

Zeitschrift: Mitteilungen / Vereinigung Schweizerischer Versicherungsmathematiker
= Bulletin / Association des Actuaires Suisses = Bulletin / Association of
Swiss Actuaries

Band: 50 (1950)

Artikel: La distribution des revenus dans la technique mathématique de la
sécurité sociale

Autor: Kaiser, Ernest

DOI: <https://doi.org/10.5169/seals-966871>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. [Siehe Rechtliche Hinweise.](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. [Voir Informations légales.](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. [See Legal notice.](#)

Download PDF: 31.10.2024

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

La distribution des revenus dans la technique mathématique de la sécurité sociale

Par *Ernest Kaiser*, Berne

Introduction

On ne saurait pratiquer de nos jours une politique sociale avisée sans connaître la répartition de la population selon le montant des revenus. Cette constatation se vérifie tout spécialement dans le domaine de la sécurité sociale qui constitue l'une des branches les plus importantes de cette politique. Il n'est dès lors pas étonnant que des sociologues et des économistes de renom aient abordé l'étude de ce problème en utilisant en premier lieu les méthodes de la statistique descriptive. Puis, l'analyse des observations faites ayant permis de déceler certaines régularités dans le phénomène étudié, on s'avisa de proposer des formules analytiques pour ajuster plus ou moins fidèlement les fréquences expérimentales. C'est ainsi que d'éminents mathématiciens — M. Fréchet, par exemple — n'ont pas dédaigné d'étudier le problème dans le cadre des mathématiques appliquées.

Les estimations financières concernant la sécurité sociale rendent particulièrement manifeste la nécessité de connaître les distributions de revenus. En effet, le but poursuivi par la sécurité sociale est, avant tout, de venir en aide aux personnes dont les ressources sont faibles. Or, la notion d'économiquement faible ne pourra être définie, au point de vue mathématique, qu'à l'aide de la répartition des revenus. Pour traiter ce problème on s'est contenté, jusqu'alors, d'utiliser la distribution des revenus propre à l'ensemble de la population. Cependant, les revenus variant sensiblement avec l'âge, qui est l'élément essentiel pour la détermination du risque, la résolution de certains problèmes rend nécessaire de connaître, non seulement cette distribution globale,

mais encore celle particulière à chaque classe d'âge. Ce fut précisément le cas lorsqu'il s'est agi de calculer les rentes moyennes de l'assurance-vieillesse et survivants suisse [22]¹⁾, pour la détermination desquelles on a supposé que les revenus des personnes appartenant à une classe d'âge donné se répartissent selon la distribution semi-normale. Nous avons constaté par la suite que la méthode utilisée dans ce cas particulier pouvait se généraliser et s'appliquer à n'importe quelle distribution. Ainsi, nous avons été amené à étudier le problème de la répartition des revenus à l'aide de deux ou même trois variables.

¹⁾ Voir liste bibliographique à l'annexe I.

Chapitre I

Bases mathématiques de la théorie

§ 1. Variables et fonctions dans l'espace bio-économétrique (t, u, x)

