). \end{cases}$$

Le premier terme de $P_n(t + \Delta t)$, dans son expression du second membre, est, à $o(\Delta t)$ près, la probabilité relative à la présence de n clients dans le système à l'instant t , et de n encore à l'instant $t + \Delta t$; le deuxième terme est relatif, de même, à $o(\Delta t)$ près, à $(n - 1)$ et n clients, et le troisième à $(n + 1)$ et n . Les probabilités afférentes aux autres possibilités sont toutes contenues dans $o(\Delta t)$. Pour $n = 0$, c'est-à-dire dans l'expression de $P_0(t + \Delta t)$, il n'y a ni $P_n(t) \mu \Delta t$ ni $P_{n-1}(t) \lambda \Delta t$, en raison respectivement de l'impossibilité d'une fin de service lorsque $n = 0$ ou de l'impossibilité de $N(t) = -1$.

Soit enfin la fonction génératrice

$$G(z, t) = \sum_{n=0}^{\infty} P_n(t) z^n.$$

On voit aisément qu'elle doit satisfaire, en raison des équations de Kolmogorov, puisque

$$\frac{\partial G}{\partial t} = \sum_{n=0}^{\infty} \frac{dP_n(t)}{dt} z^n,$$

à l'équation

$$z \frac{\partial G}{\partial t} = (1 - z) [(\mu - \lambda z) G(z, t) - \mu P_0(t)],$$

qu'on ramène, par une utilisation de la transformation de Laplace, à une équation algébrique en la transformée $G^*(z, s)$ de $G(z, t)$. Cette transformée $G^*(z, s)$ s'exprime ainsi en fonction de la transformée $P_0^*(s)$ de $P_0(t)$. Grâce à une application du théorème de Rouché, pour laquelle nous renvoyons aux ouvrages, on peut déterminer $P_0^*(s)$ en tenant compte de ce que $G^*(z, s)$ converge pour $\operatorname{Re}(s) > 0$ et $|z| \leq 1$. L'expression trouvée pour $G^*(z, s)$ est développée en série entière $\sum_{n=0}^{\infty} P_n^*(s) z^n$. L'inversion des $P_n^*(s)$ donne enfin les probabilités $P_n(t)$; elles s'expriment à partir de fonctions de Bessel modifiées de première espèce, et dépendent de n_0 .

La méthode s'applique aussi à d'autres modèles, mais l'esquisse que je viens d'en donner à propos de M/M/1 montre que, malgré la simplicité des

principes mis en œuvre, elle conduit à des opérations d'une certaine technicité.

Les choses se simplifient lorsqu'on se contente de rechercher la distribution limite, pour $t \rightarrow \infty$. Point n'est besoin, pour la déterminer, de passer par le calcul des $P_n(t)$. Les $p_n = \lim_{t \rightarrow \infty} P_n(t)$ satisfont, en effet, au système

algébrique qu'on obtient en remplaçant les premiers membres du système de Kolmogorov par 0 et les $P_n(t)$, $P_{n-1}(t)$, $P_{n+1}(t)$, $P_0(t)$ et $P_1(t)$ des seconds membres respectivement par p_n , p_{n-1} , p_{n+1} , p_0 et p_1 , c'est-à-dire, pour M/M/1, au système

$$\begin{cases} -(\lambda + \mu) p_n + \lambda p_{n-1} + \mu p_{n+1} = 0, & n \geq 1 \\ -\lambda p_0 + \mu p_1 = 0, \end{cases}$$

dont on voit immédiatement qu'il peut se résoudre par récurrence : p_1 en fonction de p_0 , p_2 en fonction de p_1 et p_0 , c'est-à-dire en fonction de p_0 , etc. On trouve, avec $\rho = \lambda/\mu$, appelé coefficient d'utilisation du système ou intensité du trafic, que

$$p_n = \begin{cases} 0 & \text{si } \rho \geq 1, \\ \rho^n (1 - \rho) & \text{si } \rho < 1. \end{cases}$$

Dans le cas $\rho \geq 1$, il n'y a donc pas, en fait, de distribution limite. Ce résultat pouvait être prévu, la file devenant inévitablement infinie s'il y a plus d'arrivées que de départs, et la probabilité $P_n(t)$ devant donc tendre nécessairement vers 0, pour tout entier n fini. Dans le cas $\rho < 1$, la distribution limite existe et est du type simple géométrique. Il convient enfin de remarquer qu'elle est indépendante de n_0 , le nombre de clients dans le système en $t = 0$.

On peut déduire de ces résultats sur les p_n d'autres résultats, relatifs par exemple aux nombres moyens $\bar{N}_f$ et $\bar{N}_s$ de clients dans la file ou dans le système (guichet compris), au temps d'attente moyen $\bar{W}_f$ dans la file ou au temps de séjour moyen $\bar{W}_s$ dans le système; on peut même en déduire la fonction de répartition du temps d'attente :

$$\text{avec } \rho < 1, \bar{N}_f = \sum_{n=2}^{\infty} (n-1) p_n = \frac{\rho^2}{1-\rho}, \bar{N}_s = \sum_{n=1}^{\infty} n p_n = \frac{\rho}{1-\rho},$$

$$\bar{W}_f = \frac{\bar{N}_f}{\lambda}, \bar{W}_s = \frac{\bar{N}_s}{\lambda}, \text{ Prob. } (W_f \leq w) = 1 - \rho e^{-(1-\rho)\mu w}.$$

Voici enfin quelques résultats numériques, en régime permanent, $\rho < 1$.

ρ	Prob. d'inactivité du serveur $= 1 - \rho$	Nombre moyen de clients dans le système $= \bar{N}_s = \rho / (1 - \rho)$	Prob. de plus de quatre clients dans le système $= \rho^5$
0,1	0,9	0,111	0,00001
0,2	0,8	0,250	0,0003
0,3	0,7	0,429	0,002
0,6	0,3	1,500	0,078
0,9	0,1	9,000	0,590

On y voit et on y mesure combien, à petite intensité de trafic, les choses sont satisfaisantes du point de vue de l'attente des clients, mais peu satisfaisantes du point de vue de l'activité du serveur, et aussi à quelle cadence elles se modifient lorsque ρ croît.

Autres modèles poissonniens.

La méthode s'applique à d'autres modèles, pourvu que l'on ne touche pas au caractère poissonnien des hypothèses b et c relatives aux arrivées de clients et durées de service. Exemples :

s serveurs (M/M/s). Il convient, dans le système de Kolmogorov, de remplacer

$$\begin{aligned} \text{lorsque } 1 \leq n < s, & \left\{ \begin{array}{l} \mu P_n(t) \text{ par } n \mu P_n(t) \\ \mu P_{n+1}(t) \text{ par } (n+1) \mu P_{n+1}(t), \end{array} \right. \\ \text{lorsque } n \geq s, & \mu P_n(t) \text{ et } \mu P_{n+1}(t) \text{ respectivement} \\ & \text{par } s \mu P_n(t) \text{ et } s \mu P_{n+1}(t). \end{aligned}$$

Le taux « global » des fins de service dépend maintenant de l'état du système. Quant à la condition de l'existence d'une distribution limite, elle est devenue $(\lambda/\mu s) < 1$.

Système à capacité finie (l). Les systèmes d'équations deviennent des systèmes finis. Il y a toujours une distribution limite.

M/M/1, mais avec λ et μ , fonctions de t .

$M/M/s$, avec source finie (m). Exemple : m machines et s réparateurs.
Les systèmes d'équations sont finis et il convient de remplacer

$$\lambda \text{ par } (m - n) \lambda,$$

$$\mu \text{ par } \begin{cases} n \mu & \text{si } n \leq s \\ s \mu & \text{si } n > s. \end{cases}$$

Résultats numériques :

	Proportion de serveurs inactifs	Proportion de clients (machines) en attente
$\lambda/\mu = 0,10; m = 6; s = 1$	0,4845	0,0549
$\lambda/\mu = 0,10; m = 20; s = 3$	0,4042	0,01694

On constate que, tant du point de vue de la proportion de serveurs inactifs que de la proportion de machines en panne, qui attendent le début de l'opération de réparation, il est plus rentable de répartir les réparateurs en groupes de 3, ayant à s'occuper de 20 machines, que de leur confier à chacun individuellement le soin de s'occuper de 6 machines seulement.

3. Modèles non poissonniens.

$M/E_k/1$. Arrivées poissonniennes et durées de service de distribution d'Erlang d'ordre k — k entier > 0 — (E_k), ce qui est la distribution d'une somme de k variables indépendantes équidistribuées, de distribution exponentielle négative. Si $k\mu$ est le paramètre de cette dernière distribution, la

moyenne de E_k vaut $\frac{k}{k\mu} = \frac{1}{\mu}$ et son écart-type $\frac{1}{\mu\sqrt{k}}$, d'où son coeffi-

cient de variation $= \frac{1}{\sqrt{k}}$.

La distribution des durées de service est donc E_k chaque fois que les services se décomposent physiquement en k phases de durées exponentielles négatives indépendantes, de même paramètre. Mais, inversement, si la distribution des durées de service est E_k , sans qu'il y ait cette décomposition physique, on peut supposer parfaitement chaque durée divisée en k phases fictives, toutes exponentielles négatives.

La méthode du bilan des probabilités s'applique dès lors au modèle $M/E_k/1$ si l'on substitue au nombre $N(t)$ de clients qui se trouvent dans le système à l'instant t , le nombre $\bar{N}(t)$ de phases de service que ces clients représentent : k par client se trouvant en file et $j \leq k$ pour le client se trouvant au guichet, si j est le nombre de phases par lesquelles il doit encore passer, y compris celle dans laquelle il se trouve. Le processus $\{\bar{N}(t)\}$ est, en effet, parfaitement markovien. *In fine*, on passera très facilement des renseignements obtenus pour $\bar{N}(t)$ à ceux désirés pour $N(t)$, sur la base de

$$\begin{cases} P[N(t) = n] = \sum_{j=1}^k P[\bar{N}(t) = (n-1)k + j], & n \geq 1 \\ P[N(t) = 0] = P[\bar{N}(t) = 0]. \end{cases}$$

Et de même, bien entendu, en régime permanent.

Une autre manière de présenter les choses est de caractériser les états du système par deux indices au lieu d'un, $P_{n,i}(t)$ représentant alors, par exemple, du moins pour $n \geq 1$, la probabilité pour qu'il y ait, à l'instant t , n clients dans le système et que le client au guichet en soit à sa i^{e} phase de service. On a

$$P[N(t) = n] = \sum_{i=1}^k P_{n,i}(t), \quad n \geq 1$$

et, en régime permanent, avec $p_{n,i} = \lim_{t \rightarrow \infty} P_{n,i}(t)$,

$$p_n = \sum_{i=1}^k p_{n,i}, \quad n \geq 1.$$

Pour $n = 0$, un second indice n'aurait pas de sens.

$M/HE/1$. Arrivées poissonniennes et durées de service de distribution hyperexponentielle, celle-ci étant apparentée à la distribution exponentielle négative par la considération d'exponentielles négatives en parallèle plutôt qu'en série. Les durées de service seraient, par exemple, hyperexponentielles si, pour chaque client se présentant au guichet, il y avait une probabilité a_h ($h = 1, \dots, r$) — avec $\sum_{h=1}^r a_h = 1$ — pour que sa durée de service soit exponentielle négative de paramètre μ_h . Contrairement à la distribution d'Er lang, la distribution hyperexponentielle est, à moyennes égales, plus dispersée que l'exponentielle négative.

La méthode du bilan des probabilités s'applique « à deux indices », $P_{n,h}(t)$ représentant, pour $n \geq 1$, $1 \leq h \leq r$, la probabilité pour qu'il y ait, à l'instant t , n clients dans le système, et que le client au guichet soit servi de la h^e manière parmi les r possibles. On a

$$P[N(t) = n] = \sum_{h=1}^r P_{n,h}(t)$$

et, en régime permanent, avec $p_{n,h} = \lim_{t \rightarrow \infty} P_{n,h}(t)$,

$$p_n = \sum_{h=1}^r p_{n,h}.$$

M/G/1. De même qu'on a considéré plus haut des exponentielles en parallèle, on peut supposer aussi des Erlang en parallèle — il faut alors trois indices — et, par un choix convenable, obtenir ainsi ce qu'on peut appeler une distribution « hypererlang » qui ait moyenne et coefficient de variation imposés, la mise en série diminuant le coefficient de variation et la mise en parallèle l'augmentant.

C'est dire que si l'on se trouve en présence d'un modèle M/G/1, à arrivées poissonniennes et à durées de service indépendantes et équidistribuées, mais de distribution quelconque G, on peut, par ce moyen, obtenir une approximation de G — deux Erlang en parallèle y suffisent d'ailleurs. Or, la méthode du bilan est applicable au modèle approché.

Autres modèles encore. Tout ce qui vient d'être dit peut se généraliser au cas de $s > 1$ serveurs. Et l'on peut aussi traiter de modèles $E_k/M/1$ ou s , à mécanisme d'arrivées en k phases exponentielles négatives, ou $HE/M/1$ ou s , ou, par approximation, de modèles $GI/M/1$ ou s , à intervalles entre arrivées indépendants et équidistribués, de distribution quelconque — GI pour « general input » — ou même, grâce à une double approximation, l'une pour les arrivées, l'autre pour les services, de modèles $GI/G/1$ ou s .

La méthode du bilan des probabilités est, en conclusion, d'applicabilité plus générale qu'il n'y paraît à première vue.

Les systèmes d'équations peuvent, bien entendu, être assez complexes. Lorsqu'ils sont infinis, on choisira, pour leur résolution numérique, un nombre v tel que les probabilités pour qu'il y ait v ou plus de v clients dans le système soient négligeables, et on les posera égales à 0. Par le choix de diverses valeurs de v , on peut extrapoler vers $v = \infty$.

Quelques livres généraux sur la théorie des files d'attente.

COX, D.R. and SMITH, Walter, L., *Queues*, Methuen (1961).

KAUFMANN, A. et CRUON, R., *Les phénomènes d'attente*, Dunod (1961).

MORSE, Philip M., *Queues, Inventories and Maintenance*, ORSA, Publications in Operations Research, Wiley (1958).

PRABHU, N.U., *Queues and Inventories*, Wiley (1965).

SAATY, Thomas, L., *Elements of Queueing Theory*, McGraw-Hill (1961).

TAKACS, L., *Introduction to the Theory of Queues*, New York - Oxford University Press (1962).

DE MOGELIJKHEDEN VAN HET STATISTISCH ONDERZOEK IN DE LINGUISTIEK

door L.K. ENGELS,

Katholieke Universiteit te Leuven

Elke wetenschap is in essentie onafhankelijk van gelijkwelke andere wetenschap. Ze is een volledige wezenheid op zichzelf. Twee verschillende wetenschappen hebben gewoonlijk verschillende wetenschappelijke theorieën. Men kan dus de methodes en de theorieën van bepaalde wetenschappen niet zonder meer toepassen of aanwenden in een andere wetenschappelijke discipline. Tegen deze regel werd de laatste jaren maar al te dikwijls gezondigd : als we spreken van mathematisatie in de linguïstiek, dan bedoelen we slechts, dat de mathematica kan *helpen* bij het descriptief, of statistisch bewerken van linguïstisch materiaal, maar we beweren niet dat we aan zogenaamde *mathematische taalkunde* zouden doen, omdat zulke taalkunde door haar notionele vermenging van twee totaal verschillende disciplines op een mislukking moet uitlopen. Het is onmogelijk de linguïstiek, die zoals de sociologie en de economie, een DESCRIPTIEVE wetenschap is, te beoefenen vanuit het standpunt van de mathematica, die een PREDICTIEVE wetenschap is, zoals de scheikunde, de fysica en de sterrekunde.

Het is ook onmogelijk de linguïstiek te beoefenen vanuit het standpunt van de psychologie of de logica, die voor een groot deel NOTIONELE wetenschappen zijn, omdat ze werken met *formele begrippen*.

De eerste voorstanders van de automatische taalkunde en automatische vertaling b.v., waren meestal mathematici en logistiekers (1). We kunnen over hun pogingen, die reeds vijftien jaar aan gang zijn, of waren, in verscheidene instituten voor research, over de gehele wereld verspreid, slechts zeggen, dat ze alle mislukt zijn : de formele aanpak van taal door middel van wiskundige predictieve modellen in plaats van door observatie van de taalfeiten, kon niet bewerken, dat de problemen van de mechanisatie en automatisatie van analyse en vertaling werden opgelost. Als men die problemen met predictieve modellen wil oplossen dan neemt men tegenover het te bereiken resultaat zo ongeveer dezelfde houding aan als de taalgeleerde die zou zoeken naar de taal die bepaalde

(1) Bar-Hillel, Carnap, Yngve, Hays, Booth, Oettinger e.a.

wezens op de planeet Mars zouden spreken. De linguïst kan zich niet veroorloven om voor zijn onderzoek van de bestaande taalfenomenen eerst een taal te creëren van buitenaards karakter (2).

Zo zullen de statistici ook niet beweren dat hun wetenschap, als empirische discipline, een onderdeel is van de mathematica. De statistiek berust wel op mathematische methodes, vooral dan die van de probabiliteitstheorie, maar ze richt zich vooral tot descriptie en verklaring van geobserveerde patronen, terwijl de mathematica veeleer een totaal abstraherende discipline is, die naar een THEORIE, naar een MODEL zoekt van de werkelijkheid. Soms stelt men de mathematica voor, als zou ze de statistiek omvatten, wat ook weer een valse opvatting is. Al die verwarringen leiden dan in de linguïstiek soms tot de naam *mathematische linguïstiek*, wanneer in feite slechts bedoeld wordt, dat de linguïstiek gebruikt maakt van bepaalde statistische methodes bij haar taalkundig onderzoek, maar het zal dan weer, zoals verder zal blijken, geen totale transpositie zijn van statistische methodes op de taalkunde, we zullen moeten zoeken naar aangepaste *statistisch-linguïstische methodes*. De evolutie in de wetenschap is oorzaak van de verwarring van de begrippen. We hebben immers nood aan elkaar: fysici interesseren zich voor de fonetiek van de taal, de taalkundige begint het nut in te zien van het gebruik van een computer om zijn steekkaarten te vervangen, de statisticus wou graag weten hoe het met sommige taalproblemen gesteld is. In het begin ziet men niet meer klaar: noties en methodes worden vermengd en verward, totdat men tot bezinning komt en de zaken weer duidelijker begint te onderscheiden (3). We zijn aan zo een inzichtelijk keerpunt geraakt, zodat we met een gerust gemoed kunnen handelen over de mogelijkheden van statistisch onderzoek in de linguïstiek.

Taalstatistiek is nog een zeer jonge wetenschap: het eerste statistische onderzoek van grote omvang, op de Duitse taal toegepast, niet met een zuiver linguïstisch doel voor ogen, maar ten dienste van de stenografie, werd gedaan door Kaeding in 1898: het was een frequentielijst van de woorden der Duitse taal op bronnen die zo maar 10.910.777 lopende woorden omvatten (4). Tussen 1920 en 1940 werd er over de gehele wereld en voor verscheidene talen (vooral het Engels) naar woordfrequentie gezocht; het beste resultaat werd bereikt in 1936 met een lijst voor het Engels: *Interim Report on Vocabulary Selection* (5). De principes, waarmee deze lijst werd uitgewerkt, vonden dan hun toepassing op vele talen.

(2) Hans Freudenthal, *Lincos, Design of Language for Cosmic Intercourse*. Amsterdam, 1960.

(3) Robert M.W. Dixon, *Linguistic Science and Logic*, Mouton, The Hague, 1963.

(4) Kaeding, *Häufigkeitswörterbuch der deutschen Sprache*, Steglitz, 1897, 1898.

(5) *Interim Report on Vocabulary Selection*, IRET, Tokyo, 1936. Leiding: H.E. Palmer.

Zo verscheen in 1939 een frequentielijst voor het Nederlands door De la Court en Vannes. Na de wereldoorlog geraakte het echt statistische onderzoek van de taal een beetje in de schaduw door de *formele* aanpak van de linguïstiek, die zou leiden naar de overmoedige proeven op automatische analyse, automatische vertaling, automatisch opvangen en ordenen van informatie. Wij weten niet of de mislukking van de pogingen tot mechaniseren van een wezen, zo wispelturig als de taal, de rechtstreekse aanleiding is geweest voor de kentering die duidelijk merkbaar is vanaf 1960 in de linguïstiek: de taalstatistiek, waarvoor de vorsers op het gebied van de automatische vertaling niets anders dan misprijzen kenden, krijgt weer meer waardering. In 1961 begint Karlgren te Stockholm met een tijdschrift voor linguïstische statistiek (6). De theoretische studies over taalstatistiek van Herdan verschijnen in 1960 en 1962 (7). In 1963 maakte Juilland zijn zevenjarenplan bekend om een aantal Romaanse talen statistisch te onderzoeken (8). Bewonderenswaardig is het resultaat van veertig jaar inspannende arbeid (zonder de hulp van een computer om de berekeningen te maken), geleverd door Helmut Meier in zijn *Deutsche Sprachstatistik* (9).

Als reactie tegen de voorstanders van de formele taalanalyse, die een uiting is van romantisch streven naar de niet te bereiken «blauwe Blume», betekent de taalstatistiek een realistische aanpak, moeizaam, langdurig, maar met zekere resultaten, omdat het descriptieve karakter van de taalkunde in wezen wordt erkend, en omdat de taalkundige zo nederig geworden is om te erkennen dat het domein van zijn wetenschap, de levende taal, slechts nauwkeuriger zal kunnen benaderd worden, als vele elementen erin MEETBAAR zijn geworden.

Een primordiale vraag vooraleer we van taalstatistiek kunnen spreken houdt dus verband met het quantitative aspect van de taal: WAT is meetbaar in de taal? Als we Istvan Fodor moeten geloven in zijn werk *The Rate of Linguistic Change*, dan blijft er schijnbaar niet veel meer over om te meten in de taal. Fodor beweert, dat de invloed van de geschiedenis, de cultuur, de geografie, de nabijheid en contact met andere talen, het nationaal karakter van een volk, hoewel EXTERNE invloeden, onmogelijk kunnen gemeten worden of quantitatief voorgesteld. Hij ziet de zaken vooral in een historisch perspectief, d.w.z. het zijn externe invloeden die men noodzakelijk meetbaar zou moe-

(6) *Statistical Methods in Linguistics*, Skriptor, Stockholm. 1 : 1961, 2 : 1963, 3 : 1964.

(7) Herdan : *Typen Token Mathematics*, Mouton, 1960 - *The Calculus of Linguistic Observation*, Mouton, 1962.

(8) A. Juilland, *Dictionnaire inverse de la langue française*, Mouton. *Frequency Dictionary of Spanish Words*, Mouton.

(9) Helmut Meier, *Deutsche Sprachstatistik*, G. Olm, Hildesheim, 1964.

ten maken om de taalveranderingen in de loop van de geschiedenis te kunnen verklaren. Maar waarom ons reeds druk maken over de verklaring van de ontwikkeling van de taal, als onze huidige geschreven of gesproken uitingen hun geheim nog niet hebben prijsgegeven? Er zijn heel wat terreinen — ik beperk me van nu af tot onze eigen Nederlandse taal — die wachten op een quantitative aanpak (om te beginnen met louter *descriptieve* statistische methodes): distributie van lettertekens, syllaben, fonemen; woordfrequentie, veranderingen in woordgebruik, neologismen, ontlending aan vreemde talen; frequenties van taalconstructies; onderzoek van morfemen (buigingsresten, derivatie, woordvorming en dergelijke).

Vermits er voor het Nederlandse taalgebied buiten de frequentielijst van Vannes (1939) nog vrijwel niet aan taalstatistiek werd gedaan zal het interessant zijn te vernemen, wat er op dit ogenblik gebeurt en welke plannen er gesmeed worden voor de toekomst. Aan het rekencentrum van de Amsterdamse universiteit maakte men bij wijze van proef een lijst met afnemende frequentie uit een reeks van 50.000 lopende woorden (uit kranten van 1 bepaalde dag). Het was een voorbeeld van descriptieve taalstatistiek, uitgevoerd door middel van een elektronische rekenmachine. De lijst bood weinig linguïstische relevantie, omdat het aantal woorden te klein was om gelijkwat te besluiten uit de resultaten; de hele operatie was dus waarschijnlijk bedoeld als een proef met een computerprogramma.