1. Les questions financières de la sécurité sociale se ramènent essentiellement à des problèmes dépendant de trois variables, *le temps* t , *le revenu* u et *l'âge* x . D'une part, les prestations ne sont souvent attribuées qu'à la portion de l'effectif $L_t(x) dx$ vivant à l'instant t et d'âge x à $x + dx$ dont les membres ne disposent que d'un revenu inférieur à une certaine limite u_1 et, d'autre part, le montant même de ces prestations est calculé selon une fonction plus ou moins compliquée du revenu u . Nous donnerons des exemples au chapitre III de ce travail. Dès lors, il ne suffit plus de connaître l'effectif $L_t(x) dx$; il faut savoir quelle est sa distribution selon le revenu u . Pour effectuer correctement les calculs, on est donc conduit à envisager des effectifs $L_t(u, x) du dx$, vivant à l'instant t , ayant l'âge x à $x + dx$ et possédant un revenu de u à $u + du$. Nous sommes donc en présence de fonctions à trois variables réelles qu'on attachera à un point $A(t, u, x)$ de l'espace, où l'on mesurera ces trois variables sur les axes de référence d'un trièdre orthogonal (voir graphique n° 1 de l'annexe II). De cet espace, nous pouvons extraire des plans particulièrement intéressants au point de vue actuariel. D'une part, nous aurons *le plan démographique ou bio-métrique* (t, x) dans lequel nous convenons de porter t en abscisse et x en ordonnée; ainsi, les effectifs vivant au même instant t seront disposés le long de droites verticales: $t = \text{constant}$; en revanche, les effectifs dont les membres sont nés au même instant t_0 , c'est-à-dire appartenant à la même génération, évolueront dans le plan biométrique le long de leur ligne de vie représentée par une droite diagonale d'équation $x = t - t_0$. Ainsi, certains auteurs parlent de problème vertical et de problème diagonal. Nous désignerons les effectifs situés sur la même diagonale de paramètre t_0 comme étant liés biométriquement entre eux. D'autre part, nous envisagerons *la famille des plans économétriques* $(u, x)_t$, ces derniers étant, au point t , perpendiculaires à l'axe du temps. Ces remarques nous autorisent à désigner l'espace entier

considéré comme *l'espace bio-économétrique*. Les effectifs $L_t(u, x)$ liés biométriquement se trouveront alors sur des plans orthogonaux au plan biométrique et dont les traces dans ce dernier seront les lignes de vie diagonales.

Pour le calcul des effectifs $L_t(x)$ l'actuaire dispose des *fonctions biométriques*, avant tout des fonctions de survie et de renouvellement. Les théories scientifiques de l'actuariat ont été presque exclusivement consacrées à ce sujet et constituent un édifice pour ainsi dire terminé. En revanche, la technique de calcul relative aux effectifs $L_t(u, x)$ n'a guère été traitée dans son ensemble. Nous y consacrerons donc l'essentiel de cette étude. Cette technique repose avant tout sur des *fonctions économétriques* que nous allons tirer entièrement de la théorie de la distribution des revenus.

D'autres critères, avant tout *le sexe et l'état civil*, jouent un rôle important dans ces questions. De même que l'on est amené sur le plan démographique à diviser la population en différents groupes, d'indice ν , relatifs au sexe et à l'état civil, et à considérer leur structure d'âge séparément, il est indiqué de procéder d'une manière analogue sur le plan économétrique et d'associer une distribution de revenus propre à chaque groupe d'indice ν .

2. Le problème de la répartition des revenus de la population entière a été considéré jusqu'ici par tous les auteurs sous l'angle d'une distribution continue à une seule variable. Comme nous l'avons déjà fait remarquer, la résolution des problèmes les plus importants relatifs à l'équilibre financier de certaines branches de la sécurité sociale nécessite de traiter cette question à l'aide de distributions continues à deux ou même trois variables. Cependant, l'aspect économétrique de notre théorie étant déterminé par l'introduction de la variable u , il est indiqué d'envisager d'abord *les distributions de revenus à une seule variable*. On sera d'ailleurs souvent amené à extraire ces dernières du problème à deux ou trois variables. Rappelons donc les propriétés fondamentales des distributions à une variable et introduisons en même temps les notations convenant à notre théorie. Nous nous bornerons toutefois à relever les fonctions qui nous seront indispensables dans la suite, tout en renvoyant le lecteur désireux de trouver de plus amples détails et de connaître les autres notions, aux traités récents sur la statistique mathématique moderne, parmi lesquels nous signalons ceux de Cramér [8], de Féraud [10] et de Linder [27].

En introduisant le revenu u comme variable indépendante, prenant toutes les valeurs de 0 à ∞ , nous envisagerons comme donnée fondamentale de la distribution des revenus *la fonction de fréquence* $f(u)$, supposée non négative, continue et dérivable. Celle-ci dépendra, en général, de un ou de plusieurs paramètres que nous introduirons au fur et à mesure des besoins. Nous aurons soin de noter $f(u)$ sous sa forme canonique ou normée, c'est-à-dire que les paramètres seront déterminés de telle sorte que soit satisfaite la loi des aires :

$$\int_0^{\infty} f(u) du = 1. \quad (1)$$

La fonction $f(u)$ ayant ainsi la dimension d'une fréquence relative, il convient de rappeler que pour passer, dans les applications pratiques, aux fréquences absolues, il faut la multiplier par le nombre L de personnes formant l'ensemble dont on étudie la distribution des revenus.