In het Instituut voor Toegepaste Linguïstiek te Leuven wordt er in samenwerking met het rekencentrum van de universiteit (onder de leiding van Professor Florin) geëxperimenteerd met computerprogramma's, die in staat moeten blijken om de taal te analyseren en te reconstrueren; zodra we erin geslaagd zijn door vernuftige maar eenvoudige coderingen aan de machine de opdracht te geven om voor ons woorden, woordverbindingen, constructies, zinsstukken, zinsdelen en dergelijke op bevel uit te zoeken, dan is al wat de machine ons op haar tabellerijsten bezorgt meetbaar, telbaar geworden en vatbaar voor descriptieve statistieken, die we naar gelang het werk vordert, zonder ophouden bekomen.

Onze eerste publicatie wordt een werk dat bepaalde statistische onderzoeken kan voorbereiden: het is een retrograad woordenboek zoals het reeds voor het Duits en het Frans is verschenen. De woorden staan er gerangschikt volgens hun uitgangen, zodat het een mooi hulpmiddel wordt voor linguïsten die belangstellen in de uitgangen van de Nederlandse woorden. De vele morfologische kenmerken van het Nederlands, voor zover ze op het einde van de woorden voorkomen kunnen dan gemakkelijker onderzocht worden.

Onze tweede opgave bestaat uit de vervaardiging van een nieuwe frequentielijst van het Nederlands. We zijn met dit groot werk begonnen, omdat

we overtuigd zijn (met Herdan) dat linguïstische elementen onderworpen zijn aan de wetten van de kansverdeling. In de woordenschat is het distributiepatroon gelijk, als een voldoende hoeveelheid woorden werd onderzocht, hoewel in iedere zin de woordenschat afhangt van inhoud, auteur, stijl. Er bestaat een grote uniformiteit van de massa, ondanks de diversiteit van de onderdelen. Fonemen en letters vertonen een grote stabiliteit van distributie, zelfs zo groot dat individuele verschillen totaal worden uitgewist. Er zijn zo'n belangrijke combinatorische wetmatigheden dat het mogelijk is informatietheoretische vergelijkingen te maken, gesteund op een binaire benadering van de onderzochte quanta. Het is begrijpelijk dat een werk zoals dit wordt aangevat, omdat de vooruitgang van de wetenschap ons daartoe aanzet. De linguïsten die de theoretische statistiek van de woordenschat hebben uitgewerkt konden bepaalde beschrijvende formules in wiskundige vorm vastleggen (Herdan, Josselson, Zipf) : dat werd gedaan voor de frequentiedistributie van de tekst, voor het rekenkundig gemiddelde, voor de standaarddeviatie, voor het variatiecoëfficiënt. We hebben de probabiliteitsberekeningen laten uitvoeren voor de resultaten van de frequentielijst van De la Court-Vannes, onze voorganger van 1939 (10), en we stelden vast, dat er gebreken waren aan die frequentielijst : de zeven radii van frequenties samen genomen telden 3.296 woorden ; de zevende radius heeft nog een frequentie van 25 met een fout van 40 %. Gesteld echter dat we een fout van 30 % willen aanvaarden, dan komt op een totaal van 1.000.000 lopende woorden het aantal woorden waarvan de frequentie nog relevant is op slechts 2.200. De zevende radius van de vroegere frequentielijst zou dus moeten vervallen, omdat hij niet voldoende zekerheid biedt. Er zijn nog andere bezwaren tegen deze lijst : Het materiaal is van zeer verschillende aard en het werd zeer ongelijk gemengd. Men had afzonderlijke lijsten moeten berekenen voor elke taalsoort of taalveld. De keuze in de tijd is niet verantwoord : naast *De Kleine Johannes*, geschreven in 1885 (met 21.000 woorden vertegenwoordigd), vinden we krantenartikels van 1930. Er is geen eenheid in de spreiding over 50 jaar. Er zijn vertalingen vertegenwoordigd uit het Engels, het Duits en het Scandinafs ; er valt geen maatstaf vast te stellen om het materiaal te kiezen : romanliteratuur varieert van 800 tot 21.000 woorden voor 1 auteur.

Het huidig frequentie-onderzoek gebruikt totaal andere maatstaven. Er worden twee verschillende werkwijzen gebruikt ; maar steeds door middel van elektronische rekenmachines. De eerste werkwijze schijnt aantrekkelijk op het eerste gezicht, omdat er zovele bewerkingen automatisch kunnen geschieden : men ponst het materiaal op band of kaarten, nummert de woorden per regel

(10) *Vocabulaire du néerlandais de base*. De Sikkel, 1939.

of per zin, samen met een code voor auteur, soort tekst en dergelijke. Men kan dan een frequentielijst maken van *vormen*: alles staat op woordkaarten die door de machine volledig automatisch kunnen gealfabetiseerd worden, geteld en hun frequenties berekend. De bewerking gaat zeer vlug, maar er is niet het minste onderscheid tussen de homoniemen. Men kan natuurlijk de woorden met hun context opnemen; de meeste homoniemen kunnen dan achterhaald worden, maar niet automatisch. De tangvormingen van het Nederlands (de gescheiden opstellingen) ontsnappen totaal aan de telling. Er is een enorme geheugencapaciteit nodig in de machine.

In het Instituut voor Toegepaste Linguïstiek te Leuven gaat men uit van een ander principe: men is daar niet zozeer bekommerd om het automatiseren van het proces, als dat moet uitgeboet worden door gemis aan degelijkheid en precisie. Er wordt gewerkt op de tabelleerlijsten die zowel het woord als de context duidelijk weergeven. Daarop wordt dan de zogenaamde pre-editie aangebracht, die door velen als zeer tijdrovend wordt beschouwd, maar die uiteindelijk betere resultaten oplevert, omdat de linguïstische beschrijving correcter en vollediger kan gebeuren en omdat er machines met kleinere capaciteit kunnen gebruikt worden. Ons linguïstisch standpunt verschilt ook sterk van dat van onze voorgangers: wij zijn van oordeel, dat het onmogelijk is bij de analyse van een tekst lexicologie en syntaxis volledig te scheiden. Zelfs bij zogenaamd zuiver lexicologisch werk (het samenstellen van een woordenboek) wordt telkens de syntaxis betrokken. Bij ons onderzoek van de lexicale constructies hebben wij dan ook syntactische elementen betrokken, en bij de behandeling van de zinsconstructie hebben we rekening gehouden met lexicale tekens. De grote moeilijkheid is steeds het vinden van een coderingsformule die soepel genoeg is om allerlei nuances te registreren, om onvoorziene gevallen op te vangen, maar tevens eenvoudig genoeg om geen al te grote inspanning te eisen van het menselijk geheugen en aangepast aan de mogelijkheden van de mechanische behandeling door een elektronische rekenmachine.

We zijn er in geslaagd om de machine alle woorden, woordgroepen, zinsdelen, bijzinnen en hoofdzinnen van de tekst te laten *reconstrueren* op bevel. Het zou ons te ver voeren om dit alles hier te tonen. Het volstaat om u te zeggen dat de hele codificering slechts vier kolommen van de ponskaart in beslag neemt: I nummercode van zin tot zin voor elk woord van de zin, die door de machine automatisch wordt aangebracht; twee kolommen voor een *lexico-syntactische woordsoort* die, in cijfers uitgedrukt, 100 verschillende woordsoorten kan differentiëren; een kolom voor een *dependentiecode* van woorden

en zinsdelen onderling in de volzin. Door het feit dat de machine kan *reconstrueren*, kan ze ook al het gevondene tellen en de descriptieve frequenties ervan berekenen en tabelleren. Het gebruikte materiaal bestaat op dit ogenblik uit geschreven taal tussen 1960-65. Op dit ogenblik zijn we met proefnemingen bezig met 5.000 zinnen van 20 verschillende Z.N. romanschrijvers, 250 zinnen per auteur (gerekend van punt tot punt). Later breiden we het materiaal uit naar andere *taelvelden* (essays, toneel, luisterspel, T.V.-interview, krantenartikels, wetenschappelijke proza van diverse aard, ge vulgariseerde wetenschap, zelfs de *kindertaal* wordt niet uitgesloten) (11). In iedere rubriek wordt natuurlijk nog onderscheiden per auteur of per bron. Het voordeel van deze handelwijze ligt in de *spreiding*: voor ieder woord kunnen we die spreiding berekenen: gemiddelde en standaardafwijking σ en variatiequotient v . Het variatiequotient moet ons de maatstaf geven om een bepaalde spreiding te aanvaarden. Wat het resultaat zal zijn van deze bewerkingen kan nu nog niet voorspeld worden. Het is zeer waarschijnlijk dat niet alles in één frequentielijst zal kunnen opgenomen worden. Er moet ongetwijfeld een gemeenschappelijke woordenschat bestaan voor alle taalvelden — een gemeenschappelijke overlapping. Maar waar zal de grens liggen?

We hebben ons tot nog toe beperkt tot geschreven taal. Allicht zullen we later trachten ook de *gesproken* taaluitingen in dit onderzoek te betrekken, als we de technische middelen hebben gevonden om werkelijk gesproken uitingen « in de vlucht » te capteren.

Het spreekt vanzelf dat we ook een onderscheid zullen maken tussen de bekomen woordenschat-frequentielijst, die alle woorden met hun frequenties en spreiding bevat, en een zogenaamde basiswoordenschat, waarin slechts een beperkt aantal woorden zal opgenomen worden, nl. de *nuttigste*, naar gelang het doel, waartoe de lijst zal moeten dienen.

De taalstatistiek krijgt na 1960 stilaan burgerrecht in een domein van de taalwetenschap, dat zelf met moeite zijn plaats als wetenschappelijke discipline aan onze universiteiten kon veroveren: de *toegepaste linguïstiek*. Zij zal meer en meer erkenning genieten, omdat ook de beschouwende filologen hebben ingezien, dat ze best met hun redeneringen vertrekken vanop vaste grond, d.w.z. vanuit precies gemeten en quantitatief berekende gegevens uit de taalwerkelijkheid. Misschien lukt het de *automatische vertaling* nog ooit door middel van de taalstatistiek tot reële in plaats van ingebeelde algoritmen te komen, die nodig zijn om transformatieregels op te stellen. Op een groot

(11) Op dit ogenblik is een onderzoek aan gang in het I.T.L. over de geschreven taalbeheersing moedertaal (opstel) van alle 11n. 1e middelbaar (technisch, 4e graad, M.O.) in het vrij ondww., Vlaams België. De statistische staaltrekking werd gemaakt door de diensten van Dr. R. Van den Bosch: *Dienst voor Statistiek en Planning*, N.S.K.O., Brussel.

aantal wetenschappelijke gebieden zal de taalstatistiek in de toekomst haar woordje kunnen plaatsen: de *dialectologie* zou er zeker goed aan doen de statistische methodes in haar onderzoek te betrekken, voor de nieuwe richting in de algemene taalwetenschap, nl. de *taaltypologie*, is statistisch onderzoek onontbeerlijk. In de psychologie, vooral dan in haar toegepaste gebieden zoals de *audiometrie* en de *logopedie* beklaagt men zich maar al te dikwijls over het gebrek aan taalstatistische gegevens om gehoorstesten, verstaanbaarheidstesten of taalvaardigheidstesten te kunnen opstellen. De linguïst moest verlegen zijn als hij psychologen moet wandelen sturen met lege handen: het onderzoek is pas begonnen, hij kan nog geen geldige resultaten meedelen. Ook de *litteraire kritiek* zou gerust haar intuïties eens mogen toetsen aan de objectieve gegevens over stijl, die door middel van quantitative methodes kunnen bekomen worden.

Het hoeft geen betoog dat de taalstatistiek vooral het vreemde-talenonderwijs kan ten goede komen. De opzet van de tellingen van het Nederlands zoals ze aan de universiteit te Leuven gebeuren hebben eerst en vooral tot doel aan het onderwijs van het Nederlands als tweede taal meer objectieve gegevens te verschaffen over woord en vooral structuurfrequentie om de doeltreffendheid van dat onderwijs te kunnen opvoeren. De taalstatistiek kan ook diensten bewijzen aan het onderwijs van het Nederlands als moedertaal: bij het opstellen van de eerste leesboekjes voor de eerste studiejaren zou men rekening moeten houden met de frequenties van lettertekens, met fonetische schrijfwijzen van de woorden, met gelijkvormigheden en ongelijkvormigheden van de spellingsconventie. Als de taalstatistiek verder gevorderd was, dan zou het veel gemakkelijker zijn een *spellingshervorming* voor te stellen voor talen met internationale spreiding maar hopeloos verouderde schrijfwijze zoals het Engels en het Frans. Laat ons hopen dat bij een volgende spellingswijziging de taalstatistiek eindelijk een helpend handje zal mogen toesteken. De taalstatistiek is ontstaan uit de stenografie — althans in Duitsland — en heeft daar haar bekroning gevonden in een 40-jaar lang onveranderlijk bewaarde eenheidsstenografie. Het spreekt vanzelf dat de huidige Duitse stenografen klagen over het feit, dat de Kaedinglijst niet meer beantwoordt aan de evolutie van de Duitse taal, vooral dan na de tweede wereldoorlog. Maar dat valt in het voordeel uit van de taalstatistiek: men wil een nieuwe Duitse frequentielijst! Tenslotte kan de taalstatistiek grote diensten bewijzen aan de verbetering van de schrijfmachines: men komt over het algemeen tot het inzicht dat niet een internationaal systeem voor alle talen nuttig is, maar dat een klavier de beste diensten bewijst als de opstelling van de toetsen beantwoordt aan de statistische spreiding van de lettertekens in een bepaalde taal. Zo bewijst de taalstatistiek diensten aan de *psychotechniek*, die zich o.a. bezig houdt met de rationalisatie van de schrijfmachine en haar bediening.

De taalstatistiek heeft tot nu toe niet vele beoefenaars. De taalstudie werd tot nu toe meestal gerekend tot de zogenaamde geesteswetenschappen: ze wordt ingedeeld aan onze universiteiten bij de *wijsbegeerte en letteren*. Het is niet gemakkelijk een *notionele* denkwijze in te ruilen tegen objectieve zienswijzen door middel van statistische methodes; het is voor de linguïst een hele stap om zijn aloude steekkaart te laten vervangen door ponskaarten, elektronische banden of tabellerlijsten; het valt hem nog moeilijker zijn regels en inzichten te abstraheren en te beschrijven door middel van mathematische formules. Maar de tijden veranderen en de jeugd die nu aan onze universiteiten met haar studies begint, denkt anders dan vroeger: het zal mogelijk zijn om bij die jongeren enthousiasme op te wekken voor een hulpwetenschap, die in de nabije toekomst heel wat diensten zal kunnen bewijzen aan de maatschappij.

HUIDIGE STAND VAN DE REGIONALE STATISTIEK

door R. DEREYMAEKER

Nationaal Instituut voor de Statistiek

1. De regionale statistiek in de E.E.G.-landen.

Een indeling van de regionale statistiek welke onmiddellijk opvalt is :
1) de statistiek welke betrekking heeft op administratieve indelingen of een som van deze indelingen en dikwijls derwijze is opgevat dat een overzicht wordt gegeven voor alle onderdelen van het land; 2) de statistiek die slaat op gebieden gekozen op basis van bepaalde criteria en het voorwerp uitmaken van ontwikkelingshulp.

De eerstgenoemde gebieden komen b.v. overeen met de arrondissementen in België, de « Kreise » in Duitsland, de departementen in Frankrijk. Deze eenheden worden verder samengevoegd tot provincies of taalstreken in België, Länder in Duitsland, programmagebieden in Frankrijk of drie grote gebieden Centrum-Noord, Centrum Noord-Oost en Zuiden plus eilanden in Italië. De aldus omschreven gebieden maken veelal het voorwerp uit van een vrij groot aantal statistieken die ook gekend zijn voor het ganse land. Men beschikt evenwel niet steeds over het geografisch produkt, zelfs niet voor de grotere onderdelen. Dit produkt is op dit ogenblik slechts berekend voor de drie grote voornoemde gebieden in Italië, in België voor de negen provincies en taalstreken, terwijl in Duitsland een produkt per hoofd is geschat voor iedere « Kreis ».

De gebieden die het voorwerp uitmaken van overheidshulp worden veelal aangeduid door ontwikkelingszones en ontwikkelingspolen of centra. Aldus worden thans in Nederland 8 ontwikkelingszones en 44 ontwikkelingscentra weerhouden die op hun beurt kunnen samengenomen worden tot drie grote gebieden, gelegen in het Noorden, het Zuidwesten en het Zuidoosten van het land. In Duitsland onderscheidt men 3 types van ontwikkelingsgebieden: het zonenrandgebied, een strook van 40 km langsheen het IJzeren Gordijn, de saneringsgebieden en een 40-tal ontwikkelingspolen (Zentrale Orte) gelegen zowel buiten als binnen de saneringsgebieden.

In Italië wordt gans het Zuidelijk gedeelte als ontwikkelingsgebied beschouwd. Het bevat een 25-tal kernen. In eigen land worden 15 ontwikkelingsgewesten onderscheiden, waaraan een aantal voordelen worden toege-

kend; 3 hiervan zijn experimentele zones : Borinage-Centrum, Zuiderkempen en Westhoek.

De criteria voor de afbakening der ontwikkelingsgewesten of saneringsgebieden zijn niet steeds dezelfde, ofschoon het nagestreefde doel gelijklopend is, nl. het bevorderen van het algemeen belang door een evenwichtige spreiding van de economische bedrijvigheid en welvaart over de onderscheiden gewesten van het land en het bestrijden van specifieke economische en sociale moeilijkheden in sommige gewesten (cfr. Belgische wet van 18 juli 1959). Als criteria worden in Nederland weerhouden de werkloosheid en de emigratie. Dit zijn ook de eerste twee kenmerken in België. De zoeven genoemde wet voegt er hieraan evenwel twee toe : seizoen-, wekelijkse of dagelijkse verplaatsingen onder ongunstige voorwaarden van een belangrijk gedeelte der arbeidskrachten en een werkelijk of nakend verval van belangrijke economische werkzaamheden. Duitsland gebruikt eveneens een met de werkloosheid gecorreleerde grootheid, nl. de industriële werkgelegenheid per 1.000 personen doch laat hieraan een directer maatstaf voorafgaan : het binnenlands (= geografisch) produkt per hoofd; als derde criterium geldt de mogelijkheid voor het gebied om zelf te voorzien in zijn financiële noden.

De ontwikkelingscentra of -polen worden gewoonlijk gecreërd in steden waar reeds een begin van industrialisatie aanwezig is; deze worden dan verder uitgebreid ten einde het omliggende sanerings- of ontwikkelingsgebied te stimuleren. Determinerende statistische criteria voor de keuze van afbakening van deze centra zijn moeilijker aanwijsbaar.

Op het statistische vlak stelt men vast dat voor de gebieden welke het voorwerp uitmaken van ontwikkelingshulp doch waarvan de grenzen niet samenvallen met één of meer administratieve eenheden, de beschikbare informatie over het algemeen schaars is. In Duitsland omvatten de saneringsgebieden en ontwikkelingspolen echter over het algemeen een geheel aantal « Kreise », zodat het er ook voor private organismen niet zeer moeilijk valt om voor de gebieden van ontwikkelingshulp heel wat gegevens samen te brengen. Hetzelfde geldt in nog grotere mate voor de 21 programmegebieden in Frankrijk. De statistische gegevens hierover bestrijken gevarieerde terreinen als de demografie, het onderwijs, de actieve bevolking, de werkgelegenheid, de woningbouw, het vervoer, het toerisme, de prijsontwikkeling, de lonen en de sociale zekerheid en het spaarwezen. Men kent evenwel niet het geografisch produkt voor deze 21 gebieden. Hieraan wordt echter sedert enige tijd gewerkt.

De genoemde informatie waarover Frankrijk beschikt voor zijn programmegebieden is in grote mate te vergelijken met deze waarover men in België beschikt en die voorkomt in de arrondissementele of provinciale

studies, welke ongeveer de helft vormen van het dertigtal regionale studies gesubsidieerd door het Ministerie van Economische Zaken en opgesteld door universitaire en andere onderzoekingscentra. De officiële informatie wordt gewoonlijk aangevuld door eigen onderzoeken en berekeningen. Dit is des te meer het geval naarmate de afbakening van de beschouwde streek minder samenvalt met de grote administratieve onderverdelingen. Naast het vastleggen van een aantal kenmerken van het onderzochte gebied wordt veelal getracht op onrechtstreekse wijze een idee te geven van het welvaartspeil. Slechts in een paar studies wordt dit peil op rechtstreekse wijze benaderd, zoals dit in de nationale rekeningen gebruikelijk is, n.l. via het geografisch produkt, de inkomensverdeling en de bestedingen. Voor het Luikse en Antwerpse wordt zelfs gegaan tot het opsporen der industriële relaties tussen diverse nijverheidstakken van de streek.

We zullen in de volgende bladzijden trachten een overzicht te geven van wat in ons land op het vlak van de regionale statistiek bestaat. We zullen dit doen in de driedubbele optiek van de produktie, de inkomensverdeling en de bestedingen. De produktie van een streek vormt de maatstaf van haar economisch en sociaal potentiaal en wijst de oorsprong en determinanten van de welvaart aan. Wegens de pendelbewegingen van een deel van de bevolking van de ene streek naar de andere kan het werkelijk inkomen van de bevolking groter of kleiner zijn dan het inkomen dat gecreeërd wordt in de beschouwde streek. De verschillen in economische structuur en de eruit volgende verschillen in sociale stratificatie leiden eveneens tot uiteenlopende inkomens voor de streken. In de bestedingsoptiek ten slotte wordt de behoeftenbevrediging concreet tot uiting gebracht evenals de regionale verschillen t.o.v. een aantal welvaartselementen.

2. De Belgische regionale statistiek in de produktie-optiek.

a. Vergelijking der niveaus van het geografisch produkt.

Het N.I.S. publiceerde in het Statistisch Tijdschrift n^o 3, 1966, en onder vorm van overdruk in Statistische en Econometrische Studiën n^o 12 de bruto toegevoegde waarde tegen factorkosten en het totale geografisch of binnenlands produkt per bedrijfstak en provincie of taalstreek voor de jaren 1955 tot 1963 en dit zowel in werkelijke als constante prijzen.

Uit deze berekeningen volgt dat in 1961 45,6 % van het bruto binnenlands produkt van België ontstond in de Vlaamse taalstreek, 31,7 % in de Waalse en 22,7 % in het arrondissement Brussel. Vergelijkt men deze percentages met deze van de actieve bevolking werkzaam in genoemde streken volgens de algemene volkstelling op 31 december 1961 dan stelt men vast

dat het produkt van Brussel boven de 21,7 der actieve bevolking ligt welke aldaar werkte en vooral de 17,3 % overtreft van de totale actieve bevolking die daar woonde. In de Vlaamse taalstreek werkte op de datum der volkstelling 47,4 der actieve bevolking terwijl er meer dan de helft woonachtig was. Voor Wallonië is de afwijking tussen het aandeel in het binnenlandse produkt en in de totale actieve bevolking zowel naar woonplaats als naar plaats van tewerkstelling minder groot.

De verdeling van de actieve bevolking op andere data dan de telling is voorlopig niet gekend voor onderdelen van het Rijk. De verdeling van het binnenlands produkt (616 miljard F) over de drie taalstreken is in 1963 geworden : 46,1 % voor het Nederlandstalig landsgedeelte, 30,7 % voor het Franstalig en 23,2 % voor het arrondissement Brussel (oude afbakening).

TABEL I

Vergelijking van aandeel en niveau van het Bruto Geografisch Produkt tegen factorkosten

	Vlaams lands- gedeelte	Waals lands- gedeelte	Arrondiss. Brussel	Het Rijk
1. B.G.P.				
a) 1963	46,1	30,7	23,2	100
b) 1961	45,6	31,7	22,7	100
2. Actieve bevolking in 1961				
a) naar plaats van tewerk- stelling — in %	47,4	31,4	21,7	100
b) naar woonplaats — in %	51,3	30,9	17,3	100
3. Saldo van de pendelbewe- ging in % v. d. plaatselijke actieve bevolking in 1961	—10,1	—4,5	+20,6	...
4. B.G.P. per actieve persoon (plaats van tewerkstelling)				
t.a.v. het Rijk in 1961	96,3	102,5	104,7	100
5. B.G.P. per arbeidsuur in de industrie in 1963 —				
index t.a.v. het Rijk	95,7	105,0	102,0	100
6. B.G.P. p. inwoner in 1963				
— index t.a.v. het Rijk	89,9	93,2	147,5	100

De vrij belangrijke afwijking tussen de procentuele verdelingen der actieve bevolking, al naar gelang de woonplaats of de plaats van tewerkstelling wordt beschouwd, is te wijten aan het positieve of negatieve saldo

van de pendelbeweging. In 1961 gingen er 175 duizend Vlaamse werknemers meer in het arrondissement Brussel, Wallonië of het buitenland werken dan er personen uit deze landsgedeelten of het buitenland in Vlaanderen kwamen werken; voor Wallonië was het saldo eveneens negatief en wel ten belope van 47 duizend eenheden; het arrondissement Brussel kende een positief saldo van 120 duizend personen.

Beschouwen we thans de verschillen in de percentages van het binnenlands produkt enerzijds en deze van de actieve bevolking anderzijds. Zij zijn natuurlijk het gevolg van de afwijkingen in produkt of toegevoegde waarde per actieve persoon. Deze waarde, welke kan worden beschouwd als een gemiddelde arbeidsproductiviteit, is het hoogst in het arrondissement Brussel en het laagst in Vlaanderen. Stelt men het produkt per tewerkgestelde actieve persoon voor het Rijk gelijk aan 100 dan bekomt men in 1961 voor het Vlaamse land 96,3, voor Wallonië 102,5 en voor het arrondissement Brussel 104,7. In Vlaanderen ligt alleen de provincie Antwerpen boven de 100-lijn, in Wallonië ligt alleen Luxemburg er onder.

De verklaring van deze verschillen ligt hoofdzakelijk in de economische structuur en in zekere mate in het verschil in beloningen van de produktiefactoren voor eenzelfde activiteit. Het Vlaamse landsgedeelte is vooral gespecialiseerd in textiel, voedings- en genotmiddelen, Wallonië meer in ijzer en staal, de extractieve nijverheid, ceramiek, glas en cement, terwijl de financiële diensten en de handel bijzonder sterk vertegenwoordigd zijn te Brussel. In de volgorde waarin deze bedrijfstakken werden genoemd neemt ook de toegevoegde waarde toe per eenheid van arbeid. Het is dus de tertiaire sector, welke in het arrondissement Brussel de twee derden van het produkt verzekert, en de financiële diensten en de handel in het bijzonder die de hoge toegevoegde waarde per eenheid van arbeid tot stand brengt.

Laten we als laatste punt in de vergelijking der niveau's van het binnenlandse produkt de toegevoegde waarde per inwoner beschouwen. Zoals blijkt uit de cijfers vermeld op regel 6 ligt in 1963 het produkt per inwoner in het arrondissement Brussel nagenoeg 50 % hoger dan in de andere taalstreken. T.a.v. de berekening van het B.G.P. per hoofd van de bevolking moet evenwel een voorbehoud worden gemaakt omdat de teller slaat op het produkt dat in een gebied ontstaat ten dele door medewerking van actieve bevolking woonachtig buiten het gebied en ten dele door actieve bevolking van het gebied terwijl de noemer alleen betrekking heeft op de bevolking van het beschouwde gebied. Toch blijkt uit de vergelijking van deze verhoudingen in welke mate er in het arrondissement Brussel een concentratie van produktie-eenheden en economisch potentieel is, buiten verhouding met de bevolking welke er woont.

b. *Ontwikkeling tijdens de jongste jaren.*

Blijft per actieve persoon het produkt in het Vlaamse landsgedeelte nog ruim 6 % beneden dit van Wallonië, dan moet anderzijds worden vastgesteld dat het globale geografisch produkt evenals het geografisch produkt per hoofd van de bevolking sterker stijgt in Vlaanderen dan in het Waalse landsgedeelte. Zowel globaal als per hoofd is de groei het sterkst in het arrondissement Brussel. Het verschil van dit arrondissement met de andere

TABEL 2

*Jaarlijkse groei (in %) van het Bruto Geografisch Produkt
in constante prijzen en enkele verklarende elementen*

	Vlaams lands- gedeelte-	Waals lands- gedeelte-	Arrondiss. Brussel	Het Rijk
1. Groei van het globale B.G.P.				
a) 1955-1963	4,0	2,2	4,3	3,5
b) 1961-1963	5,4	4,1	6,1	5,1
2. Groei van het B.G.P. per inwoner				
a) 1955-1963	3,2	2,0	3,3	2,8
b) 1961-1963	4,5	4,0	5,0	4,5
3. Groei van het B.G.P. per arbeidsuur in de industrie				
a) 1955-1963	4,9	3,6	4,3	4,1
b) 1961-1963	5,6	3,9	2,7	4,4
4. Bruto investeringen in % van het B.G.P., in de industrie voor 1955-1959	13,5	14,1	15,2	14,0
5. Exportgerichtheid v. d. industriële produktie 62-63				
a) Leveringen a. h. buitenl. in % v. d. totale lever.	32,5	28,5	23,9	29,8
b) Aandeel in het Rijks-totaal	50	41	9	100
6. Immigratiesaldo in % v. d. bevolking – 1962-1965	0 à 0,3	0,2 à 0,6	0,7 à 1,5	0,2 à 0,5
7. Natuurlijk accres v. d. bevolking in % – 1962-65	0,7 à 0,8	0,1 à 0,2	0 à 0,1	0,5

landsgedeelten wordt dus steeds groter. Dit blijkt uit de regels 1 van tabel 2. Wallonië kent na de lage groei, voornamelijk wegens de mijnsluitingen, een expansie welke dichter komt bij deze in het Vlaamse landsgedeelte.

De groei van B.G.P. per inwoner vindt men op de regels 2 van tabel 2. De verschillen in groei per hoofd zijn kleiner dan deze in absolute groei. Dit hangt samen met de evolutie van het % actieve bevolking in de totale bevolking. In 1961 was 36,7 % der bevolking actief (werklozen en miliciens niet meegeteld) in Vlaanderen; het overeenkomstig percentage was 34,9 voor Wallonië en 40,7 voor het arrondissement Brussel.

Laten we thans een poging doen om de voornaamste determinanten te achterhalen van de geschetste ontwikkeling. Als directe verklarende veranderlijken kunnen worden beschouwd de arbeidsproductiviteit en de actieve bevolking vermits het globale geografisch produkt kan worden geschreven als het produkt per eenheid van arbeid vermenigvuldigd met het aantal arbeidseenheden. Over de *arbeidsproductiviteit* zijn, zoals reeds vastgesteld, per provincie en taalstreek gegevens beschikbaar voor de industrie. Hieruit volgt dat vooral in het Vlaamse landsgedeelte de toegevoegde waarde per arbeidsuur sterk is gestegen tijdens de periode 1955-1963, nl. met 4,9 % gemiddeld per jaar tegen 3,6 % voor Wallonië, 4,3 % voor het arrondissement Brussel en 4,1 % voor het Rijk. Dit hangt samen met de industriële activiteiten waar de expansie voornamelijk gelocaliseerd was, omdat ook gewoonlijk in de meest expansieve bedrijfstakken de vooruitgang van de arbeidsproductiviteit het grootst is.

De arbeidsproductiviteit op haar beurt is afhankelijk van een aantal factoren, w.o. de hoeveelheid en kwaliteit van de produktiemiddelen of kapitaalgoederen en de intensiteit van de vraag op de afzetmarkten, al is die samenhang niet zo gemakkelijk te meten. De kapitaalgoederenvoorraad is nog niet gekend voor het Rijk, dus ook niet voor onderdelen hiervan. Wel kan men proberen op grond van de gecumuleerde investeringen een benadering hiervan op te bouwen. In het Statistisch Tijdschrift n^o 4, 1962 en Statistische en Econometrische Studiën n^o 3 werden voor de nijverheid de bruto-investeringen in de jaren 1955 tot 1959 gepubliceerd per bedrijfstak en verdeeld naar provincies en taalstreken. Het N.I.S. is voornemens dit onderzoek zodra mogelijk bij te werken tot 1964; vanaf 1962 zal het ook mogelijk zijn de investeringen in machines en materieel ook zuiverder te benaderen door de voertuigen af te zonderen. Toch moet worden vastgesteld dat, mede te oordelen naar de ervaring van andere landen, de macro-economische samenhang tussen groei en investeringen nog niet voldoende nauwkeurig is bepaald.

Het Vlaamse landsgedeelte en in mindere mate ook Wallonië hebben voor hun expansie de jongste jaren, gekenmerkt door een bijzonder sterke buitenlandse vraag, voordeel gehaald uit de *exportgerichtheid* van hun industrie. Op grond van de nijverheidsstatistieken van de jaren 1962 en 1963 werden per bedrijfstak de leveringen aan het binnenland enerzijds en het buitenland anderzijds gedepouilleerd per provincie en taalstreek. Hieruit blijkt dat 32,5 % van de totale industriële produktie van het Vlaamse landsgedeelte rechtstreeks afgezet wordt in het buitenland; voor Wallonië is dit 28,5 %, voor het arrondissement Brussel slechts 23,9 % en voor het Rijk 29,8 %. Volgens genoemde aangiften van de industriële bedrijven (niet te verwarren met de aangiften in de douanestatistiek, waarop de statistiek van de buitenlandse handel is gebaseerd) zou 50 % van de Belgische industriële afzet in het buitenland afkomstig zijn uit de Vlaamse taalstreek, 41 % uit de Waalse en slechts 9 % uit het arrondissement Brussel. In Wallonië leveren de ijzer- en staalnijverheid en de metaalverwerkende nijverheid bijna 70 % van de uitvoergoederen. In Vlaanderen daarentegen is het exportpakket gevarieerder : de metaalverwerkende nijverheid (inclusief scheepsbouw), de weefnijverheid en de chemische nijverheid nemen er samen iets minder dan 60 % voor hun rekening.

We zegden hoger dat de ontwikkeling van het geografisch produkt behalve door de productiviteit ook gedetermineerd wordt door de *evolutie* van de *actieve bevolking*. Zoals eveneens reeds aangestipt zijn over de evolutie van deze bevolking spijtig genoeg voorlopig geen jaarlijkse gegevens beschikbaar voor onderdelen van het Rijk. Wel weet men door de volkstellingen dat voor het Rijk als geheel de actieve bevolking einde 1961 praktisch op hetzelfde peil lag als einde 1947. Sedert 1962 is zij gestegen, hoofdzakelijk door buitenlandse immigratie en dan vooral in 1964 en 1965. Wat de verdeling over de bedrijfstakken betreft moet vooral de zeer belangrijke daling in de landbouw worden vermeld. Deze factor zal vermoedelijk ook nog spelen in de toekomst. Er mag ook worden gerekend met een zekere toename van de deelneming van de vrouwelijke bevolking aan het economisch leven. De stijgende scolariteit anderzijds zal de activiteitsgraad der bevolking blijven drukken.

Keren we nog een ogenblik terug tot de *verhuizingsbeweging*. Sedert 1962 kenden alle taalstreken een immigratieoverschot waar dit voorheen gewoonlijk alleen het arrondissement Brussel het geval was. Tijdens de jaren 1962-1965 overtrof in Wallonië en in het arrondissement Brussel de verhuizingsbeweging merkkelijk de natuurlijke aanwas, zoals blijkt uit de regels 6 en 7 van tabel 2.

3. De Belgische regionale statistiek in de inkomensoptiek.

Men weet dat de toegevoegde waarde welke in een produktie-eenheid en in het geheel der produktie-eenheden van een bepaalde streek ontstaat gelijk is aan de inkomens welke aan de produktiefactoren (arbeid, kapitaal, ondernemersfunctie, ...) ten deel vallen voor hun bijdrage in het produktie-proces. Een goede werkwijze om de regionale verdeling te onderzoeken in de inkomensoptiek zou er dus in bestaan de verschillende soorten inkomens, welke voor het Rijk in het raam van de nationale boekhouding werden berekend, regionaal uit te splitsen. Dit is tot hiertoe voor het N.I.S. nog niet mogelijk gebleken doch staat wel op het programma. Voorlopig zullen we dus wat meer empirisch moeten te werk gaan en ons afvragen over welke gegevens we beschikken om het niveau der inkomens te vergelijken en in zekere mate de ontwikkeling te beschrijven.

De hoger vermelde cijfers van het B.G.P. per actieve persoon en per inwoner kunnen reeds als een aanduiding beschouwd worden van de regionale inkomensverschillen. Spijts een aantal bezwaren, w.o. de eliminatie van de lagere inkomens of de inkomens die wegens gezinslasten niet belastbaar geworden zijn en de forfaitaire belasting van sommige beroepen, lijkt de fiskale statistiek der belastbare netto-inkomens van de natuurlijke personen op dit ogenblik nog als het voornaamste gegeven te moeten gebruikt worden om de inkomensverschillen te meten. Deze statistiek heeft voor het dienstjaar 1964 (inkomen 1963) betrekking op het geglobaliseerde inkomen en desgevallend op het inkomen van man en vrouw samen. Het is wel merkwaardig dat de niveauverschillen per streek van dit belastbare inkomen per aangifte een veel betere overeenstemming vertonen met het B.G.P. per inwoner (dat zoals hoger gezegd met een zeker voorbehoud moet worden gehanteerd) dan met het B.G.P. per ter plaatse tewerkgesteld actieve persoon. Stelt men het belastbare inkomen per aangifte voor het Rijk gelijk aan 100 dan bekomt men 91 voor Vlaanderen, 99 voor Wallonië en 141 voor het nieuwe arrondissement Brussel-hoofdstad.

Behalve uit de fiskale statistiek kan over de inkomensverschillen nog een en ander afgeleid worden uit het gezinsbudgetonderzoek van 1961 (zie Statistisch Tijdschrift n^o 6, 1964 of Statistische en Econometrische Studiën n^o 7). Op grond van dit onderzoek werd het beschikbaar inkomen berekend, d.w.z. het inkomen uit economische activiteit plus het inkomen uit vermogen, plus het inkomen uit sociale voorzieningen minus de belastingen, bijdragen voor sociale voorzieningen en andere inkomensoverdrachten. Het inkomen is berekend per gezin en kan ook omgerekend worden per verbruikseenheid, dit is bij de opstelling van de cijfers van regels 2 van tabel 4 gebeurd

volgens de E.E.G.-schaal. Bedoelde cijfers bevestigen in grote mate de verschillen vastgesteld door middel van het B.G.P. per inwoner en de fiskale statistiek. Bovendien blijkt per verbruikseenheid het inkomensverschil van de Waalse en Vlaamse bedienden groter te zijn dan dit van de Waalse en Vlaamse arbeiders. Dit schijnt bevestigd te worden door de resultaten

TABEL 3
Vergelijking der inkomens volgens diverse bronnen

	Vlaamse taalstreek	Waalse taalstreek	Brussel hoofdstad (1) of vroeger arr.	Het Rijk
1. Belastbaar globaal netto-inkomen – gemiddelde per aangifte, dienstjaar 1964 (inkomen 1963) Index t.o.v. het Rijk	91	99	141 ⁽¹⁾	100
Percentage der aangiften boven 100.000 F	21	27	39 ⁽¹⁾	25
2. Beschikbaar inkomen per verbruiks- eenheid volgens het gezinsbudget- onderzoek van 1961 – Index t.o.v. het Rijk				
Arbeiders	96	99	121	100
Bedienden	93	100	116	100
Niet-actieven	97	90	151	100
3. Totale vergoeding van de arbeid per werknemer (arbeiders en bedienden) volgens de nijverheidsstatistiek 1963 – Index t.o.v. het Rijk	92	113	99	100
4. Bruto-uurloon van de werklieden in de nijverheid (ijzer- en steenkool- nijverheid inbegrepen) Index t.o.v. het Rijk				
a) in oktober 1963	95	107	99	100
b) in oktober 1965	97	106	99	100
5. Verdiensten der mannelijke bedien- den in de handel volgens de R.M.Z.- statistiek 1963 – Index t.o.v. h. Rijk	91	89	106	100
6. Stijging van de reële vergoeding van de arbeid per werknemer (arbeiders en bedienden) in de nijverheid 1955-1963	4,4	4,2	4,6	4,2

vermeld op de regels 3 en 4. De niet-actieven in het arrondissement Brussel (oude omschrijving) zouden een beschikbaar inkomen hebben dat meer dan de helft hoger ligt dan in de andere taalstreken. De Vlaamse niet meer actieven zouden spijs lagere inkomens uit sociale voorzieningen (voornamelijk pensioenen) doch hogere inkomens uit economische activiteit en vermogen een totaal beschikbaar inkomen hebben dat dit van hun Waalse collega's met zowat 7 % overtreft.

Zowel uit de jaarstatistiek van de nijverheid als uit de loonenquête van oktober blijkt dat de lonen en wedden in de nijverheid het hoogst zijn in Wallonië. De wedden van de mannelijke bedienden in de handel daarentegen zouden volgens de R.M.Z.-gegevens van 1963 iets lager liggen in Wallonië dan in Vlaanderen en in het arrondissement Brussel meer dan 15 % hoger zijn dan in de beide andere taalstreken.

Zowel de gegevens over het bruto-uurloon der werklieden (regels 4) als het gemiddelde der jaarlijkse veranderingen (telkens berekend t.o.v. het voorgaande jaar) van de reële lonen en wedden (dit zijn de nominale gedeeld door de prijsindex) wijzen erop dat in de nijverheid de verschillen tussen de Waalse en Vlaamse landsgedeelten kleiner worden; de ontwikkeling in het arrondissement Brussel lijkt eerder divergent.

4. De regionale statistiek in de bestedingsoptiek.

Een belangrijke statistische bron om na te gaan hoe de behoeftebevrediging concreet verloopt is het gezinsbudgetonderzoek. Dergelijke officiële onderzoeken werden uitgevoerd in België op eigen initiatief in 1961 en in E.E.G.-verband in 1963. De hierna vermelde verschillen steunen op het onderzoek van 1961, waarvoor onder regionaal oogpunt de beschikbare informatie vollediger is.

Er werd reeds vastgesteld dat globaal gezien de gezinnen per verbruikseenheid een beschikbaar inkomen hebben dat in de Vlaamse taalstreek lager ligt dan in Wallonië en in deze streek nog heel wat lager dan in het arrondissement Brussel (cfr. lijnen 2 van tabel 3). Desniettemin staande is in Vlaanderen de consumptiequote lager en de spaarquote bijgevolg hoger dan in de andere landsdelen. Uit de lijnen 1 van tabel 4 kan men afleiden dat in 1961 de spaarquote in Vlaanderen en Wallonië respectievelijk was 9,5 % tegen 8,1 % bij de arbeiders, 10,7 % tegen 8,2 % bij de bedienden en 8,1 % tegen 5,9 % bij de niet-actieven. Het verschil is dus het grootst bij het niet meer actieve deel der bevolking en ook groter bij de bedienden dan bij de arbeiders ofschoon het inkomensverschil bij de bedienden het grootst is.

Zoals te verwachten op grond van de inkomensverschillen liggen, voor de drie sociale groepen opgenomen in het budgetonderzoek, de uitgaven in het arrondissement Brussel voor het merendeel der uitgavencategorieën merkelyk hoger dan in de twee andere taalstreken. Vrijwel hetzelfde geldt voor Wallonië t.a.v. het Vlaamse land.

In overeenstemming hiermede is dat in Vlaanderen voor alle sociale groepen meer uitgegeven wordt voor *brood* en andere produkten op basis van graangewassen. Bij de arbeiders blijkt het *vlees*verbruik gelijk te zijn in Vlaanderen en in Wallonië; bij de Waalse bedienden is het evenwel merkelyk hoger dan bij hun Vlaamse collega's terwijl de Vlaamse niet meer actieven wat meer uitgeven dan de Waalse. Het arrondissement Brussel ligt steeds hoger.

Het verbruik van *boter* is in Wallonië veel hoger dan in Vlaanderen (ca. 40 % voor de arbeiders en bedienden, ca. 30 % voor de niet-actieven); het overtreft in eerstgenoemde streek ook merkelyk (15 à 20 %) dit in het arrondissement Brussel. Voor *groenten en fruit, dranken en tabak* geven de Waalse arbeiders zowat 10 % meer uit van de Vlaamse; bij de bedienden bereikt het verschil ruim 20 % voor gealcoholiseerde dranken en alcohol. In het arrondissement Brussel ligt het verbruik voor genoemde produkten nog veel hoger dan in Wallonië.

De enige uitgavenrubriek waar de Vlaamse taalstreek voor de drie sociale categorieën hoger komt dan de Waalse is « *kleding, schoeisel en ander persoonlijk goed* »; het verschil is van de orde van 10 %. Ook hier stelt men in het arrondissement Brussel een veel aanzienlijker verbruik vast dan in beide andere streken.

De op de regels 4 van tabel 4 vermelde uitgaven voor *buisvesting* slaan op huur (werkelijke betaalde en toegerekende), water, verwarming en verlichting, duurzame huishoudartikelen en onderhoud. Ook hier zijn de uitgaven in Wallonië groter dan in Vlaanderen. De verklaring van de afwijking ligt vooral in de verwarming, waarvoor de Waalse gezinnen heel wat meer uitgeven dan de Vlaamse. Dit wordt bevestigd door het groter percentage woningen met centrale verwarming. Zeer opvallend is evenwel het grote verschil in de bouw van nieuwe huizen. 72 % van de nieuwe of wederopgebouwde particuliere huizen werden in 1964 gebouwd in het Vlaamse land; slechts 24 % in Wallonië en 4 % in het arrondissement Brussel-hoofdstad.

Gaan we nu over naar de rubriek « *persoonsverzorging en hygiëne* » die de uitgaven omvat voor persoonsverzorging (zeep, tandpasta, kapper, reukwerk), geneeskundige zorgen, farmaceutische en para-farmaceutische produkten. Het verschil is klein bij de arbeidersgezinnen. Het is veel groter

bij de bedienden en niet meer actieven. Toch zijn de afwijkingen op het vlak van de uitgaven klein vergeleken bij de verschillen in medisch personeel. Per 10.000 inwoners waren er in 1964 10 geneesheren in de 4 Vlaamse provincies, 15 in de 4 Waalse en meer dan 20 in Brabant. Het aantal beschikbare bedden in de verzorgingsinstellingen daarentegen is 40 % hoger in de Vlaamse provincies dan in de Waalse terwijl Brabant een middenpositie inneemt.

Voor aankoop en gebruik van *eigen vervoermiddelen* gaf de Waalse arbeider in 1961 ca. 75 % meer uit dan de Vlaamse. Bij de bedienden was het verschil 37 %. Gans anders lag de toestand evenwel bij de niet meer actieven : het Vlaamse gezin gaf 88 % meer uit dan het Waalse. Voor het geheel der bevolking lijken deze resultaten vrij goed in overeenstemming met het aantal personenwagens per 100 inwoners in de drie taalstreken in 1962, nl. 8,4 in de Vlaamse streek, 10,7 in de Waalse en 13,4 in het arrondissement Brussel. Wat het aantal nieuwe wagens betreft was in 1964 de achterstand bijna ingehaald.

Uit de algemene volkstelling blijkt dat einde 1961 in Vlaanderen en Wallonië nagenoeg eenzelfde fractie der bevolking volledig *onderwijs* van het secundair niveau had gevolgd. In Wallonië liep evenwel een groter gedeelte middelbaar onderwijs, in Vlaanderen een groter gedeelte technisch onderwijs. Het aantal gediplomeerden van het hoger secundair niveau lag in het arrondissement Brussel evenwel twee derden hoger dan in de twee andere taalstreken. De verhoudingen liggen nagenoeg hetzelfde voor het niet universitair hoger niveau. Op het vlak van het universitair onderwijs evenwel telde de Vlaamse taalstreek een aantal gediplomeerden dat, voor eenzelfde aantal inwoners, 20 % lager lag dan in Wallonië. In het arrondissement Brussel is het aantal academisch geschoolden 2,5 maal zo groot als in Vlaanderen en ruim 2 maal zo groot als in Wallonië.

De regels 7 van tabel 4 betreffende de nog schoolgaande bevolking bevestigen dat in Wallonië en in het arrondissement Brussel de jeugd tussen 12 en 18 jaar meer, resp. veel meer middelbaar onderwijs volgt dan in Vlaanderen. Het omgekeerde is waar voor het beroeps- en secundair, technisch onderwijs. Voor de twee soorten onderwijs samen was de scolariteitsgraad in Vlaanderen op weinig na dezelfde, nl. 62 %; in het arrondissement Brussel was hij evenwel 71 %. Bijna 4 % van de bevolking van de passende leeftijd volgde lager of middelbaar normaal onderwijs in Vlaanderen; in het arrondissement Brussel was dit ruim 4,5 % en in Wallonië 5,5 %. Zowel in het hoger technisch als in het universitair onderwijs had het Vlaamse landsgedeelte in 1961 nog ca. 10 % jongeren van 18 tot 22 of 23 jaar minder in opleiding dan Wallonië. Het arrondissement Brussel

TABEL 4
Welvaartverschillen volgens het gezinsbudgetonderzoek van 1961 en een aantal andere indicatoren

	Vlaams lands- gedeelte	Waa's lands- gedeelte	Arrond. Brussel of Brussel hoofdstad (1)	Rijk
1. Totale consumptieuitgaven per verbruikseenheid				
a) In % van het beschikbaar inkomen				
Arbeiders	90,5	91,9	91,5	91,3
Bedienden	89,3	91,8	92,8	91,1
Niet-actieven	91,9	94,1	90,6	92,6
b) Index t.o.v. het Rijk				
Arbeiders	95	99	122	100
Bedienden	90	101	119	100
Niet-actieven	96	91	145	100
2. Uitgaven voor voeding per verbruikseenheid (Index t.o.v. het Rijk)				
Arbeiders	95	102	112	100
Bedienden	92	105	109	100
Niet-actieven	95	98	127	100
3. Uitgaven voor kleding, schoeisel en ander persoonlijk goed, per verbruikseenheid				
Arbeiders	101	89	117	100
Bedienden	101	93	108	100
Niet-actieven	100	89	116	100
4. Huisvesting				
a) Uitgaven per verbruikseenheid in 1961				
Arbeiders	95	100	129	100
Bedienden	89	101	125	100
Niet-actieven	95	92	159	100
b) Aantal particuliere huizen in 1964	55,1	37,9	(1) 7,0	100
c) Nieuw gebouwde of wederopgebouwde particuliere huizen in 1964	72,1	23,9	(1) 4,0	100
d) Percentage der woningen met centrale verwarming einde 1961	7,3	8,1	(1) 30,3	100
e) Percentage der woningen met badkamer of priv. stortbad einde 1961	21,4	20,8	(1) 38,8	100

5. <i>Persoonsverzorging en hygiëne</i>				
a) Uitgaven per verbruikseenheid in 1961 (Index)				
Arbeiders	96	98	123	100
Bedienden	81	102	133	100
Niet-actieven	90	98	152	100
b) Medisch personeel per 1.000 inwoners in 1964 voor resp. de 4 Vlaamse prov., de 4 Waalse prov. en Brabant				
Geneesheren	1,03	1,45	2,12	1,40
Tandartsen en lic. tandheelkunde	0,12	0,16	0,28	0,17
Apothekers	0,42	0,82	0,77	0,62
Aantal bedden in verzorgingsinstellingen, per 1.000 inwoners in 1960	5,12	3,64	4,26	4,47
6. <i>Vervoer met eigen vervoermiddelen</i>				
a) Uitgaven per verbruikseenheid in 1961 (Index)				
Arbeiders	75	131	111	100
Bedienden	81	111	119	100
Niet-actieven	124	66	91	100
b) Totaal aantal personenwagens in 1962 per 100 inwoners	8,4	10,7	13,4	10,0
c) Aantal nieuwe personenwagens aangekocht in 1964 per 100 inwoners	2,0	2,0	(1) 3,2	2,1
7. <i>Schoolbevolking (einde 1961)</i>				
- Middelbaar onderwijs	30,6	35,0	47,4	34,2
- Beroeps- en secundair technisch onderwijs (in % van de bevolking van 12- < 18 jaar)	30,9	27,0	23,5	28,6
- Lager normaal onderwijs (in % van de bevolking van 15- < 19 jaar)	4,0	5,8	5,0	4,7
- Middelbaar normaal onderwijs (in % v. d. bevolk. v. 19- < 21 jaar)	3,7	5,2	4,2	4,2
- Hoger techn. en Kunstonderwijs (in % v. d. bevolk. v. 18- < 22 jaar)	4,4	4,8	6,8	4,8
- Universitair en gelijkg. onderwijs (in % v. d. bevolk. v. 18- < 23 jaar)	5,4	6,0	11,8	6,5

telde voor de twee vormen van hoger onderwijs zelfs 70 à 100 % studenten meer dan Vlaanderen en Wallonië.

5. Slotbeschouwingen.

In plaats van een opsomming te geven van de bestaande statistiek op het regionale plan hebben we getracht ze in elkaar te passen tot een economisch en sociaal overzicht van onze taalgemeenschappen. Dit bood ons de gelegenheid om de voornaamste gegevens te vermelden. Een aantal evenwel moest onvermijdelijk onbesproken blijven. Het is duidelijk dat ook de methode welke bij onze bespreking werd gevolgd op vele punten kan verbeterd worden.

De reeds aangestipte leemten in onze regionale statistiek hebben betrekking op de indeling van de bevolking naar leeftijd en activiteit. Zij zullen zo spoedig mogelijk moeten aangevuld worden. Zoals dit reeds gebeurd is met het binnenlandse produkt moeten ook de inkomens en de nationale bestedingen regionaal uitgesplitst worden, om te komen tot een nationale boekhouding op regionaal niveau. Concreet betekent dit dat de verschillende soorten factorinkomens en hun grote onderdelen evenals de private consumptie, de overheidsconsumptie, de investeringen en het saldo van de betalingsbalans regionaal moeten verdeeld worden. Verschillende der hogergenoemde statistische bronnen zullen daartoe kunnen aangewend worden. Allerlei aanvullende onderzoeken zullen evenwel nodig zijn, o.a. in verband met de onderlinge economische relaties der streken.

De vraag kan worden gesteld of in een perspectief van enkele jaren ook reeds het opstellen van regionale input-outputrekeningen moet worden betrokken. Is hun nut voldoende groot opdat daartoe vrij spoedig op het vlak van de officiële statistiek de vereiste middelen zouden moeten worden gevraagd of is dit veeleer een terrein dat aan regionale onderzoekingscentra kan worden overgelaten? Wat het antwoord ook zij, de opstelling van nationale rekeningen op regionaal niveau is een werk dat eerst dienst te gebeuren.

REFLEXIONS EN VUE D'UNE VENTILATION REGIONALE DES STATISTIQUES

par E. NOLS
L'Equerre, Liège

On m'a demandé d'examiner quelles pourraient être les bases d'une présentation régionalisée de la statistique belge. Encore que la réponse à donner à cette question puisse varier en fonction du type de statistiques à ventiler, je me suis volontairement placé dans une optique générale.

Si la solution que je suggérerai pourra paraître simple, sinon simpliste, je dois à la vérité de dire qu'elle m'a été dictée par quelques considérations de divers ordres que je voudrais évoquer succinctement.

Depuis près de vingt ans que je m'occupe de régionalisme, j'entends discuter de la définition et de la délimitation des régions, chacun ayant sa méthode infailible, son découpage irrécusable; le tout donnant lieu à de sérieuses empoignades. Pour corser davantage le débat, la politique s'en est mêlée. Ceux d'entre nous qui ont suivi la préparation des lois sur l'aménagement du territoire et la décentralisation économique et administrative savent à quel point la question éveille les susceptibilités, — électorales et autres, — révèle les rancunes de clocher, attise les appétits. Notre pays, malgré la coloration linguistique qu'il ne peut se passer d'ajouter à ce genre de problème, n'a pas seul l'apanage de ces querelles byzantines où la science finit toujours par céder le pas aux préoccupations des groupes; le phénomène a été tout aussi marquant en France et il est très possible qu'on le retrouverait en d'autres pays.

Il faut d'autre part remarquer que l'attitude devant la délimitation des régions est fondamentalement différente dans le chef de l'économiste, du géographe et du géonome.

Raymond BARRE note (1) * que « l'économiste est moins soucieux que le géographe des différenciations qualitatives réalisées par l'homme sur l'espace; il s'intéresse à l'espace organisé en vue du rendement le plus élevé et du coût le plus faible; il définit un espace économique à partir de plans établis par les sujets économiques, comme le périmètre d'action et d'influence d'une unité économique, comme le système de relations qui constituent la vie

* Les notes infrapaginales et bibliographiques sont reportées en fin de texte.

de cette unité ». Citant Jeanneney, Barre poursuit : « Le rôle du géographe me paraît être d'étudier les phénomènes économiques dans leurs rapports avec les données physiques (naturelles ou créées par l'homme) localisées. Le rôle de l'économiste est de découvrir le jeu des mécanismes économiques et de mesurer leurs effets. L'espace du géographe est sillonné de lignes de partage inscrites par la nature et l'homme sur le sol même; celui de l'économiste a pour frontière les limites d'action des flux fondamentaux distingués par la théorie économique » (2).

Aussi me rallierai-je assez facilement à la thèse défendue, voici quelques années, par Léon DERWA et selon laquelle les régions économiques sont fondamentalement indéterminées (3). Sans que cela n'entame en rien la démonstration de Derwa, BOUDEVILLE qui, comme lui d'ailleurs, se fonde sur les travaux de PERROUX, a depuis lors, apporté quelques précisions à ce point de vue en distinguant *espace* et *région* économiques, concepts qui ne sont nullement synonymes.

Pour Boudeville, l'espace économique, — dont les régions-plans ou régions de programme sont l'illustration concrète, pratique, — est l'application d'un espace mathématique, c'est-à-dire d'un modèle sur ou dans un espace géographique. L'espace économique est discontinu alors que la région économique est, elle, continue; elle se différencie de l'espace par une contrainte de contiguïté dont la justification est d'être une réalité observable (*in natura rerum*) (4).

Aussi l'économiste, s'il a besoin, sans conteste possible, d'un espace géographique pour situer ses instruments d'analyse et ses modèles opérationnels, a moins le souci du périmètre qu'on lui propose que le géonome, par exemple, dont la démarche conduit à des implantations, à des modifications de ligne de partage, à la création de zones, bref à une série d'opérations qui s'inscrivent directement sur le sol et qui sont directement fonction, quantitativement et qualitativement, de l'espace retenu. Il appartiendra d'ailleurs souvent à l'économiste d'en modifier le contour en vue d'optimiser la solution des problèmes qu'on lui proposera.

* * *

Un autre point de vue à verser au dossier concerne la liaison qui s'établit entre la manière de concevoir une région et la délimitation qu'on lui impose (5). S'il s'agit d'une *unité géographique*, elle se détermine par des caractéristiques physiques : géologiques, orographiques, hydrographiques, etc. S'il s'agit d'une *unité administrative*, elle est considérée comme le champ d'action d'une administration décentralisée ou d'un pouvoir subordonné. On l'entend encore comme une aire de *similitude relative de structure écono-*

mique basée sur l'uniformité du plus grand nombre possible de caractéristiques (revenu par habitant, degré d'industrialisation, répartition de la population active, etc.). Enfin, comme *une unité spatiale de vie sociale* déterminée par la densité des relations avec une ville-centre ou avec un pôle; c'est ce que Perroux appelle *la région polarisée*, support naturel de la *région-plan* et qu'il détermine en mesurant l'intensité des flux économiques échangés avec le pôle principal (6).

D'ailleurs, il est plausible de déterminer à l'intérieur d'un même espace géographique plusieurs périmètres correspondant chacun soit à une fonction à conférer à la région, soit à des problèmes internes à résoudre. Ainsi, en considérant Liège comme une *métropole régionale*, au sens où F. GRAVIER, par exemple, entend cette expression, est-on amené à y distinguer une zone à dominance urbaine, une zone à la fois urbaine et industrielle, un périmètre d'agglomération, une zone d'expansion et d'espaces verts à préserver. A l'inverse, on peut définir la Tennessee Valley comme une région correspondant aux limites du bassin hydrographique lorsque l'attention se porte sur les flux du système des eaux, puis de modifier et d'élargir cette limite lorsque l'attention retient les flux énergétiques distribués par les centrales hydro-électriques qu'administre la T.V.A. (7).

W. ISARD partage également l'opinion selon laquelle les régions varient suivant le problème envisagé et sont, en ce sens, uniques (8).

* * *

Il faut en outre tenir compte que les limites d'une région sont appelées à varier dans le temps soit que les structures se modifient, soit que les échanges s'amplifient, se restreignent ou que de nouveaux flux apparaissent. En 1948, déjà SUAVET écrivait : « Tandis que les unités de la géographie physique ne changent qu'avec les siècles et celles de l'administration à un rythme aussi variable qu'imprévisible, les unités réelles de vie humaine, modelées par le progrès technique, industriel et agricole et surtout des transports, sont en constante évolution et il est plus important encore de déterminer le sens de cette évolution que les limites à un instant donné de telles ou telles unités » (9). VIDAL DE LA BLACHE a émis un point de vue identique remarquant que les fonctions régionales sont soumises aux lois de l'évolution, « qu'elles se défont et se recréent suivant les changements qui se produisent dans les relations des hommes ».

Aussi, à moins d'être nanti d'un solide dogmatisme, peut-on difficilement admettre que le découpage d'une nation puisse convenir quelle que soit l'optique où l'on se place et puisse être immuable. Si la division de la France en départements signifiait quelque chose au temps des diligences (10), il

ne peut plus convenir à l'âge de l'automobile et de l'aménagement du territoire. Cependant, ce pays n'a pu encore se résoudre à remodeler ses divisions et s'est contenté d'organiser ses régions en conservant la base départementale. En 1957, le Conseil économique et social émettait à ce sujet un avis de prudence qu'il n'est certes pas sans intérêt de rappeler en l'occurrence (11) :

« Il serait peu raisonnable, disait le Conseil, surtout dans nos incertitudes actuelles, de fixer définitivement, à coups de décisions réglementaires, un découpage régional que les forces économiques ne tarderaient pas à disloquer. Plus qu'un territoire, c'est un ensemble de problèmes communs qui déterminent la région, et la nécessité où nous sommes d'inscrire nos actions dans un cadre spatial ne doit en aucun cas constituer un frein ou un obstacle à l'expansion régionale. Il convient donc de recommander en matière de découpage régional (par exemple, pour l'élaboration de programmes régionaux) la plus grande prudence et la plus grande souplesse. Si les découpages retenus pour le 3^{me} plan soulèvent des difficultés, il faudra le modifier par la suite. Il serait souhaitable d'éviter tout malentendu au départ en déclarant nettement que l'étendue territoriale des programmes régionaux actuels n'engage pas l'avenir ».

Comme l'administration joue un rôle important dans l'expansion régionale, les Français ont retenu les circonscriptions administratives pour délimiter les cadres d'application des mesures préconisées. Là où les limites administratives ne coïncident pas, ils ont choisi le tracé qui était commun au plus grand nombre de circonscriptions (12).

Comme on le sait, les 90 départements français ont été regroupés en 21 régions de programme. Ce regroupement est, comme bien l'on pense, contesté; certains l'admettent avec cette nuance que les 21 régions précitées sont des *régions de programme à moyen terme* et qu'il faudrait les regrouper en 9 régions, polarisées chacune par une *métropole régionale pour les programmes à long terme* : « ce sont les régions-plans à long terme dont les bassins fluviaux et la conception des métropoles régionales sont deux exemples privilégiés », dit Boudeville (13).

Après ces dernières remarques, c'est sans doute l'instant d'indiquer qu'à notre sens la province belge peut suffire pour l'application d'un programme quinquennal; elle est, par contre, trop étroite pour la conception et la réalisation de programmes à long terme, d'autant que quatre de nos provinces, au moins, ne comptent aucun centre urbain qui puisse revendiquer avant longtemps le rôle assigné aux métropoles régionales. Ceci dit par parenthèse.

Le pragmatisme dont ont fait preuve les Français en conservant leurs unités administratives comme base de ce groupement régional, ce pragmatisme n'est pas unique. Il semble tout au contraire qu'il soit l'attitude qu'aient

adoptée tous les Etats promouvant des actions régionales. En voici quelques exemples :

— *Aux Etats-Unis* : les études et les actions ont porté sur de grandes régions telles que le Sud, la Nouvelle-Angleterre, l'Upper Midwest (partie septentrionale des plaines du Middlewest); d'autres sur des Etats, par exemple la Californie et le Colorado; des régions plus restreintes ont été assimilées soit à un « county » (Sacramento, en Californie, Leadville-Lake au Colorado) ou sur des groupes de « counties » appartenant à un ou plusieurs Etats; enfin des villes et des zones métropolitaines ont également été choisies comme support à l'étude et à l'action régionale; c'est d'ailleurs à cet échelon que l'analyse régionale paraît avoir atteint sa diffusion le plus large et connu les applications théoriques les plus poussées (14);

— *Au Royaume-Uni* : si l'on excepte l'Irlande du Nord qui est considérée dans son ensemble comme une région à développer, la politique régionale se fonde sur les comtés et les municipalités de comtés; ce sont ces aires qui sont retenues pour la délimitation des zones déprimées à reconvertir ou à développer;

— *En Allemagne occidentale* : encore que l'entité retenue ne soit pas clairement définie, c'est le « kreiss » ou canton qui est pris comme unité d'étude et d'action régionales (15); les programmes portent sur un nombre plus ou moins élevé de cantons;

— *Aux Pays-Bas* : l'étude et l'action s'organisent indifféremment au départ des districts de province ou d'un groupe de communes;

— *En Italie* : le Mezzogiorno est limité, au Nord, par la frontière du « Midi historique », c'est-à-dire les anciennes frontières du Royaume des Deux Siciles et des Etats de l'Eglise (16); à l'intérieur de cet ensemble, des zones plus homogènes ont été déterminées dont l'étendue est toujours inférieure à celle des régions traditionnelles (par exemple, dans les Abruzzes) (17);

— *En Espagne* : ce sont soit des provinces, soit des bassins hydrographiques (pour les problèmes d'irrigation) qui sont adoptés;

— *En Yougoslavie* : les échelons vont de l'Etat fédéré à la commune selon la nature, la portée et la durée du programme à mettre en œuvre.

* * *

La statistique a-t-elle épousé cette grande diversité des entités administratives retenues pour les études et l'action régionales ?

D'après les sources statistiques que j'ai pu consulter, je relèverai que c'est généralement la subdivision administrative majeure qui a été adoptée : l'Etat aux Etats-Unis, le département en France, le comté en Grande-Bretagne, le Länder en Allemagne. Ce n'est que dans des études particulières ou dans les publications des pouvoirs régionaux que l'on trouve des données davantage

régionalisées. Encore ces dernières ne concernent-elles généralement que la démographie, l'emploi, le chômage et, parfois, les transports.

En me fondant sur ces quelques réflexions, je vais à présent tenter de répondre à la question qui m'a été posée, savoir sur quelles bases organiser la statistique belge. Il ne s'agit donc pas, j'y insiste, de présenter un découpage opérationnel ni de suggérer les bases d'une action régionale. Il s'agit de rechercher une présentation de nos statistiques qui puisse utilement servir aux chercheurs et qui puisse aussi, — cet aspect est également important, — être élaborée dans des délais raisonnables.

Lorsqu'on examine, avec un certain détachement des contingences socio-politiques, l'articulation de l'économie belge, il apparaît assez rapidement qu'elle s'organise autour de cinq pôles, cinq points forts animant des régions plus ou moins vastes, ces dernières étant particulièrement malaisées à délimiter parce que maintes d'entre elles se recoupent, se chevauchent. Il semble d'ailleurs qu'en ce qui les concerne la délimitation à retenir serait davantage *zonale* que *linéaire*.

Les cinq pôles qui, eux, tranchent nettement sur l'ensemble sont : Bruxelles, Anvers, Liège, Gand et Charleroi. Il s'agit, dans chaque cas, de complexes industriels et urbains, centrés sur une ville où les activités tertiaires se sont développées à des degrés divers, parfois en symbiose plus ou moins étroite avec les besoins de l'industrie ou de l'économie nationale, et aussi selon l'intensité des besoins quantitatifs et qualitatifs des populations regroupées dans ces complexes.

Par l'implantation des activités qui les animent et les dimensions des agglomérations qu'ils ont suscitées, il me paraît qu'en *première analyse du moins* ces pôles peuvent s'identifier, sans trop d'inconvénients pour leur appréhension et leur analyse, avec les arrondissements qui portent leur nom. Aussi proposerai-je que cette entité administrative soit retenue comme base géographique des statistiques qui les concernent. Les mouvements migratoires, les mouvements pendulaires de la main-d'œuvre, la densité de la population, la localisation de la majeure partie des centrales électriques, la concentration des activités tertiaires sont autant de facteurs qui indiquent, sans conteste, que ces arrondissements jouent dans l'économie nationale un rôle privilégié. En sorte que la compréhension de cette dernière, l'analyse de ses faiblesses, la détection des grandes tendances socio-économiques qui suscitent son évolution ne peuvent guère être saisies totalement si on ignore comment se comportent les régions majeures sur lesquelles cette économie s'appuie. D'autre part, si la politique économique régionale s'est, jusqu'à présent, en Europe du moins, surtout attachée aux régions retardées ou en déclin, il ne paraît

pas raisonnable qu'elle néglige les problèmes de croissance ou de décélération que connaissent ces régions majeures, les zones motrices elles-mêmes dont il est bien connu que leur évolution conditionne celle d'espaces débordant largement leur propre périmètre. C'est un point qu'il importe de souligner : hormis Bruxelles, il est très malaisé de suivre à travers nos statistiques l'état de santé des régions de Liège, Anvers, Gand et Charleroi. C'est un problème aisé à résoudre; je crois qu'il mérite d'être examiné et d'être rapidement élucidé.

Le problème est beaucoup plus complexe si l'on cherche à délimiter les zones polarisées par ces centres moteurs. On se trouve en présence de phénomènes d'attraction et de relations socio-économiques dont l'expression simplifiée pourrait être ébauchée au départ soit de la loi de Reilly évoquant la formule physique de la gravitation, soit sur les travaux de Harris associant les critères d'accessibilité et les coûts de transport. Encore que la délimitation de ces zones varie, nous l'avons vu, selon le point de vue où l'on se place, elle est, en outre, appelée à se modifier dans le temps. Enfin, il y a fort à parier qu'aucune délimitation préconisée ne fasse jamais l'unanimité, soit toujours contestée et contestable à l'un ou l'autre égard.

Je ne vois donc pas que l'on puisse articuler notre appareil statistique sur des bases géographiques aussi fragiles, aussi mouvantes, aussi incertaines. Aussi serai-je enclin à suggérer qu'en ce domaine l'on se contente du *statu quo ante*, à savoir que la *province*, dont les limites sont quand même moins fluctuantes malgré certains rajustements récents, reste la base de la ventilation géographique de nos statistiques; il n'y aurait en cela aucune singularité : les Américains, malgré leur abondance statistique et les moyens dont ils disposent, se contentent des données relatives à leurs Etats.

Je n'ignore pas les inconvénients de cette solution. Les provinces sont des agglomérats géographiques, démographiques et économiques hétérogènes; aucune étude sérieuse ne peut se contenter de les considérer comme des entités indifférenciées. Même nos programmes quinquennaux d'expansion économique régionalisés y trouveraient tout juste une base valable de réflexion et d'action : le cadre est soit trop grand, soit trop petit pour l'application d'une politique concertée de développement et d'aménagement du territoire à long terme.

Si, malgré ces inconvénients, je préconise la base provinciale, c'est qu'on peut y apporter deux correctifs. Tout d'abord, on pourrait aisément la compléter par un regroupement effectué selon les régimes linguistiques comme cela se fait déjà pour les valeurs ajoutées et le produit intérieur brut. En second lieu, des répartitions plus limitées dans l'espace pourraient être tout aussi aisément élaborées pour répondre aux besoins d'études particulières :

nos statistiques étant collectées, soit à l'échelon de la commune, soit à l'échelon des établissements, il n'y a guère de difficulté à les ajuster à des entités géographiques plus spécifiques. Pour autant bien sûr qu'on ne se heurte pas au sacro-saint secret statistique, — qu'on n'évite même pas à l'échelon provincial, — et que l'Institut national dispose du personnel nécessaire.

* * *

En sorte que, pour me résumer, je crois pouvoir suggérer les ventilations suivantes :

- les arrondissements de Bruxelles, Liège, Anvers, Charleroi et Gand;
- les neuf provinces;
- les régions linguistiques flamande et wallonne.

L'I.N.S. devrait en outre nous promettre de fournir, comme il le fit par le passé, des séries statistiques relatives à des entités géographiques ne correspondant pas aux limites provinciales ni d'arrondissement lorsque des études les requièrent.

Je crois que tant que les pouvoirs publics n'auront pas pris position sur la définition des régions belges, — et encore faudra-t-il être prudent, — je crois qu'il ne serait pas réaliste de prendre une autre attitude.

* * *

- (1) R. Barre : *Economie politique*, t. I, p. 17, Paris 1964.
- (2) J.M. Jeanneney in *Revue de Géographie alpine*, 1955, t. XLIII, fasc. I, p. 222.
- (3) L. Derwa : *L'indétermination fondamentale des régions économiques*, in *Mélanges Graulich*, Liège 1959, pp. 577 et suivantes.
- (4) J.R. Boudeville : *Les notions d'espace et d'intégration*, actes du Congrès de Bâle sur l'aménagement des territoires (octobre 1965).
- (5) A. Schreurs : *Contribution aux études d'économie régionale en Wallonie*, in *Revue des Sciences économiques*, juin-sept. 1954. - I. Divoy : *Note sur une politique économique régionale*, sept. 1958 (étude stencillée).
- (6) W. Leontief la détermine au départ des échanges extérieurs, ce qui paraît moins probant.
- (7) J. Libert : *Le récent développement de l'analyse régionale aux Etats-Unis* (étude publiée par le Conseil central de l'Economie).
- (8) W. Isard : *Location and Space-Economy*, John Wiley, New York 1956. - *Regional Science, the concept of Regional structure* - Regional Science Association, Papers and Proceedings, 1956.
- (9) Th. Suavet : *Comment déterminer les unités territoriales*, Economie et Humanisme, mars-avril 1948, p. 158.
- (10) On devait pouvoir aller en une journée de diligence du chef-lieu à tous les points du département.
- (11) Avis du 22 janvier 1957.
- (12) Décret 55-873 de juin 1955 relatif à l'établissement des programmes régionaux.
- (13) J.R. Boudeville, *loc. cit.*
- (14) J. Libert, *op. cit.* pp. 37 et 38.
- (15) P. Romus : *Expansion économique régionale et communauté européenne*, Leyde 1958, pp. 159 et 160.
- (16) Ibidem, p. 126.
- (17) *Planification économique régionale*, actes du Congrès de Bellagio, A.E.P.-O.E.C.E. p. 80.

REGIONAAL-ECONOMISCHE MODELLEN EN DE EISEN DIE ZE STELLEN AAN DE STATISTIEK

door J.F. van WATERSCHOOT
Katholieke Universiteit te Leuven

1. — Inleiding.

1.1. We onderstellen dat niet hoeft stil gestaan bij de definitie van het universeel denkinstrument, dat we een *model* noemen. Het gaat in onze opvatting om een simultaan stelsel van symbolische relaties, die de samenhang van de economische problematiek weliswaar vereenvoudigd, maar toch zo getrouw mogelijk weergeven. Een *regionaal model* beoogt dit te realiseren t.o.v. een regionaal-economische problematiek, waarbij de *endogene of te verklaren elementen* vooral een regionaal karakter zullen hebben, terwijl de *exogene of verklarende grootheden* vooral slaan op de grotere geografische eenheid (nationaal of interregionaal), waarin het gewest gesitueerd word.

1.2. Het voorafgaande onderscheid, dat bij alle modellen dient gemaakt tussen de *wiskundige* en *econometrische* modellen, wordt ook hier toegepast. Bij de wiskundige modellen worden de relaties wiskundig gespecificeerd tot een simultaan samenhangend denkschema. Deze modellen kunnen (en hebben in het verleden) grote diensten bewezen, daar waar het erom gaat het economisch inzicht te verbeteren. In dit opzicht werd reeds heel wat weg afgelegd van de eerste, ruwe Keynesiaanse tot de huidige, zeer gedetailleerde macro- en micro-economische modellen.

In het *econometrisch* model wordt de deductie, die in het wiskundige model wordt beoefend, gecombineerd met de waargenomen werkelijkheid ten einde een operationeel schema te verkrijgen, op basis waarvan rationele beslissingen kunnen getroffen worden. De coëfficiënten die in de econometrische modellen voorkomen, dienen te worden *geschat* op basis van de statistische *waarnemingen*.

1.3. Hier duikt dan het *probleem van de schattingsmethoden* op, dat de jongste jaren meer aandacht verworven heeft dan welk ander statistisch probleem ook ! Hierbij dient vastgesteld dat de econometrie bevruchtend heeft ingewerkt op de ontwikkeling van de wiskundige statistiek, o.m. dank zij de methodologische strijd die zich toespitst op de Kleinste Kwadratenmethode, op de « Maximum Likelihood »-methode en nu ook, als « Dritte im Bunde » op de Bayesiaanse benadering.

Nochtans kan men zich afvragen of daardoor niet te weinig aandacht werd besteed aan eensdeels de specificatie der relaties (bijv. onze allergie t.o.v. niet-lineaire relaties) en anderdeels aan de beschikbaarheid der statistische gegevens en hun kwaliteit.

1.4. Principieel dienen de data zich natuurlijk te ontwikkelen naar het model en de ervaring heeft uitgewezen, bijv. bij de nationale programmatie, dat het model de vraag naar statistische gegevens systematisch en samenhangend maakt.

In de praktijk zal men nochtans realistisch moeten zijn en zal men bijv. bij de keuze tussen twee nagenoeg gelijkwaardige modellen de voorkeur geven aan het model dat statistisch realiseerbaar is.

De gunstige wisselwerking tussen model en statistische data kan vastgesteld worden op het regionale vlak. Zonder dat men in België formele regionale modellen heeft uitgewerkt (met uitzondering van input-output tabellen voor sommige agglomeraties), heeft men toch reeds een aanloop daartoe genomen. Een stimulerende kringloop ontstond met de inspanningen van het N.I.S. tot berekening van de bruto geografische produkten (1), waardoor ons land over meer gedetailleerde regionale produktiedata beschikt dan de partnerlanden in de E.E.G.

2. — Wiskundige modellen.

2.1. Wiskundige modellen hebben het voordeel dat ze, los van wat op een gegeven ogenblik statistisch mogelijk is, analyserend kunnen gebruikt worden. Na verloop van tijd blijkt zulks bijzonder nuttig te zijn om een grondiger inzicht te bekomen in de regionaal-economische problematiek die als het ware door de werkelijkheid werd opgedrongen en daardoor nog steeds in overwegende mate een empirisch, op actie gericht karakter heeft.

Het zou ons te ver voeren, indien we alle modeltypes met regionale inslag zouden overlopen. Daarom worden slechts enkele principële gedachten uiteengezet en enkele selectieve voorbeelden gekozen.

Het ware ideaal indien men over « regionale rekeningen » kon beschikken om de *economische fluxen of stromen intra* (in de streek) in *inter* (met andere streken) te meten. R. Stone en W. Isard o.m. hebben hieromtrent heel wat geschreven en gepresteerd. Vanuit een Belgisch standpunt gezien zou een betere kennis van de mensenstromen (pendelbeweging) tussen de regio's (provincies) reeds een belangrijke aanwinst zijn. Uiteindelijk zou

(1) Voor het bruto geografisch produkt per provincie en taalstreek (1955-1963) raadplege men bijv. het *Statistisch Tijdschrift*, Nr. 3, 1966 en *Statistische en Economische Studiën*, Nr. 12, 1966.

men er moeten toe komen tussen de regio's een matrix van de goederen- en dienstenstromen en een matrix van de geld- en kapitaalstromen op te stellen.

2.2. Eerstgenoemde reikt inhoudelijk niet zover als een *interregionale input-outputmatrix* voor de negen provincies. In ons land hebben nochtans sommige onderzoekers zich onderscheiden door het opstellen van een regionale transactiematrix, o.m. voor Luik en Antwerpen, terwijl in het buitenland analoge matrices werden opgebouwd, o.m. voor Sicilië, Berlijn en enkele gewesten in de Verenigde Staten.

Een treffende illustratie van de afstand tussen een wiskundig en een econometrisch model wordt hier geboden. De regionale transactiematrix voor één streek, wanneer statistisch gerealiseerd, blijkt een machtig instrument te verschaffen voor de structurele analyse van die streek. Aan de andere kant zijn interregionale input-outputtabellen aangewezen, wanneer een inzicht vereist is in de structurele verwevenheid van meerdere streken, een inzicht dat onontbeerlijk is voor een geharmoniseerde regionale programmatie. Hun statistische verwezenlijking stelt evenwel nog omvangrijke problemen, die het N.I.S. niet hebben afgeschrikt, zoals blijkt uit het referaat, door Directeur R. Dereymaeker gehouden op de Statistische Dag 1966.

2.3. De regionale transactiematrix kan worden gecombineerd met allerlei bijkomende analyse-instrumenten :

(i) *het groeipoolconcept* : (Aujac, Chenery, Rasmussen e.a.) hieronder verstaat men de (tendentieel driehoekige) structuur van de tabel, die wijst op de sterkere of zwakkere economisch-technologische impulsen, voorwaarts of achterwaarts werkend, van de ene sector naar de andere. Andere criteria zoals produktiviteit en diversificatie kunnen bij het groeipoolconcept betrokken worden.

(ii) In dezelfde orde van gedachten kunnen de *comparatieve kostenstudies* vermeld worden, die door W. Isard werden geïntegreerd met de input-outputanalyse voor de zgn. « greater New York-Philadelphia Industrial Region ». Het bekomen van de comparatieve kostengegevens blijkt ook hier nog ernstige moeilijkheden op te werpen.

(iii) Ten slotte kan men de *lineaire programmering* combineren met een regionaal input-output model, waardoor men de optimale regionale toewijzing der produktie-activiteiten kan berekenen, uitgaande van een te minimaliseren kostenfunctie (met o.m. de transportkosten) of een te maximaliseren objectieffunctie (bijv. het maximaliseren van de produktie onder vorm van toegevoegde waarde).

De kunst — « l'art de faire des modèles » — bestaat er evenwel in te roeien met de riemen die men op een gegeven ogenblik heeft. Er zijn

heel wat pogingen aan te duiden (en ze groeien steeds) om methoden te ontwikkelen die minder « sophisticated », doch meer realistisch zijn. Die evolutie houdt o.m. verband met de toenemende druk van de concrete noden aan regionale ontwikkeling en reconversie, waardoor het (toegepast) econometrisch model prioriteit krijgt op het wiskundige (theoretisch) model.

3. — Econometrische modellen.

3.1. In de eerste plaats kunnen modellen van lineaire programmering worden vermeld, die los van enige input-outputtabel worden opgesteld. Doorgaans wordt in deze modellen een maximale tewerkstelling beoogd, die moet voldoen aan minimale normen van produktie, terwijl de binding met een nationaal expansieprogramma bijkomende grensvoorwaarden oplegt.

3.2. Verder kan het definiëren op een concrete basis worden aangestipt van het veel gebruikte concept « groeitak ». De inhoud van dit voor de regionaal-economische politiek zo belangrijke instrument is echter alles behalve uniform. In een rapport, opgesteld door het Centrum voor Economische Studiën te Leuven, werden een viertal voorwaarden aan een sector opgelegd, opdat hij voor een werkelijke « groeitak » zou kunnen doorgaan (2).

(i) Het ontwikkelingsritme van de bedrijfstak dient 20 pct. hoger te liggen dan het totaal industrieel groeiritme.

(ii) De sector moet in de bedrijfsstructuur een voldoende belangrijke positie innemen. Deze belangrijkheid kan o.m. worden gemeten door het procentueel aandeel dat de bedrijfstak in de totale produktie inneemt.

(iii) Deze belangrijkheid is functie van de dispersiekracht, bijv. gemeten door de dispersiecoëfficiënt van Rasmussen, die de bedrijfstak in het net der interindustriële relaties kenschetst.

(iv) Vanuit het standpunt van de inkomensvorming is een hoge toegevoegde waarde per actief persoon noodzakelijk.

In het vermelde rapport werden voor de periode 1953-1959 in de Vlaamse gewesten vier groeitakken onderscheiden, m.n. metaalverwerking, scheikunde, petroleum en elektriciteit, tegenover — steeds op basis van dezelfde criteria — drie in de Waalse gewesten, m.n. de siderurgie, de elektriciteit en de scheikunde.

3.3. De reconversie van de Borinage en het Centrum eensdeels, de ontwikkeling van de Zuiderkempen en de Westhoek anderdeels, vormen de eerste concrete pogingen tot regionale programmatie in ons land. Aan de

(2) *De economische structuur der Vlaamse gewesten*, C.E.S., 1963, (gestencild) blz. 105-114.

basis liggen weliswaar géén formele regionale modellen, maar alleszins consistente denkschema's, die toelaten een aantal gekwantificeerde objectieven (bijkomende arbeidsplaatsen of een bepaald niveau van actieve bevolking) te relateren tot de invloed van diverse instrumenten, zoals o.m. de regionale investeringen en de onderscheidene infrastructuur.

3.4. Op internationaal vlak constateren wij dezelfde tendentie naar een operationele aanpak, geïnspireerd door de modeltechniek, zonder volledige formele uitwerking van het model. De E.E.G.-diensten hebben deze benadering toegepast bij het opstellen van een industrialisatieprogramma voor het gewest Bari-Tarente in Zuid-Italië. In opdracht van de O.C.E.D. heeft de Rotterdamse hoogleraar J. Klaassen regionale « feasibility »-rapporten opgesteld omtrent de regionale inplanting van enkele moderne bedrijfstakken in de Verenigde Staten, waarbij hij het belangrijk onderscheid invoert of die activiteiten *niet* regionaal gebonden (« footloose ») of *wél* gewestelijk gedetermineerd zijn.

*
* *
*

4.1. In verband met de beschikbaarheid en de kwaliteit der statistische gegevens, waardoor de econometrische modellen dienen gevoed te worden, kunnen wij slechts besluiten met de woorden, die een der grote voorlopers van de modeltechniek neerschreef : « ... toute la certitude est dans l'évidence des données » (Quesnay).

Op stuk van regionaal-economische modellen dient er zowel in als buiten ons land nog heel wat gepresteerd te worden inzake omvang en kwaliteit der gegevens, wil men de keur van wiskundige modellen, waarover men reeds beschikt, kunnen omzetten in operationele instrumenten ten dienste van de « policy-makers ».

Van regeringswege zou men de wetenschappelijke vooruitgang en meteen de welvaartspromotie dienen, indien men in het raam van de regionale reconversie- en ontwikkelingspolitiek enkele operationele modellen deed opstellen, in eerste instantie gericht op het detecteren van de meest geschikte industriële orientaties.

De grote inspanningen door het N.I.S. de jongste jaren verricht op stuk van de regionale data, zouden daardoor verder gecatalyseerd worden en tot hun volledige benutting gebracht.

DISCRIMINANTANALYSE BIJ HET OPSPOREN VAN PATIENTEN MET TURNER-SYNDROOM

door E. MEULEPAS

Katholieke Universiteit te Leuven

Het Turner-syndroom is een samenloop van verscheidene afwijkingen bij sommige vrouwen, o.m. sexuele onvolwassenheid en dwerggroei; ook een korte nek (« webbed neck ») is typisch. Het werd door Turner in 1938 beschreven. De verklaring van het ontstaan van dit syndroom heeft nog wel een twintigtal jaren op zich laten wachten. Het gaat waarschijnlijk om een genetische stoornis.

Zoals men weet bevatten de celkernen van de mens 23 paren chromosomen, waarvan 22 paren zgn. autosomale chromosomen, gelijk in eenzelfde paar, en één paar geslachtschromosomen, van het type XX bij de vrouw en XY bij de man. Het Y-chromosoom is determinerend voor het mannelijk geslacht, één X-chromosoom minstens is nodig voor leefbaarheid.

Bij vrouwen kan het nu in zeldzame gevallen voorkomen dat één van beide X-chromosomen ontbreekt. Dit XO-type blijkt nu haast steevast het Turner-syndroom te vertonen. Men heeft nu anderzijds bij bepaalde genetische aandoeningen, bvb. bij mongolisme, vastgesteld dat afwijkingen in het huidlijstenpatroon (« vingerafdrukken ») voorkomen. Ook voor XO-afwijkingen werd dit nu in onderhavige studie nagegaan en vastgesteld.

Men gebruikte hiervoor gemakkelijk te kwantificeren kenmerken van het huidlijstenpatroon. Men kwam aldus tot zeven veranderlijken waarvan er vijf aantallen huidlijsten tussen bepaalde punten van de handpalm voorstellen, en de overige een deelverhouding en een hoek. Deze veranderlijken werden gemeten bij 300 vrouwen en 40 Turner-patiënten.

Op grond van deze waarnemingen kan men nu twee vragen beantwoorden, welke nauw samenhangen :

1° is er verschil in huidlijstenpatroon tussen normale vrouwen en Turner patiënten ?

2° indien er verschil is, hoe kan men van een vrouw uit haar huidlijstenpatroon haar genetisch type bepalen ?

We zullen in de nu volgende uiteenzetting deze problemen gradueel oplossen, en met de eenvoudige gevallen beginnen.

Eerste geval : één kwalitatieve dichotomische veranderlijke welke de groepsindeling verwekt.

Onderstel dat de classificatie betekent : toewijzing aan het XX- of het XO-type, waarbij de veranderlijke t resp. de waarde 0 of 1 aanneemt.

Is de veranderlijke t waargenomen, dan is de classificatieregels triviaal. Is ze niet waargenomen, maar kent men de populatieverdeling, waarin $P(XO) = P(Z = 1) = 1/10.000$ bvb., dan is de regel : aanzie de (niet onderzochte) vrouw als normaal. De kans op een foute conclusie is dan $1/10.000$.

Tweede geval : één kwalitatieve dichotomische veranderlijke, gecorreleerd met de indelings variabele.

Weze $Z = 1$ bij vrouwen met Turner-syndroom,
 $Z = 0$ bij vrouwen zonder Turner-syndroom.

Op basis van Z moet een vrouw ingedeeld worden bij het XX- of het XO-chromosomentype.

Veronderstel dat bij $m = 300$ XX-vrouwen en bij $n = 40$ XO-vrouwen Z werd waargenomen, waarbij we Z door X in de eerste en door Y in de tweede groep zullen voorstellen. Dan stellen de gemiddelden $\bar{X}$, $\bar{Y}$ de waargenomen proporties p_1 , p_2 in beide groepen afzonderlijk voor, en $\bar{Z} = p$ de proportie over beide groepen samen.

De eerste, preliminaire, vraag leidt nu tot de t -toets voor het verschil tussen twee proporties. Is er geen verschil, dan komt het Turner-syndroom in gelijke mate bij XX- en XO-vrouwen voor en kan het niet dienen als basis voor classificatie.

We zullen dit geval toch volledig uitwerken omdat we de essentie van discriminantanalyse reeds tot uiting kunnen brengen.

Voor ongepaarde steekproeven berekent men :

$$t^2 = \frac{p_1 - p_2}{\frac{m p_1 q_1 + n p_2 q_2}{m + n - 2} \left(\frac{1}{m} + \frac{1}{n} \right)} \quad (1)$$

We stellen dit nu in andere notaties voor :

men weet dat

$$s_w^2 = \frac{m p_1 q_1 + n p_2 q_2}{m + n - 2} \quad (2)$$

de binnenvariantie van Z voorstelt, d.i. het gewogen gemiddelde van de varianties $\frac{m p_1 q_1}{m - 1}$ en $\frac{n p_2 q_2}{n - 1}$ van resp. X en Y in beide groepen afzon-

derlijk (het moge wel vermeld worden, tegen veelverbreide formules in, dat de geschatte varianties niet $p_1 q_1$, $p_2 q_2$ zijn, en de geschatte variantie van een geschatte proportie niet $\frac{p q}{n}$ is, maar $\frac{p q}{n-1}$, d.i. $\frac{s^2}{n}$ mits $s^2 = \frac{n p q}{n-1}$); verder kan

$$s_B^2 = m(p_1 - p)^2 + n(p_2 - p)^2, \quad (3)$$

de tussenvariantie, omgewerkt worden tot

$$s_B^2 = \frac{(p_1 - p_2)^2}{1/m + 1/n} \quad (4)$$

uit (1) tot (4) volgt dan

$$t^2 = \frac{s_B^2}{s_W^2} = F(1, m + n - 1) \quad (5)$$

Stelt men nu $d = p_1 - p_2 = \bar{X} - \bar{Y}$

dan noemt men $D^2 = \frac{d^2}{s_W^2} \quad (6)$

het afstandskwadraat van Mahalanobis tussen beide groepen, en

$$T^2 = \frac{D^2}{1/m + 1/n} = t^2 \quad (7)$$

de T^2 van Hotelling; in het speciale geval van één veranderlijke kan t zelf als een gemiddelde afstand beschouwd worden.

Men bepaalt nu een indelingsveranderlijke L , de zgn. (lineaire) discriminantfunctie door

$$L = aZ \quad (8)$$

met a bepaald uit $s_W^2 \cdot a = d \quad (9)$

Haar gemiddelden zijn

$$\bar{L}_X = a\bar{X} = ap_1 \quad \text{in de eerste groep}$$

$$\bar{L}_Y = a\bar{Y} = ap_2 \quad \text{in de tweede groep}$$

$$\text{en } \bar{L} = a\bar{Z} = ap \quad \text{voor beide groepen samen.}$$

Men bemerkt dat

$$\bar{L}_X - \bar{L}_Y = a(p_1 - p_2) = ad = \frac{d^2}{s_W^2} = D^2 \quad \text{wegens (9) en (6).}$$

Is nu voor een gegeven individu de L-waarde zó dat bvb. $\bar{L}_X < L < \bar{L}$ $< \bar{L}_Y$ dan behoort het tot de eerste groep, terwijl in het geval $L > \bar{L}$ met $\bar{L}_X < \bar{L} < \bar{L}_Y$ het tot de tweede behoort.

De gevallen $L < \bar{L}_X < \bar{L} < \bar{L}_Y$ en $\bar{L}_X < \bar{L} < \bar{L}_Y < L$ zijn evident.

Cijfervoorbeeld : $\bar{X} = p_1 = 0$, $\bar{Y} = p_2 = 0,9$

dan is

$$d = 0,9$$

$$p = \frac{300 p_1 + 40 p_2}{340} = \frac{36}{340}$$

$$s_w^2 = \frac{40 \times 0,9 \times 0,1}{338}$$

$$a = \frac{338}{4} = 84,5$$

zodat

$$\bar{L}_X = 0, \quad \bar{L}_Y = 76,5, \quad \bar{L} = 8,95;$$

een vrouw met Turner-syndroom ($Z = 1$) krijgt een L-waarde $L = 84,5 \times 1 = 84,5$ waardoor ze in de tweede groep valt.

Derde geval : één continue veranderlijke.

Het vorige geval gaat onverminderd door, mits de binnenvariantie s_w^2 op de geëigende wijze te berekenen; men kan weer een L-veranderlijke bepalen, wat echter overbodig is (zoals overigens in het vorige geval) omdat ze slechts op een schaalfactor na van Z verschilt.

Vierde geval : twee ongecorrelleerde continue veranderlijken.

Het is pas nu dat discriminantanalyse nuttig wordt, omdat ze de indeling op basis van meerdere veranderlijken vervangt door deze op basis van één L-veranderlijke.

Men bekomt deze L uit twee veranderlijken Z_1, Z_2 met gemiddelden $\bar{Z}_1, \bar{Z}_2$ over beide groepen samen, en $\bar{X}_1, \bar{X}_2$ in de eerste, $\bar{Y}_1, \bar{Y}_2$ in de tweede groep, en binnenvarianties $(11)_w, (22)_w$ door te stellen

$$L = L_1 + L_2$$

waarin L_1 en L_2 beide volgens (8) en (9) uit Z_1 resp. Z_2 bepaald worden, zodat

$$L = a_1 Z_1 + a_2 Z_2 \quad (10)$$

Hieruit volgen de gemiddelden

$$\bar{L}_X = a_1 \bar{X}_1 + a_2 \bar{X}_2$$

$$\bar{L}_Y = a_1 \bar{Y}_1 + a_2 \bar{Y}_2$$

$$\bar{L} = a_1 \bar{Z}_1 + a_2 \bar{Z}_2$$

waarmede classificatie op basis van $L \geq \bar{L}$ kan gebeuren zoals in het tweede geval.

De proportie P van misclassificaties wordt gegeven door

$$P = 1 - F(D/2)$$

waarin $F(x) = P(X \leq x)$ de cumulatieve Gaussverdeling voorstelt, en D volgt uit

$$D^2 = \bar{L}_X - \bar{L}_Y.$$

De preliminaire toets voor verschil steunt nu op

$$T^2 = t_1^2 + t_2^2 \quad (11)$$

De critische waarde voor T^2 volgt uit de betrekking

$$\frac{T^2}{m+n-2} \cdot \frac{m+n-p-1}{p} = F(p, m+n-p-1) \quad (12)$$

waarin $p = 2$ het aantal veranderlijken voorstelt; vermits nu

$$F(2, m+n-3) < F(1, m+n-2) < 2 F(2, m+n-3)$$

$$\text{is} \quad \frac{T^2}{2} \cdot \frac{m+n-3}{m+n-2} < t^2 < T^2 \cdot \frac{m+n-3}{m+n-2}$$

$$\text{zodat} \quad t_a < T_a < t_a \sqrt{2} \quad (13)$$

In een (t_1, t_2) -assenstelsel ligt het critisch gebied voor de marginale toetsen op basis van t_1 en t_2 samen buiten een vierkant met middelpunt in de oorsprong en halve zijde t_a , maar voor de gezamenlijke toets op basis van T^2 buiten de cirkelomtrek $t_1^2 + t_2^2 = T_a^2$ welke wegens (13) de zijden van het vierkant snijdt. Een zeer belangrijk gevolg hiervan is dat T^2 significant kan zijn terwijl t_1 en t_2 het beide niet zijn en omgekeerd.

Vijfde geval: twee continue gecorrelleerde veranderlijken.

L wordt nog steeds door (10) bepaald, maar a_1 en a_2 zijn nu de oplossingen van het stelsel

$$(11) a_1 + (12) a_2 = d_1$$

$$(21) a_1 + (22) a_2 = d_2$$

waarin (12) de binnencovariantie tussen Z_1 en Z_2 voorstelt.

Nu gaat (11) over in

$$T^2 = \frac{1}{1-r^2} (t_1^2 - 2r t_1 t_2 + t_2^2) \quad (14)$$

Het kritisch gebied voor de T-toets ligt nu buiten een ellips met vergelijking $T^2 = T_a^2$ waarin T^2 door (14) gegeven is; deze ellips heeft assenlengten $2 T_a \sqrt{1 \pm r}$ in de richtingen $\pm 45^\circ$ ingeschreven in een vierkant met zijden $2 T_a$.

Algemeen geval : p continue veranderlijken.

Notaties : $\underline{Z}$ kolomvector der veranderlijken

$\underline{Z}$ » » gemiddelden

$\underline{d}$ » » gemiddelde verschillen

$\underline{a}$ » » a-coëfficiënten

S binnen-variantie-covariantiematrix.

Dan is

$$\underline{a} = S^{-1} \underline{d}$$

$$\underline{L} = \underline{a}' \underline{Z}$$

$$D^2 = \underline{a}' \underline{d} = \underline{d}' S^{-1} \underline{d}$$

$$T^2 = \frac{D^2}{1/m + 1/n}$$

$$F(p, m+n-p-1) = \frac{m+n-p-1}{p} \cdot \frac{T^2}{m+n-2}$$

$$\frac{\frac{T^2}{m+n-2} / p}{1/(m+n-p-1)}$$

De laatste vorm voor F is analoog aan de F-toets voor het door p veranderlijken gezamenlijk verklaarde gedeelte van de kwadraatensom van een afhankelijke veranderlijke in meervoudige regressie, waarin de residuele

variantie aan 1 gelijk zou zijn (wat door schaalverandering trouwens steeds te bereiken is). Ook andere toetsen in meervoudige regressie kunnen door een soort dualiteitsprincipe omgewerkt worden tot toetsen voor discriminant-analyse.

Deze toetsen kan men aanwenden om bepaalde belangrijke vragen te beantwoorden. Het algemeen geval roept nl. nog verscheidene problemen op.

Een eerste volgt uit het algemeen wetenschappelijke principe dat overtollige of nutteloze gegevens buiten beschouwing moeten gelaten worden. Daarom werden de berekeningen stapsgewijze uitgevoerd, waarbij in een voorafbepaalde volgorde (nl. van verwacht meest- tot minst-discriminerende) de veranderlijken in de analyse werden betrokken. Bij iedere stap werd een discriminantfunctie L berekend, en werd getoetst of weglaten van telkens één der reeds opgenomen veranderlijken verantwoord kon worden; ook toetsen voor het weglaten van telkens méér dan een veranderlijke zijn mogelijk, maar werden niet uitgevoerd.

Een tweede probleem van vereenvoudiging is, of de berekende a -coëfficiënten door afgeronde getallen mogen vervangen worden zonder significant verlies aan informatie. Men toetst dan een opgelegde L -functie.

Tenslotte zijn er nog twee problemen van meer technische aard: de variantie-covariantiematrices van beide groepen mogen niet verschillen, wat reeds op eerste zicht waar bleek te zijn, en de veranderlijken moeten normaal verdeeld zijn. Daarom werden normaliteitstoetsen uitgevoerd met behulp van lineaire functies van resp. de derde en vierde momenten, functies die zelf standaardnormaal verdeeld waren. Deze toetsen brachten geen significante afwijkingen van normaliteit aan het licht, wat op grond van het groot aantal waarnemingen wel te verwachten was.

BIBLIOGRAPHIE

- H. Van den Berghe, *Het zogenaamde syndroom van Turner*, Standaard-Wetenschappelijke Uitgeverij, 1966.
- E. Meulepas, *Discriminant analysis and consumer price index*, in Tijdschrift voor Economie (te verschijnen).
- T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, 1958.
- C.R. Rao, *Advanced Statistical Methods in Biometric Research*, Wiley, New York, 1952.
- M.G. Kendall, *The Advanced Theory of Statistics*, vol. II, Griffin, London, 1948.
- C.C.Li, *Introduction to Experimental Statistics*, Mc Graw-Hill, New York, 1964.
- A. Hald, *Statistical Theory with Engineering Applications*, Wiley, New York, 1952.

HET VERWERKEN VAN KANKERSTERFTESTATISTIEKEN

door J. V. JOOSSENS,

Katholieke Universiteit te Leuven.

De gegevens over kankersterfte die tegenwoordig, vooral dank zij de Wereldgezondheids Organisatie, ter beschikking van de onderzoekers gesteld worden, zijn zo talrijk (gegevens per land, jaartal, geslacht en leeftijd) dat ze totaal onoverzichtelijk worden. Het is dan ook nodig speciale technieken te gebruiken om de informatie uit deze gegevens te halen, in een vorm die biologische waarde heeft.

Kankersterfte gedraagt zich, tussen bepaalde leeftijden die voor de clinicus van overwegend belang zijn (namelijk 40 tot 64 jaar), als een lineaire functie van de leeftijd (op dubbel logaritmische schaal). Hiervan kan gebruik gemaakt worden om een belangrijke reductie van het aantal gegevens te bereiken. Als R_x de sterfte is op leeftijd x dan volgt uit het voorgaande dat

$$R_x = R_1 x^b \quad (1)$$

waar R_1 de berekende sterfte is op leeftijd één en b de helling is van de regressielijn.

Een belangrijk biologisch fenomeen is nu dat $\log R_1$ en b zeer sterk gecorreleerd zijn en dit voor gegevens uit verschillende landen en beide geslachten (Joossens 1963). Nu zal wel met achterdocht opgekeken worden naar een techniek die $\log R_1$ en b correleert, aangezien voor regressielijnen berekend op verschillende « random samples » van éénzelfde land, jaartal en geslacht er een evident louter mathematisch verband bestaat tussen $\log R_1$ en b . Al deze lijnen zullen immers kruisen op $\log R_x$ en $\log x$. Deze critiek geldt echter niet noodzakelijk voor de correlatie tussen $\log R_1$ en b afkomstig van verschillende landen, jaartallen en geslachten. Indien er hier een correlatie bestaat dan kan deze van biologische oorsprong zijn, namelijk door het feit dat de sterfteleeftijdslijnen gemiddeld kruisen op een bepaalde leeftijd. Deze leeftijd wordt dan de isomortaliteitsleeftijd genoemd (Joossens 1963). De regressie tussen $\log R_1$ en b zal twee constanten opleveren namelijk $\log \alpha_1$ en β_1 . Het is

Met de hulp van het Fonds voor Wetenschappelijk Geneeskundig Onderzoek, Brussel.

gemakkelijk in te zien, dat als de correlatie uitgerekend wordt tussen $\log R_x$ en b (met constanten $\log \alpha_x$ en β_x) of in het algemeen tussen $\log R_x$ en b (met constanten $\log \alpha_x$ en β_x), $\alpha_1 = \alpha_x = \alpha_x$, terwijl $\beta_1 = -\log X_m$, waar X_m de gemiddelde isomortaliteitsleeftijd is en

$$\beta_1 = \beta_x - \log x \text{ of } \log X_m = -\beta_x + \log x \quad (2)$$

Meulepas en Joossens* hebben recent aangetoond dat dit immers voortvloeit uit de duale eigenschappen van lijn en punt. Het probleem van een ge-

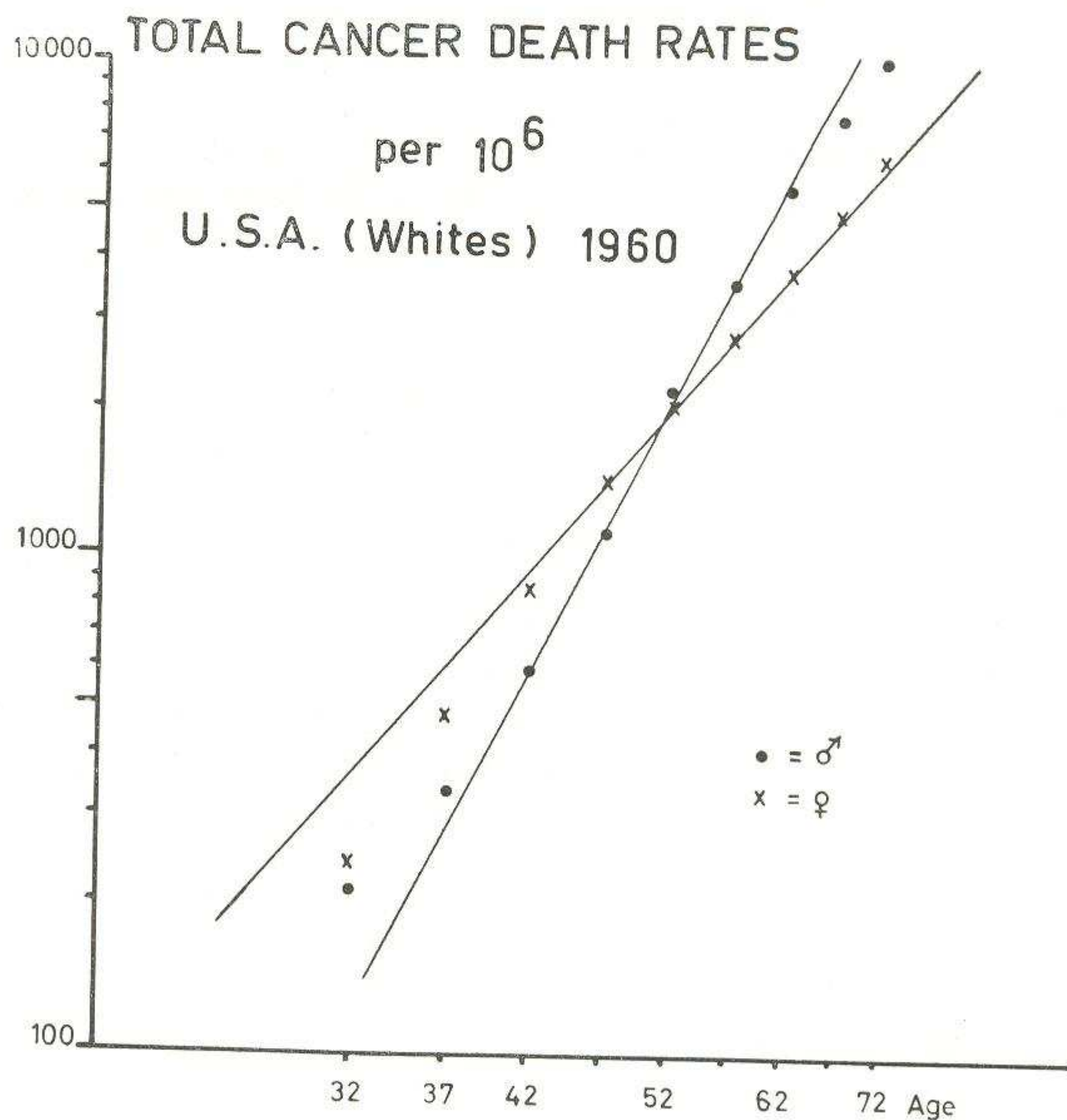


Fig. 1.

middeld snijpunt te zoeken van een bundel regressielijnen met de verzameling van de lijn parameters van deze lijnen (in casu de $\log R_1$'s en de b 's) is dan equivalent met het zoeken van één regressielijn tussen de verzameling van de punten bepaald door de $\log R_1$'s en de b 's, m.a.w. de correlatie tussen $\log R_1$ en b .

* In publicatie.

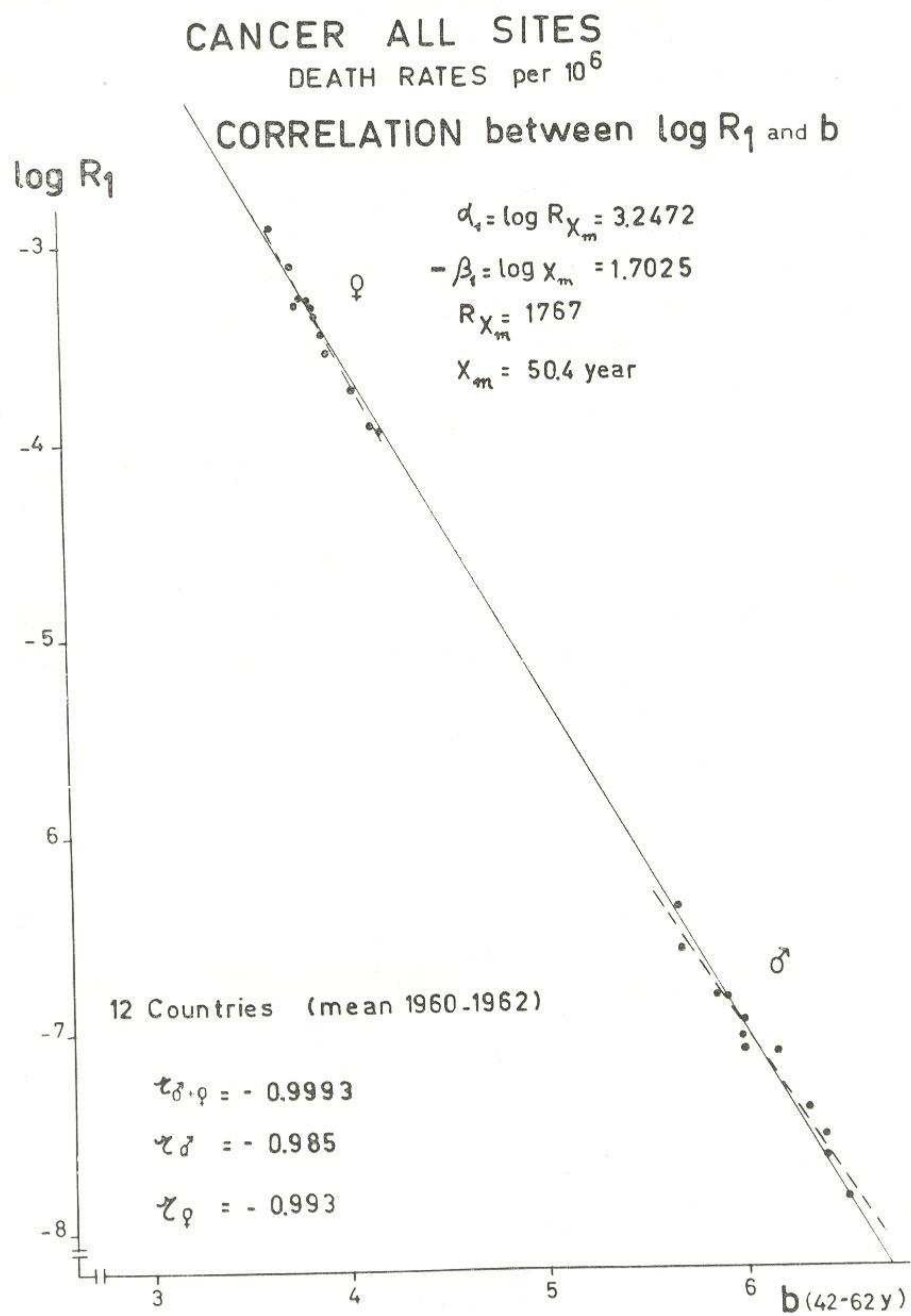


Fig. 2.

De berekening bevestigt dat de correlatie tussen $\log R_1$ en b en tussen $\log R_x$ en b (welke laatste mathematisch niet gecorreleerd zijn) identische waarden oplevert voor X_m , rekening houdend met de vergelijking (2).

Een tweede methode om aan te tonen dat het bestaan van X_m een biologisch fenomeen is en geen technisch artefact kan gezocht worden in de berekening van het snijpunt * van slechts twee sterfte-leeftijd lijnen van éénzelfde land en jaartal, maar van beide geslachten. Fig. 1 geeft de kankersterfte weer voor beide geslachten in de Verenigde Staten in 1960. De lijnen snijden elkaar in de omgeving van 50 jaar. Welnu de berekening bewijst dat voor n landen en k

$$\text{jaartallen } \log X_m = \frac{\sum_{i=1}^n \sum_{j=1}^k \log X_{ij}}{n \cdot k}.$$

Fig. 2 geeft de correlatie weer tussen $\log R_1$ en b voor mannen en vrouwen. Eerst voor beide geslachten afzonderlijk dan, aangezien $b \delta$ niet significant verschillend is van $b \varphi$, ook voor beide samen. De berekende isomortaliteitsleeftijd, X_m , is hier 50,4 jaar. Ook voor andere aandoeningen kan een isomortaliteitsleeftijd berekend worden; zo is deze voor infectieziekten rond de 30 jaar, voor diabetes rond de 50 jaar en voor atherosclerosis (degeneratieve hartaandoeningen) rond de 85 jaar.

Door X_1 of X_m in vergelijking (1) te brengen bekomt men :

$$R_{X_1} = R_1 X_1^b \quad (3) a$$

$$\text{of } R_{X_m} = R_{X_m} X_m^b \quad (3) b$$

Uit (1) en (3) volgt

$$R_x = R_{X_1} \left(\frac{x}{X_1} \right)^b \quad (4) a$$

$$\text{of } R_x = R_{X_m} \left(\frac{x}{X_m} \right)^b \quad (4) b$$

Vergelijking (4) is biologisch veel interessanter dan (1), aangezien R_1 een fictieve waarde is, terwijl R_{X_1} of R_{X_m} meestal rechtstreeks observeerbaar is.

Welke van de twee vergelijkingen (4) zal verkozen worden hangt af van het feit of X_1 in alle landen en jaartallen een constante is. In dat geval is (4) b te verkiezen. Door verder onderzoek is het echter gebleken dat X_1 niet constant is, noch in de tijd, noch onder verschillende landen. Daarom zal (4) a de beste keuze zijn.

* De leeftijd overeenkomend met dit snijpunt wordt dan X_1 genoemd.

(4) a is ook van nut voor de studie van de sex ratio van de sterfte (verhouding mannelijke op vrouwelijke sterfte). Vermits R_{X_1} , per definitie, gelijk is voor beide geslachten, herleidt de sex ratio zich tot

$$\left(\frac{x}{\bar{X}_1} \right) b_{\delta} - b_{\varphi} \quad (5)$$

ten minste voor het gebied van x waar aan (1) voldaan wordt.

Door Joossens en Van den Berghe en door Joossens en Meulepas werd recent een hypothese voorgesteld waarbij b , de exponent van de leeftijd, als genetisch van oorsprong beschouwd wordt. b , equivalent met de elasticiteit van de economisten, zou volgens deze hypothese gedragen worden door de X chromosomen en namelijk dominant zijn voor degeneratieve cardiovasculaire aandoeningen en diabetes, maar recessief voor kanker en infectieziekten. Nu is het gemakkelijk te bewijzen dat b_{δ} in dit geval de waarde van b_{φ} zal zal bepalen en namelijk zo dat $b_{\delta} - b_{\varphi}$ des te groter zal zijn in absolute waarde naarmate de variance van b_{δ} groter is. Voor een rechthoekige verdeling en een dominant verschijnsel zoals in atherosclerose (tussen 50 en 84 jaar) is

$$b_{\delta} - b_{\varphi} = -\frac{\sigma_{b_{\delta}}}{\sqrt{3}} \quad (6)$$

voor een recessief, zoals kanker tussen 40 en 64 jaar,

$$b_{\delta} - b_{\varphi} = +\frac{\sigma_{b_{\delta}}}{\sqrt{3}}$$

Voor een normale verdeling is $b_{\delta} - b_{\varphi}$ iets kleiner voor éénzelfde variance, dan voor een rechthoekige verdeling.

Deze formule werd afgeleid uit een meer algemeen werk van Moriguti (1951).

Hieruit volgt, als de variance voor kanker gelijk is aan die van atherosclerose, dat $(b_{\delta} - b_{\varphi})_{ca} = -(b_{\delta} - b_{\varphi})_{at}$.

Onder dezelfde voorwaarde is dan

$$b_{at\delta} + b_{ca\delta} = b_{at\varphi} + b_{ca\varphi}$$

Figuur 3 toont dat aan de eis van gelijke variance voor $b_{ca\delta}$ en $b_{at\delta}$ ongeveer voldaan wordt als het gemiddelde van 12 landen genomen wordt en dit voor alle onderzochte jaren tussen 1956 en 1962.

b is een vrij grote waarde, gemiddeld ongeveer 6 voor kanker bij mannen (4 bij vrouwen) en ongeveer 7 voor atherosclerose bij mannen (9 bij vrouwen). Daaruit volgt dat b_{at} vooral de maximaal bereikbare levensduur duidelijk beperkt. b kan dan ook beschouwd worden, in de hoger gemaakte hypothese een eigenschap van de X chromosoom, als de genetisch limiterende factor van de levensduur en dit karakter zou dan voor elke diersoort verschillend zijn. Bomen, die geen X chromosomen bezitten zoals dieren, hebben theoretisch dan ook geen beperkte levensduur.

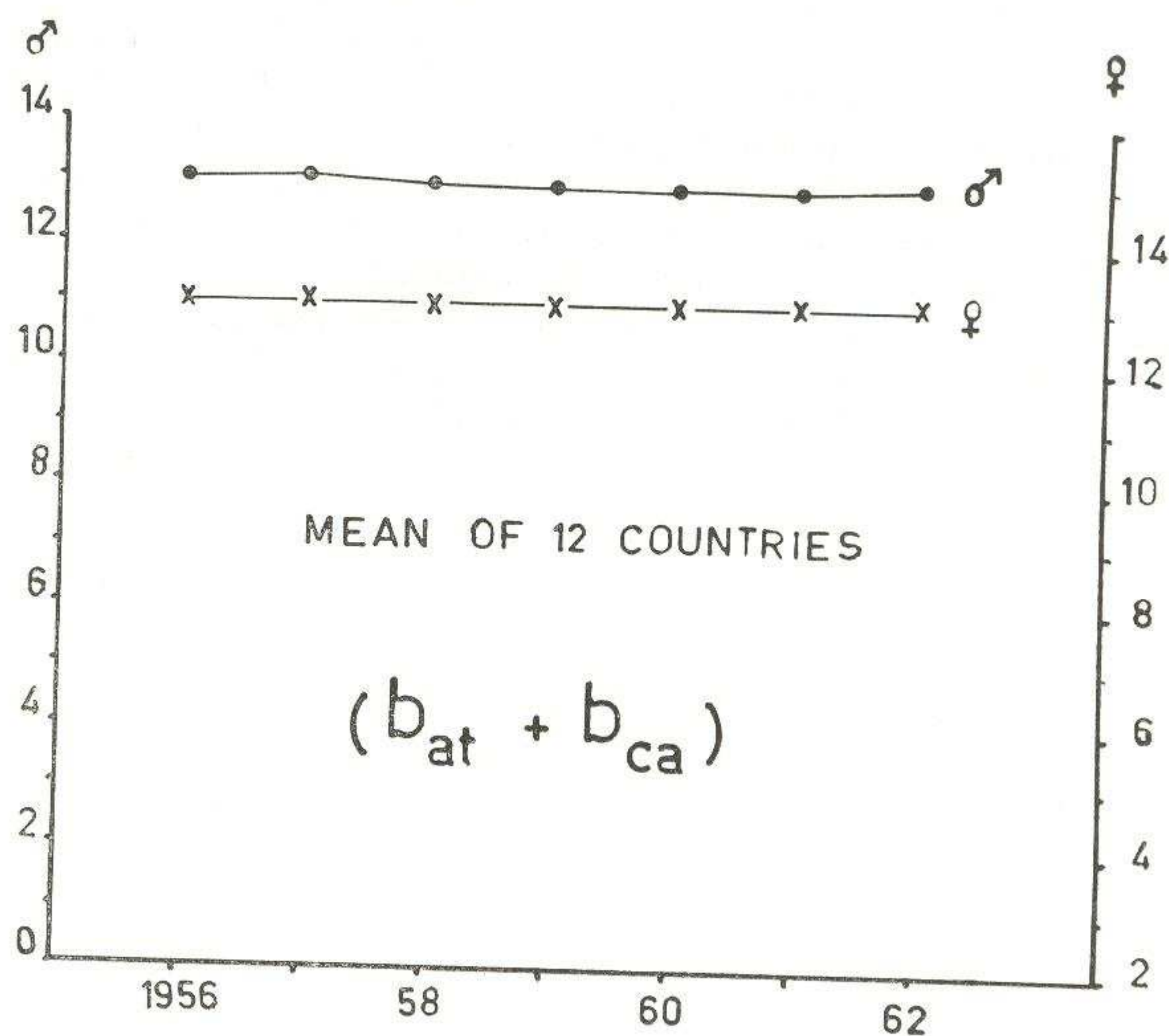


Fig. 3.

Een gans andere methodiek om kankersterftestatistieken, meer bepaald sommige kankerlocalisaties, te verwerken is de volgende (Joossens 1965).

Bepaalde localisaties kunnen per leeftijd, tussen 40 en 74 jaar, uitgedrukt worden in % van de totale kankersterfte op die leeftijd. Voor elk van de 21 onderzochte landen, blijkt nu dat de som van de % van maag-, long- en geslachtskankers ongeveer gelijk is op elke leeftijd van 40 tot 74 jaar, op voorwaarde dat de gegevens van mannen en vrouwen samengeteld worden. Fig. 4

geeft hiervan enkele voorbeelden. Voor mannen stijgt het % met de leeftijd, voor vrouwen daalt het juist evenveel. Dit wijst erop dat kankersterfte biologische informatie uitwisselt tussen beide geslachten. Dit is dan een ander argument voor de hogergemaakte hypothese.

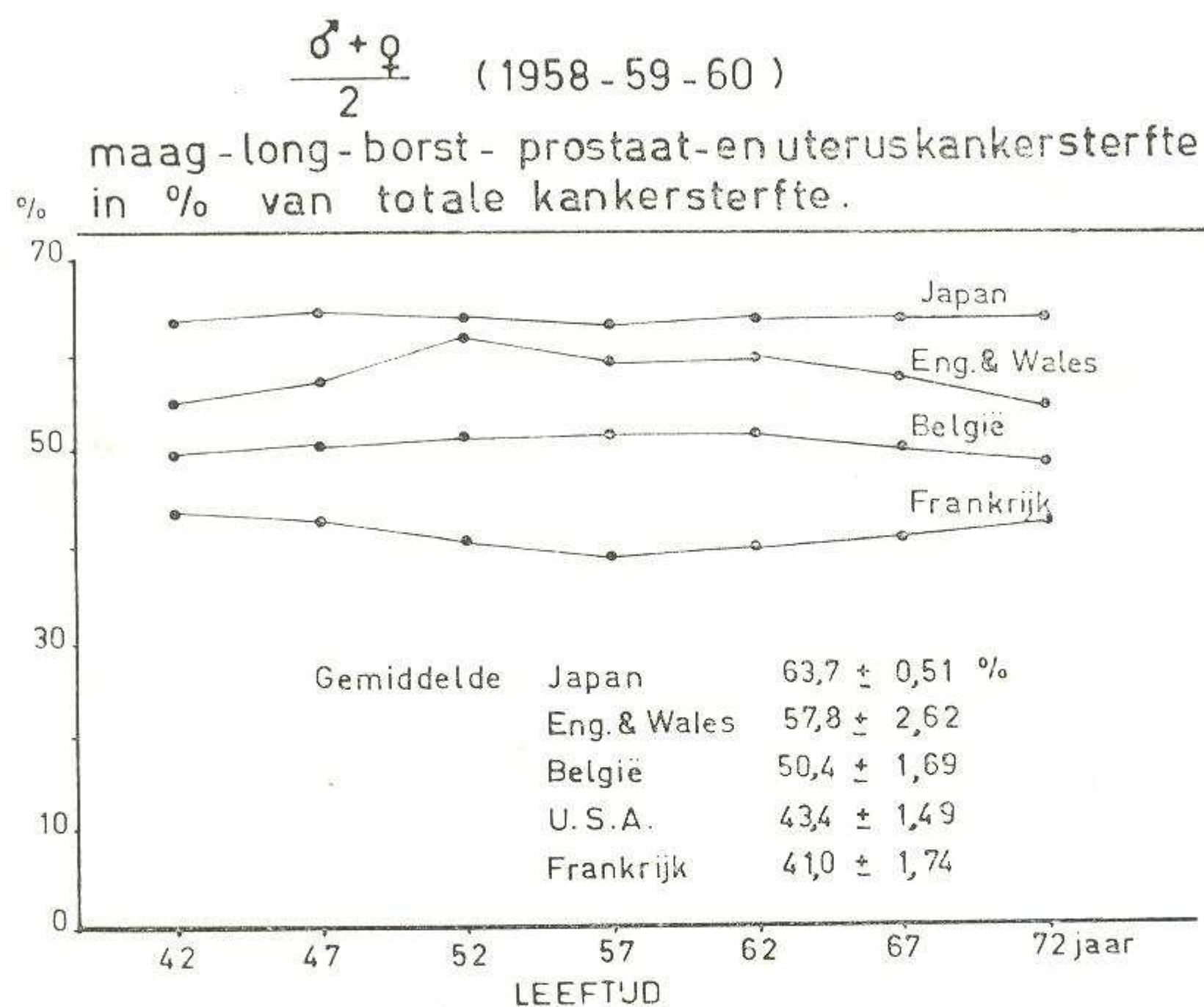


Fig. 4.

Het belang van maag-, long- en geslachtskankersterfte ($\pm 50\%$ van de totale kankersterfte) wordt door de volgende regressie-analyse aangetoond. Gegevens van 20 of 21 landen van de jaren '58, '59 en '60 en waargenomen op leeftijden gaande van 42 tot 72 jaar ($n = 427$) werden in een veelvoudige lineaire regressie verwerkt. De regressie werd eerst per leeftijd uitgevoerd ($n = 61$) om de vergelijkbaarheid van de bekomen regressiecoëfficiënten na te gaan van leeftijd tot leeftijd. Dit bleek het geval te zijn voor de coëfficiënten van maag-, long- en geslachtskankersterfte, maar niet voor de constante factor. Deze werd namelijk groter met de leeftijd volgens een machtsfunctie van de leeftijd. Daarom werd in de regressieanalyse een machtsfunctie van de leeftijd ingevoerd. De waarde van deze exponent werd door de elektronische ordinator door « error and trial » gezocht totdat de R^2 maximaal was en zo mogelijk tevens de nieuwe constante factor niet significant verschillend was van nul.

De waarde van de regressiecoëfficiënten wordt in tabel I aangegeven.

TABEL I

Beste regressie-vergelijkingen

 $n = 427$

$$\delta \quad R_x = 0,811 \left(\frac{x}{48,9} \right)^{4,90} + 0,796 M + 1,155 L + 1,380 G (*) \quad R^2 = 0,991$$

$$\varphi \quad R_x = 0,819 \left(\frac{x}{48,9} \right)^{4,90} + 0,873 M + 1,607 L + 1,696 G - 0,258 \quad R^2 = 0,986$$

$$\frac{\delta + \varphi}{2} \quad R_x = 0,414 \left(\frac{x}{48,9} \right)^{5,40} + 0,872 M + 1,260 L + 1,993 G \quad R^2 = 0,993$$

In tabel II wordt één regressievergelijking gebruikt om de kankersterfte per leeftijd te berekenen in drie verschillende rassen, die niet in de oorspronkelijke regressieberekening bevat waren. De overeenkomst met de waargenomen waarden is bevredigend.

TABEL II

Totale kankersterfte ‰ berekend per leeftijd uit de maag-, long- en genitaalkankersterfte (met één vergelijking, de laatste van tabel I)

$$\frac{\delta + \varphi}{2} \quad 1959$$

Leeftijd	Verenigde Staten (niet-blanke bevolking)		Israël		Nieuw Zeeland	
	Waarge- nomen	Berekend	Waarge- nomen	Berekend	Waarge- nomen	Berekend
42	1,17	1,15	0,71	0,73	0,64	0,73
47	1,78	1,75	1,13	1,05	1,42	1,22
52	2,86	2,64	1,99	1,63	1,79	1,87
57	4,47	4,32	2,90	2,72	2,71	2,83
62	5,79	5,67	3,82	3,96	4,62	4,68
67	8,60	8,78	6,57	5,89	6,03	6,63
72	9,44	10,91	9,19	8,27	8,52	8,98

(*) R_x = leeftijd-specifieke sterfte ‰ voor totale kanker op leeftijd x , M = maagkankersterfte ‰, L = longkankersterfte ‰, G = borst- en prostaatkankersterfte bij de man, borst- en uterus kankersterfte bij de vrouw ‰, alle waargenomen op leeftijd x .
 R^2 is het kwadraat van de totale veelvoudige correlatie-coëfficiënt tussen de verschillende veranderlijken.

LITERATUUR

JOOSSENS, J.V. : Mortaliteit en leeftijd. Verh. Kon. Vlaamse Akad. voor Gen. van Belg. 1963 : 25, 175.

JOOSSENS J.V. : Het probleem van de kankersterfte. Verh. Kon. Vlaamse Akad. voor Gen. van Belg. 1965 : 27, 489.

JOOSSENS, J.V. en H. VAN DEN BERGHE : Exogenous and endogenous factors in atherosclerosis. Acta Cardiologica (in druk).

JOOSSENS, J.V. en E. MEULEPAS : The sex ratio of mortality. A genetic hypothesis. Kon. Vlaamse Akad. voor Gen. van Belg. (in druk).

MORIGUTI, S. : Extremal properties of extreme value distributions. Ann. Math. Stat. 1951 : 22, 523.

De hier aangehaalde berekeningen werden door het « Rekencentrum » van de Leuvense Universiteit uitgevoerd (Prof. H. Florin en J. Mainguet). Dr G. Peeters en Dr A. Vuylsteke voerden de programmaties uit.

APPLICATIONS ASTRONOMIQUES ET ASTROPHYSIQUES DE LA STATISTIQUE

par R. COUTREZ
Université libre de Bruxelles

L'exposé fait dans le cadre de cette Journée de la Statistique a pour but de montrer dans quel sens s'est développé depuis quelques décades un ensemble de méthodes appliquées couramment à l'investigation du monde extérieur et groupées sous le titre très vaste d'Astronomie Statistique. Pour l'astronome, en effet, la statistique est devenue un moyen de choix qui lui permet d'accéder à trois fins essentielles, communes d'ailleurs à toutes les sciences de la nature :

a) *caractériser* par un petit nombre de paramètres convenablement définis l'« état » ou structure générale de vastes ensembles d'informations numériques obtenues par l'observation de nombreux objets similaires, *comparer* de telles assemblées de manière à constituer des collections suffisamment homogènes et représentatives d'un phénomène astrophysique déterminé (ex. : l'étude des étoiles de la Galaxie dans le but de reconnaître l'évolution stellaire ou celle du système);

b) *dégager* les relations plus ou moins étroites entre ces paramètres afin que la donnée arbitraire d'un ou plusieurs d'entre eux puisse servir à délimiter les autres (ex. : l'étude des lois de structure galactiques, la reconnaissance des lois du mouvement stellaire);

c) *rechercher* une explication de ces lois empiriques les groupant dans une synthèse théorique nécessairement imprégnée de concepts statistiques (ex. : la dynamique stellaire, véritable mécanique statistique du « gaz » d'étoiles).

On se limitera volontairement au domaine de la Statistique et de la Dynamique stellaires, celles-ci ne constituant qu'une partie de l'astronomie statistique. Remontant en effet dans le passé, on pourrait faire débiter le sujet à Kepler, dont les trois lois fondamentales du système solaire tirées des nombreuses observations planétaires de Tycho-Brahé ne sont que des lois empiriques trouvant leur explication fonctionnelle dans la mécanique

de Newton, avec une définition corrélatrice des cadres inobservables d'espace et de temps. Lorsqu'on s'adresse à des ensembles beaucoup plus étendus d'objets sidéraux comme la Galaxie (10^{11} étoiles) dont seule une faible partie est observée, on conçoit tout d'abord la nécessité d'une caractérisation bien nette, d'une homogénéisation des données numériques, de la recherche statistique des distributions spatiales et cinétiques, de l'étude statistique des corrélations entre les facteurs qui les caractérisent, enfin d'une explication théorique où les concepts statistiques joueront un rôle majeur. Il faut tenir compte aussi de la lenteur (à notre échelle) de l'évolution et des mouvements apparents des objets célestes. En substituant aux observations étendues dans le temps celles faites en quelque sorte dans l'espace, en s'adressant à une multitude d'objets qui en représentent les phases, on réussit à définir avec assez bien de précision la manière dont cette évolution ou ces mouvements s'accomplissent. Enfin, les observations elles-mêmes sont difficiles : il faut quelques dizaines d'années pour obtenir un bon « mouvement propre » stellaire ou une parallaxe, les déplacements *relatifs* d'étoiles étant observés. Les données numériques sont entachées d'erreurs inévitables ; elles dépendent des méthodes et des instruments employés et doivent être rendues cohérentes, et c'est probablement dans ce domaine que la statistique s'est introduite pour la première fois en astronomie (il fut un temps où la théorie des erreurs d'observation et la méthode des moindres carrés étaient strictement l'apanage des astronomes et géodésiens).

Parmi les données stellaires les plus immédiatement représentatives, il convient de citer : a) le *type spectral* et la *magnitude apparente*, données physiques dont la dernière est convertie en *magnitude absolue* ou « éclat intrinsèque » si la distance est connue, b) la position sur la sphère céleste définie par les *coordonnées angulaires* (ascension droite, déclinaison), la distance déterminée par la *parallaxe* dont il existe divers types : parallaxes trigonométriques, spectroscopiques, dynamiques, hypothétiques ; données géométriques servant à préciser l'emplacement de l'étoile dans un trièdre choisi, c) les *mouvements propres* angulaires (variations des coordonnées) et la *vitesse radiale* (données cinématiques). La plupart de ces quantités sont difficilement mesurées et toutes n'ont pas la même abondance ; les parallaxes sont rares et peu précises, les vitesses radiales basées pourtant sur des mesures spectroscopiques pratiquement indépendantes de l'absorption interstellaire n'ont pas toujours la qualité ni l'homogénéité désirables. En outre, l'hémisphère austral est beaucoup moins connu que l'hémisphère boréal du ciel et les données y sont moins nombreuses. Les imperfections instrumentales et les difficultés d'observation, surtout s'il s'agit d'étoiles faibles, font que les *listes individuelles* de données provenant des divers observatoires sont

sans valeur pour les études d'ensemble de la Galaxie parce qu'hétérogènes. Au prix de comparaisons statistiques, ces listes sont groupées en *catalogues relatifs* ou secondaires déjà meilleurs (élimination d'erreurs systématiques, réduction d'erreurs accidentelles). Enfin, une étude très serrée et demandant des précautions considérables permet de constituer des *catalogues absolus* ou fondamentaux, véritables étalons, dont la précision interne et externe est excellente. De tels catalogues ne se rapportent qu'aux étoiles brillantes; il va de soi qu'un catalogue absolu contiendra peu de données. Les autres catalogues s'y rattachent au prix d'une étude statistique; on peut ainsi ramener un catalogue relatif au catalogue absolu correspondant s'il existe, enrichir le catalogue absolu au moyen des données du catalogue relatif, améliorer la précision de ce dernier. Ce problème de la comparaison statistique des catalogues astronomiques est capital, car c'est généralement sur les catalogues relatifs que portent les investigations de la statistique stellaire. Dans un autre domaine, les études photométriques portent souvent sur les clichés astrographiques, les magnitudes stellaires de référence étant empruntées aux catalogues de base.

Nous envisagerons dans la suite, et afin de donner une idée du sujet, quelques exemples particulièrement simples et représentatifs de ce genre d'investigation en essayant d'en montrer les implications.

Le diagramme Hertzsprung Russell et l'évolution stellaire.

Un bon exemple de diagramme statistique dont l'interprétation a été capitale dans l'évolution des idées sur le phénomène stellaire est le diagramme H.R., ou diagramme *magnitude absolue-type spectral*. La magnitude absolue représente la quantité totale de puissance rayonnée par l'étoile. Le type spectral donne *grosso modo* la température de l'atmosphère stellaire. Bien que l'atmosphère d'une étoile ne représente que la dix-millionième partie de la masse, on observe que les points s'y groupent en *séquences* caractérisées (séquence principale, région des naines blanches, branche des géantes, super-géantes, etc.). Il va de soi que le diagramme est construit à l'aide d'étoiles dont on connaît la magnitude apparente et la parallaxe, ce qui permet d'obtenir la magnitude absolue. Dès lors, une *extrapolation* fournit un premier usage pratique du diagramme : si l'on ne connaît que le type spectral et la magnitude apparente, il suffira de reconnaître la « classe de luminosité » (appartenance à l'une des séquences, ce qui se fait à l'aide de critères spectraux) pour pouvoir estimer la magnitude absolue. La formule classique $M = m - 5 \log_{10} (r_{pc}/10)$ donne alors la distance. D'autres classifications (notamment celle à trois paramètres de Chalonge-Barbier) donnent M

avec plus de précision à partir de critères uniquement spectraux mais n'ont pas conduit à autant d'implications sur le plan de l'évolution stellaire que le diagramme H.R. En effet, la séquence principale n'est autre qu'un ensemble de *points d'arrêt* dans l'évolution, l'étoile formée par condensation gravitationnelle à partir d'hydrogène (éventuellement enrichi d'éléments lourds) évoluant très vite, de haut en bas, pour atteindre la séquence principale où elle se fève en raison des processus d'interaction nucléaire, l'hydrogène brûlant en hélium. Elle quitte ensuite la séquence principale en se déplaçant vers le haut dès qu'il n'y a plus suffisamment d'hydrogène à brûler. On assiste alors à la combustion de l'hélium, du carbone, etc., l'étoile gagnant la branche des géantes. D'une manière générale, les régions vides de points sont rapidement traversées, l'étoile recherchant une autre interaction nucléaire. Elle s'effondre ensuite, par interaction gravitationnelle, vers la région des naines blanches. Ainsi le diagramme H.R. représente la synthèse progressive des éléments chimiques et il existe d'ailleurs une correspondance très nette entre ce diagramme, le tableau de Mendéléev, le diagramme charge et masse nucléaire. Ainsi le diagramme H.R., dont Hertzsprung disait avec modestie qu'il l'avait constitué parce qu'il ne disposait à cette époque que de deux éléments à mettre en parallèle, est fondamental. On peut établir maintenant au cœur des étoiles, dans des conditions impossibles au laboratoire (températures de quelques millions à des milliards de degrés, densités de 100 à 10^{16} g/cm³), la synthèse des éléments qui constituent l'univers et y lire à la fois la structure des noyaux atomiques.

Etudes géométriques et cinématiques : la statistique stellaire.

A côté des régressions d'ordre astrophysique dont bon nombre ont été étudiées (relation période-luminosité pour les Céphéides, relation masse-luminosité, etc.), il s'agit de reconnaître a) la distribution des étoiles dans l'espace, b) celle de leurs mouvements. On obtiendra ainsi les grandes lois de structure qui gouvernent la Galaxie. L'interprétation mécanique de ces lois conduira aux forces qui tiennent le système stellaire ensemble.

L'astronome ne fait aucune hypothèse *a priori* sur la forme des distributions statistiques qu'il étudie. Il est confronté avec un phénomène physique qui détermine la distribution. Les variations aléatoires qu'il suppose n'interviennent pas assez fréquemment, dans le domaine stellaire, ni avec assez d'intensité, pour faire tendre la distribution vers une structure gaussienne ou équivalente. Malgré les imperfections des données observées, il ne dispose que de celles-ci pour reconnaître sa structure. Aussi est-il conduit à utiliser

largement la *méthode des moments*. Il procède en général comme suit (pour fixer les idées, nous considérons une distribution à une variable) :

a) *établissement des moments*

$$\mu_n = \frac{1}{N} \sum_{k=1}^N (x_k)^n;$$

identification de ces moments à ceux de la population observée; construction formelle de la première fonction caractéristique

$$\phi(p) = \int_{-\infty}^{+\infty} f(x) e^{ipx} dx = \sum_0^{\infty} \frac{(ip)^n}{n!} \mu_n$$

$f(x)$ étant la densité;

b) *inversion de Fourier*; si tous les moments étaient connus, la densité serait

$$f(x) = \frac{1}{N} \sum_1^N \delta(x - x_k)$$

c) *adoucissement*; celui-ci consiste à rechercher une densité $f_r(x)$ continue positive ayant pour premiers moments $\mu_1 \dots \mu_r$ sans hypothèse sur les suivants (leur qualité laissant d'ailleurs à désirer). On peut adoucir en développant en série *limitée* de polynômes orthogonaux, la fonction-poids étant choisie, ou par limitation de la série représentative de $\phi(p)$, ou mieux de la *seconde* fonction caractéristique, développée en série des cumulants

$$\phi(p) = \exp S(p) = \exp \left\{ \sum \frac{(ip)^n}{n!} \lambda_n \right\}$$

d) *vérification*, sur l'échantillon, et extension à des échantillons plus vastes.

Cette méthode essentiellement pragmatique a effectivement pour but d'inclure la densité ou ses représentations (éventuellement son inverse de Fourier ou analogue) dans des équations différentielles ou intégrales dont on tentera la résolution.

A) *Distribution des étoiles dans l'espace.*

La relation entre magnitude apparente et absolue sera notée ici

$$M = m - R \quad (1)$$

R étant le « module de distance » ($-\infty < R < +\infty$). Les trois variables statistiques M , m , R sont liées fonctionnellement par (I) et distribuées statistiquement suivant une densité f_{mMR} . On peut écrire

$$f_{mMR} = f_m \cdot f_{(m)MR} = f_R \cdot f_{(R)M} \cdot f_{(RM)m}$$

$f_m(m)$ est accessible à l'observation car il suffit de compter photométriquement les étoiles (sur un cliché ou dans un catalogue). $f_R(R)$ représente la distribution radiale (dans une direction supposée déterminée). $f_{(R)M}$, distribution conditionnelle de M , jouit de la propriété presque toujours vérifiée d'être indépendante de R car il n'y a aucune raison autre que cosmogonique que l'éclat intrinsèque d'une étoile dépende de sa position dans la Galaxie. Enfin, $f_{(RM)m} = \delta(m - M - R)$ en vertu de (1), liaison fonctionnelle.

D'où

$$f_m(m) = \int f_R(R) \cdot f_M(m - R) dR \quad (2)$$

convolution entre limites infinies. Une seconde convolution s'obtient en introduisant une caractéristique $\beta(R)$ (p. ex. $1/r$, la parallaxe) dont la moyenne conditionnelle

$$\bar{\beta}(m) = \int \beta(R) \cdot f_{(m)R}(m, R) dR$$

peut être relevée; d'où

$$\bar{\beta}(m) \cdot f_m(m) = \int \beta(R) \cdot f_R(R) \cdot f_M(m - R) dR \quad (3)$$

Les convolutions (2, 3) sont algébrisées moyennant la transformation de Fourier, qui introduit les fonctions caractéristiques qu'on développe en termes des cumulants. Ceux-ci sont donc calculés par récurrence pour f_R et f_M . Dès lors, on connaît à la fois la distribution radiale et la « fonction de luminosité » stellaire. On voit ainsi combien les deux problèmes sont intimement liés, la distribution géométrique des étoiles dans l'espace s'obtenant en même temps que la distribution physique de leur état intrinsèque. On ne sait trop pour quelles raisons les équations (2, 3) ont été appelées « équations fondamentales » de la statistique stellaire.

B) Cinématique stellaire.

Seules les vitesses stellaires relativement au *Soleil* sont mesurées. On peut définir évidemment la vitesse d'une étoile relativement au repère galactocentrique V_{EA} (au centre de la Galaxie), relativement au centroïde local V_{EB} (au centre du groupe *local*), et l'on a

$$V_{ES} = V_{EB} + V_{BS} = V_{EA} + V_{AB} + V_{BS}$$

V_{EB} est la vitesse « résiduelle », $V_{SB} = -V_{BS}$ celle de l'*apex* solaire (mouvement du Soleil relativement au groupe local), V_{BA} n'est autre que la vitesse de courant d'étoiles (mouvement hydrodynamique). De telles relations très simples donnent des correspondances évidentes entre moments d'ordre successifs. Les moments

$$\mu_i(S) = \langle V_{ES}^i \rangle, \mu_{ik}(S) = \langle V_{ES}^i \cdot V_{ES}^k \rangle, \text{ etc.}$$

sont en principe obtenus à partir des observations. Ensuite

$$\mu_i(B) = 0, \mu_{ik}(S) = \mu_{ik}(B) + \mu_i(S) \mu_k(S), \text{ etc.}$$

$$\mu_{ik}(A) = \mu_{ik}(B) + \mu_i(A) \mu_k(A), \text{ etc.}$$

$\mu_i(A)$ ou bien V_{SA} n'est autre qu'un apex particulier, celui du Soleil relativement à une population fixe dans la Galaxie (ex. : la population II du halo galactique).

Dès que les moments sont connus, on peut obtenir la distribution correspondante pour les composantes de vitesse. A ce stade, on observe pourtant la nécessité de disposer de l'*ensemble* des données numériques (coordonnées angulaires, parallaxes, mouvements propres, vitesses radiales) pour chaque étoile de la population. De telles études effectuées sur les vitesses spatiales ne pourraient toucher que les étoiles proches ou intrinsèquement très brillantes. Pour les autres étoiles, les parallaxes sont inconnues et les mouvements propres généralement très imprécis. Heureusement, il est possible de remplacer les données manquantes par des données directionnelles et c'est là l'objet de la *méthode des aires*. La sphère céleste est partagée en un certain nombre d'aires limitées auxquelles on attache un trièdre local (Y) et l'ensemble est rapporté au trièdre central (p. ex. héliocentrique) désigné par (X); la matrice de passage O est connue pour chaque aire, de sorte que, pour les composantes de vitesse, $\dot{Y} = O \cdot \dot{X}$, et pour chaque aire (p), $\dot{Y}(p) = O(p) \cdot \dot{X}$. Dès lors, pour les moments $\mu_{ijk...}(p)$, (*)

$$\mu_{ijk...}(p) = O_{ii'}(p) \cdot O_{jj'}(p) \cdot O_{kk'}(p) \dots \mu_{i'j'k'...}(\dot{X})$$

Ordinairement, les vitesses radiales sont utilisées (moments $\mu_{III...}(p)$ pour chaque aire). Dès lors, un nombre suffisant d'aires permettra d'écrire des équations de condition, lesquelles, résolues par moindres carrés conformément au principe de Gauss-Markov, fourniront les moments cherchés des composantes spatiales dans le trièdre central. Divers perfectionnements peu-

(*) nous sommions sur les indices répétés.

vent être tentés, mais l'exemple, simplement esquissé, du calcul de l'apex solaire à partir des vitesses radiales suffit à faire comprendre le procédé. On n'aura ici à résoudre que les équations de condition

$$\mu_i(p) = O_{1i}(p) \mu_i(\dot{X})$$

d'où les équations normales

$$\sum_k [O_{1i} \ O_{1k}] \mu_k(\dot{X}) = [O_{1i} \ \mu_1]$$

L'exposé attire aussi l'attention sur les résidus de vitesse radiale et sur leur allure systématique dans le cas de certaines étoiles (notamment le type spectral B), les équations de condition devant être écrites sous la forme

$$\mu_i(p) = O_{1i}(p) \mu_i(\dot{X}) + K$$

de manière à absorber la valeur moyenne non nulle des vitesses radiales; ce terme K n'a pas reçu d'interprétation cohérente à l'heure actuelle, bien que les tentatives d'explication soient légion. On attire également l'attention sur le fait que les composantes transversales (mouvements propres) peuvent être utilisées et sur le perfectionnement qui en résulte pour les « constantes de précession », améliorant ainsi la mécanique du système solaire par des observations portant sur les étoiles.

L'usage des moments jusqu'au second ordre conduit déjà à des constatations intéressantes en cinématique de la Galaxie; à peu de choses près, tous les vecteurs d'apex relativement à des groupes distincts d'étoiles ont leurs extrémités alignées sur une droite du plan galactique et perpendiculaire à la direction Soleil-Centre du système (région du Sagittaire), ce qui s'explique aisément en admettant que les divers centroïdes sont animés de vitesses de *rotation*, le mouvement général du système stellaire s'effectuant autour d'un axe central normal au plan galactique avec une période calculable (env. 220 millions d'années). Quant aux vitesses résiduelles, elles sont distribuées autour de leur moyenne nulle d'une manière pratiquement ellipsoïdale; l'ellipsoïde des vitesses a un grand axe pointé vers le centre de la Galaxie et il s'agit ici d'une loi importante dont l'explication doit être recherchée en dynamique stellaire. Cette rotation d'ensemble est différentielle, les couches situées loin du centre tournant plus lentement, et ce nouveau fait requiert lui aussi une explication dynamique. L'utilisation des moments d'ordre supérieur conduit à diverses améliorations (déviation des vertex de Kapteyn, excès et asymétries). Une nouvelle relation importante est celle qui existe entre la dispersion et la vitesses de rotation, les systèmes les plus dispersés tournant plus lentement. Pour finir, le phénomène d'*asymétrie* constaté dans la distri-

bution mutuelle des ellipsoïdes peut s'interpréter par des considérations de moments angulaires orbitaux et de *stabilité*; la « seconde loi » de Strömberg montre que les étoiles tournant très vite sur leurs orbites galactiques ont beaucoup de chances de ne pas demeurer dans le système, le Soleil lui-même n'étant pas loin de pouvoir s'en échapper définitivement.

De telles lois empiriques ont un contenu pratique évident. En effet, si l'on a pu repérer le caractère différentiel de la rotation galactique, on connaît par le fait même la dépendance entre la vitesse radiale d'un constituant galactique, la distance et la direction. Dès lors, les observations de vitesse radiale faites dans une direction déterminée fournissent un *critère de distance* tandis que les observations d'intensité fournissent un *critère de densité*. On peut utiliser de telles relations pour interpréter notamment les profils observés de la raie interstellaire sur 21 cm de longueur d'onde (hydrogène neutre interstellaire), la déviation Doppler à partir de la longueur d'onde typique au laboratoire donnant la vitesse radiale et par suite la distance, et la hauteur du profil donnant la densité. Or, une telle raie radioastronomique est très peu sensible à l'absorption et les sondages galactiques effectués par ce moyen ont de ce fait un pouvoir pénétrant considérable. Nous voyons ainsi comment les observations astrophysiques modernes jointes aux résultats de la statistique stellaire ont permis d'établir le plan général de la Galaxie dans presque toute son étendue, alors que les observations stellaires elles-mêmes sont fortement limitées en profondeur par l'absorption des nuages cosmiques (gaz, poussières et grains interstellaires).

Interprétations théoriques.

Alors que les décomptes d'étoiles fournissent, du moins en principe, la densité de distribution des coordonnées, l'étude statistique des vitesses stellaires fournit la distribution locale des composantes de vitesse. Soient respectivement $f_r(r)$ et $f_{(r)v}(r, v)$ ces densités marginale et conditionnelle. Effectivement

$$f_{rv}(r, v) = f_r \cdot f_{(r)v}$$

est la *densité en phase*. Celle-ci est donc (partiellement) accessible aux observations. Il est donc évident que toute description théorique du système stellaire comportant, comme la Galaxie, un grand nombre d'individus, doit mettre en œuvre des concepts de mécanique statistique.

L'interprétation basée sur le problème des N corps (N étoiles ponctuelles s'attirant mutuellement suivant la loi de Newton) ne fournit que des indications générales sur la structure et l'évolution de la Galaxie. En parti-

culier, le théorème du viriel et le critère de Jacobi nous montrent qu'un système stellaire autogravitant ne peut être *stable* (c'est-à-dire se maintenir en lui-même dans l'espace sans se dissocier) que si l'énergie totale des mouvements stellaires (énergies cinétique plus potentielle) est *négative*. Cette condition nécessaire de stabilité donne une condition nécessaire sur la masse du système

$$M > M_{\text{crit}} = \bar{r} \cdot \bar{v}^2 / G$$

$\bar{r}$ étant la distance moyenne interstellaire, $\bar{v}^2$ le carré de la vitesse quadratique moyenne et G la constante de gravitation. Mais le problème des N corps ne nous apprend rien sur la structure géométrique et cinétique du système. Pour progresser, des concepts de mécanique statistique sont indispensables.

Reprenons les équations du mouvement d'une étoile individuelle dans un champ de potentiel $V(r)$; ces équations ont la forme

$$\dot{r} = v; \dot{v} = -\partial_r V$$

Si $f(r, v) = f_{rv}$ désigne la densité dans l'espace des phases, la conservation du nombre de particules par unité d'extension en phase donne ici l'équation aux dérivées partielles

$$\partial_t f + v \cdot \partial_r f - (\partial_r V) \cdot \partial_v f = 0 \quad (4)$$

à condition qu'il n'y ait entre étoiles que des interactions rapprochées négligeables (ceci est généralement adéquat pour la Galaxie, mais ne s'applique qu'aux systèmes peu denses). Le système différentiel est complété si l'on tient compte du fait que c'est la densité spatiale qui crée le potentiel :

$$\Delta V = 4 \pi G M \int f d^3 v \quad (5)$$

Ainsi les équations (4, 5) constituent un système différentiel qui doit permettre en principe de calculer la structure du système stellaire à tout instant si l'on connaît cette structure à un instant considéré comme initial. L'intégration n'en est pas commode et n'a été réussie que dans les cas où il existe des symétries (symétries centrales, symétries axiales), celles-ci donnant des intégrales premières du mouvement. D'une manière générale, la tendance actuelle consiste à utiliser ces intégrales premières dans des développements en série du type

$$f = \sum a_k I_k + \sum \sum a_{kl} I_k I_l + \dots$$

Ces développements sont injectés dans (4, 5) et les coefficients sont calculés de manière à satisfaire à la fois à l'équation de Liouville-Boltzmann sans

second membre (4) et à l'équation de Poisson (5). Mais on peut rechercher aussi des intégrales premières qui soient liées au caractère observé de la distribution des vitesses. Par exemple, le cas simple où la distribution est supposée ellipsoïdale est traité facilement

$$f = \phi (a_{ik} v_i v_k + \dots)$$

la forme quadratique inhomogène aux composantes de vitesse sous la fonction arbitraire ϕ devant être elle-même une intégrale première pour un système stationnaire (∂_t négligés). On obtient ainsi, par (4), un système différentiel aux caractéristiques $a_{ik}, \dots$ qui peut être intégré. En voici la solution pour un système stellaire à symétrie axiale (Q désigne cette forme quadratique) :

$$Q = \frac{V_r^2}{\alpha^2} + \frac{(V_\theta - V_{\theta 0})^2}{\beta^2} + \frac{V_z^2}{\gamma^2} + Q_0(r, z)$$

$$\alpha^2 = \frac{1}{a_0}; \quad \beta^2 = \frac{1}{a_0 + a'_0 r^2}; \quad \gamma^2 = \frac{1}{a''_0} \quad (a_0, a'_0, a''_0 \text{ constantes})$$

$$V_{\theta 0} = - \frac{a'_0 r}{a_0 + a'_0 r^2}$$

V_r, V_θ, V_z désignent les composantes de vitesse respectivement dans la direction radiale du plan galactique, dans la direction transversale et suivant la verticale du plan. On voit ainsi que l'ellipsoïde de vitesse pointera toujours vers le centre galactique, que cet ellipsoïde est centré sur le point moyen $(0, V_{\theta 0}, 0)$ et que la vitesse transversale moyenne est bien une vitesse de rotation autour d'un axe central. Si l'on écrit cette dépendance sous la forme

$$V_{\theta 0} = \Omega(r) \cdot r$$

alors on constate sur la solution que

$$\frac{r}{2} \frac{d\Omega}{dr} = \Omega \left(1 - \frac{\beta^2}{\alpha^2}\right)$$

Or, la quantité au premier membre est accessible aux observations (rotation différentielle de la Galaxie), tandis que les différents termes du second membre sont aussi mesurables. Il existe donc une relation entre le caractère différentiel de la rotation galactique et les dispersions, cette relation pouvant servir au contrôle de la théorie.

Quelques remarques s'imposent quant au caractère du potentiel V introduit dans les équations. Pour un système de masses discrètes distribuées d'une

manière aléatoire, ce potentiel doit-il encore être considéré comme une fonction ordinaire des coordonnées, ou bien ne faut-il pas l'envisager plutôt comme une variable aléatoire ?

La situation paraît bien la suivante : V est lié à la densité massique ρ par l'équation de Poisson

$$\Delta V = 4 \pi G \rho$$

Pour un système de masses discrètes (toutes égales à m) occupant des emplacements $\underline{r}_a$ ($a = 1, 2, \dots, N$) la densité est

$$\rho = m \sum_{a=1}^N \delta(\underline{r} - \underline{r}_a)$$

En représentant les fonctions δ par des intégrales de Fourier

$$\rho = \frac{m}{(2\pi)^3} \sum_{a=1}^N \int e^{-i\underline{k}(\underline{r} - \underline{r}_a)} d^3 \underline{k}$$

on obtient aisément la solution correspondante

$$V(\underline{r}) = - \frac{4\pi G m}{(2\pi)^3} \sum_1^N \int \frac{d^3 \underline{k}}{|\underline{k}|} e^{-i\underline{k}(\underline{r} - \underline{r}_a)} = - G m \sum_1^N \frac{\sigma |\underline{r} - \underline{r}_a|}{|\underline{r} - \underline{r}_a|} \quad (6)$$

où σ est la fonction-signe (notons en passant que la self-énergie gravitationnelle du système est ainsi complètement évitée).

Mais on adopte une description statistique dans laquelle la distribution spatiale des masses est représentée par $f_r(\underline{r})$. En conséquence, toutes les variables $r_1 \dots r_N$ sont distribuées suivant f_r , et comme on peut permuer chaque étoile mutuellement sans changer le système, ces variables sont stochastiquement indépendantes, la distribution couplée étant simplement

$$F = \prod_{a=1}^N f_r(\underline{r}_a)$$

Nous avons donc à considérer $3N$ variables $r_1 \dots r_N$ et une variable supplémentaire $V(r, r_1 \dots r_N)$ liée fonctionnellement à celles-ci par (6). En résultat, les valeurs du potentiel en chaque point spécifié r sont distribuées suivant la densité

$$f_v(V) = \int \prod_{a=1}^N f_r(\underline{r}_a) \cdot \delta(V + G m \sum \frac{\sigma(\underline{r} - \underline{r}_a)}{|\underline{r} - \underline{r}_a|}) d^3 r_1 \dots d^3 r_N$$

On peut en déterminer la fonction caractéristique

$$\phi_v = \int f_v e^{ipv} dp = \left\{ \int f_r(\underline{r}') \exp(-ip G m \frac{\sigma |\underline{r} - \underline{r}'|}{|\underline{r} - \underline{r}'|}) d^3 \underline{r}' \right\}^N$$

et les cumulants en posant

$$\phi_v = e^{\sum \frac{(ip)^n}{n!} \lambda_n(v)};$$

$$\int f_r(\underline{r}') e^{(-ip G m \frac{\sigma (\underline{r} - \underline{r}')}{|\underline{r} - \underline{r}'|})} d^3 \underline{r}' = e^{\sum \frac{(ip)^n}{n!} X_n(r)}$$

d'où simplement

$$\lambda_n = N X_n$$

Soit

$$J_1 = \int \frac{\rho(\underline{r}') d^3 \underline{r}'}{|\underline{r} - \underline{r}'|}; \dots; J_n = \int \frac{\rho(\underline{r}') d^3 \underline{r}'}{|\underline{r} - \underline{r}'|^n}; \bar{\rho} = Nm f_r(\underline{r})$$

Alors

$$\lambda_1 = -G J_1; \quad \lambda_2 = \frac{G^2}{N} \{M J_2 - J_1^2\};$$

$$J_3 = \frac{G^3}{N^2} \{-M^2 J_3 + 3 M J_2 J_1 - 2 J_1^3\}, \text{ etc.}$$

Les cumulants peuvent être ainsi évalués à partir d'intégrales simples calculables si l'on connaît la densité moyenne. Notons que λ_1 satisfait à l'équation de Poisson

$$\Delta \lambda_1 = 4 \pi G \bar{\rho}$$

M représente la masse totale Nm finie du système; aussi λ_2, λ_3 etc. tendront-ils généralement vers zéro quand N augmente indéfiniment. Ainsi les valeurs locales du potentiel présentent-elles un « bruit de fond » qui s'atténue à mesure que le système devient plus peuplé. Le système est évidemment coupé dès que les valeurs du potentiel deviennent positives, c'est-à-dire dès que le bruit de fond coupe l'axe r dans le diagramme (V, r) . Ceci définit pour le système stellaire un diamètre effectif. De plus, les trajectoires stellaires elles-mêmes, gouvernées par le potentiel, deviennent aléatoires en ce sens qu'à une situation initiale correspondra une distribution statistique de positions successives, celle-ci créant $f_r(r)$.

En résumé, cette conception statistique paraît offrir une vision plus vaste de la dynamique stellaire, mais demanderait à être traitée avec plus de détails.

PUBLICATIONS REÇUES

ONTVANGEN PUBLICATIES

- 1) Cuadernos de Estadística Aplicada e Investigación Operativa (Barcelona), Vol. IV Fasc. 1-2-3, 1965.
- 2) INTIN (Instituto Nacional de Tecnología Industrial) N° 13 Año III - N° 14, Año IV (Buenos-Aires).
- 3) Trabajos de Estadística y de Investigación Operativa, Vol. XVII, Cuaderno I, Madrid.
- 4) Journal of the Operations Research Society of Japan Vol. 8, N°s 3 & 4, June 1966 & October 1966.
- 5) Boletín de la Sociedad Argentina de Investigación Operativa, (Buenos-Aires), n°s 12, 13 et 14, (3^e et 4^e trimestres 1965 - 1^{er}, 2^e et 3^e trimestres 1966).
- 6) Revue Française de Recherche Opérationnelle N°s 40 et 41 (3^e et 4^e trimestres 1966).
- 7) Tijd, Economisch Weekblad uitgegeven door het Vlaams Economisch Verbond, N° 1/67.
- 8) Oost-Vlaanderen groeit (Tijdschrift van de economische raad voor Oost-Vlaanderen) N° 4, 2^e jaargang.
- 9) Annales de Sciences Economiques Appliquées N°s 3 et 4 septembre-novembre 1966.
- 10) Revue IBN (Institut belge de Normalisation) N°s 9-10-11-12 - 1 septembre - octobre - novembre - décembre 1966 - janvier 1967 — BIN Revue (Belgisch Instituut voor Normalisatie), n°s 9, 10, 11, 12, 1 september, oktober, november, december 1966, januari 1967.
- 11) Bulletin de l'AFCIQ (Association Française pour le Contrôle industriel de la Qualité) Vol. II N°s 3 & 4.
- 12) Revue Suisse pour l'Organisation Industrielle (Institut d'organisation Industrielle de l'Ecole Polytechnique Fédérale à Zürich) N°s 8-9-10-11-12 (1966) - N° 1 (1967).
- 13) Quality EOQC Journal, Rotterdam, Vol. X, n° 3, Autumn 1966.
- 14) Bulletin Economique pour l'Europe (Nations Unies) Vol. 17, N°s 1 & 2.
- 15) Operation Research Quality, London, 17, n° 4, 1966 et 18, n° 1, 1967.
- 16) Revue française de statistique appliquée, Paris, 14, n°s 3 et 4, 1966.
- 17) Revue de l'Institut international de Statistique, La Haye, 34, n°s 1, 2 et 3, 1966.

DEMANDE D'EMPLOI

LIC. SC. ECON. Appliquées, 27 ans, cherche emploi en GESTION SCIENTIFIQUE. Expérience approfondie en Programmation linéaire, Techniques budgétaires, Analyse des résultats et Méthodes statistiques.

Pratique la programmation en FORTRAN, Simulation et PERT. - Connaissance parfaite du Français, du Néerlandais et de l'Anglais, bonne connaissance de l'Allemand.

Prière de faire offre sous double enveloppe au Secrétariat de la revue qui transmettra.

OUVRAGE REÇU**ONTVANGEN BOEK**

B.L. Van der WAERDEN, *Statistique mathématique*, traduit de l'Allemand par C. Guinchat, Documentaliste au Centre d'études sociologiques et A. Degenne, Statisticien au Centre d'études sociologiques. Dunod - Editeur - 92, rue Bonaparte - Paris 6ème — 384 pages 16 X 25, avec 39 figures, 1967. Relié toile sous jaquette. Prix : 68 F.

L'ouvrage, récemment publié chez DUNOD, expose les principes théoriques en statistique mathématique et débouche sur les applications.

En effet, cet ouvrage est construit de telle manière que les parties théoriques se trouvent rassemblées dans un ensemble de chapitres et les aspects pratiques développés dans des chapitres distincts. Ainsi, les lecteurs qui n'ont pas une formation suffisante pour suivre les développements théoriques peuvent se contenter des chapitres d'application.

On y trouve une discussion des tests statistiques les plus couramment utilisés et un exposé des conditions strictes de leur emploi.

De nombreux exemples d'application illustrent chaque exposé théorique ; choisis dans les différentes branches où les statistiques sont d'une utilisation constante (météorologie, biologie, médecine, économie, industrie... etc.), ils sont réels.

Toutes les catégories d'utilisateurs et spécialement les étudiants en sciences, pour préparer les certificats de statistique, les enseignants et les chercheurs en sciences humaines, physiques ou biologiques, les ingénieurs et techniciens qui ont à utiliser la statistique, trouveront dans le livre de Van der Waerden un exposé, d'un ton nouveau en langue française, de l'essentiel des fondements théoriques et des principaux modèles qui constituent la statistique mathématique.

Prix de vente

Au numéro : Belgique 75 FB
Etranger 90 FB
Abonnement : Belgique 250 FB
(4 numéros) Etranger 300 FB

Tarif de publicité
(4 numéros)

La page : 5.000 F
La 1/2 page : 3.000 F
Le 1/4 page : 2.000 F

Les frais de clichés sont à charge de l'annonceur.

Publications d'articles

- 1) La Revue est ouverte aux articles traitant de statistique pure et appliquée, de recherche opérationnelle et de « quality control ».
- 2) Les manuscrits seront dactylographiés et peuvent être envoyés au secrétariat de la Revue : 66, rue de Neufchâtel, Bruxelles 6.
- 3) Les auteurs d'articles techniques recevront 25 tirés à part de leurs textes.
- 4) La responsabilité des articles n'incombe qu'à leurs auteurs.

Verkoopprijs

Per nummer : België 75 BF
Buitenland 90 BF
Abonnement : België 250 BF
(4 nummers) Buitenland 300 BF

Advertentietarief
(4 nummers)

Per bladzijde : 5.000 F
Per 1/2 bladzijde : 3.000 F
Per 1/4 bladzijde : 2.000 F

De cliché-onkosten vallen ten laste van de adverteerders.

Publicaties van artikels

- 1) Het Tijdschrift neemt artikels aan over wiskundige statistiek en toepassingen, over operationeel onderzoek en kwaliteitszorg.
- 2) De teksten dienen getipt gestuurd te worden naar het secretariaat van het Tijdschrift : 66, Neufchâtelstraat, Brussel 6.
- 3) De auteurs ontvangen 25 overdrukken van de technische artikels.
- 4) De auteurs zijn alleen verantwoordelijk voor de inhoud van hun teksten.