Partant de $f(u)$, on définit *la fonction de répartition ou fonction cumulative* $F(u)$ de la manière classique suivante :

$$F(u) = \int_0^u f(\eta) d\eta; \quad (2)$$

sous cette forme, elle donnera la proportion des personnes dont le revenu est inférieur ou égal à u . Il est évident que $F(u)$ est une fonction jamais décroissante, bornée inférieurement par $F(0) = 0$ et supérieurement par $F(\infty) = 1$. Comme nous le verrons au chapitre II à propos de la distribution de Pareto, il est parfois plus simple d'opérer avec *la fonction cumulative complémentaire* $H(u)$ qui donne la proportion des personnes dont le revenu est supérieur à u et qui est donc définie par :

$$H(u) = \int_u^{\infty} f(\eta) d\eta = 1 - F(u). \quad (2')$$

Aussi bien au point de vue théorique qu'au point de vue pratique, les moments de la distribution jouent un rôle primordial. *Les moments d'ordre* k : ${}_aM_k$ par rapport à un point a de l'abscisse u , sont donnés par l'expression :

$${}_aM_k = \int_0^{\infty} (u - a)^k f(u) du. \quad (3)$$

On sait que la suite complète des moments peut caractériser entièrement une distribution. Lorsque les moments seront pris par rapport à l'origine, nous omettrons l'indice $a = 0$ et nous noterons M_k ; de plus nous négligerons l'indice $k = 1$ de sorte que M indiquera la moyenne de la distribution par rapport à l'origine. Les moments par rapport à $a = M$ seront notés par la lettre gothique \mathfrak{M}_k ; rappelons que la dispersion (écart quadratique moyen, «standard deviation») est définie par $\mu_2 = \sqrt{\mathfrak{M}_2}$. Dans les applications, on sera souvent obligé de passer des moments pris par rapport à un point a à ceux pris par rapport à un autre point b et inversement. En supposant $b > a$, les formules de passage peuvent s'écrire de la manière symbolique suivante [10]:

$${}_bM_k = [{}_aM - (b - a)]^k, \quad (3')$$

$${}_aM_k = [{}_bM + (b - a)]^k; \quad (3'')$$

on aura soin de substituer dans le développement des binômes les moments M_i aux puissances M^i . Relevons le cas particulier important où $a = 0$ et $b = M$, pour lequel (3') donnera \mathfrak{M}_k à l'aide des moments M_i , ($1 \leq i \leq k$), par rapport à l'origine. Rappelons à ce propos, la relation bien connue: $\mathfrak{M}_2 = M_2 - M^2$.

En plus de ces notions, importantes pour tous les problèmes relevant du calcul des distributions, nous introduirons encore deux quantités ayant une signification spéciale pour les distributions de revenus. Les fonctions $\Phi(u)$ désignant la somme réduite des revenus $\leq u$ et $\Theta(u)$ la somme réduite complémentaire des revenus $> u$, doivent être calculées à l'aide des deux formules suivantes:

$$\Phi(u) = \int_0^u \eta f(\eta) d\eta, \quad (4)$$

$$\Theta(u) = \int_u^\infty \eta f(\eta) d\eta = M - \Phi(u); \quad (4')$$

il est évident que $\Phi(\infty) = \Theta(0) = M$. Rappelons que nous avons désigné par L le nombre total des personnes composant la population considérée; la somme effective des revenus $\leq u$ sera alors donnée par $L\Phi(u)$, la somme des revenus $> u$ par $L\Theta(u)$ et le revenu total de toute la population envisagée par LM . C'est pourquoi nous avons introduit plus haut l'appellation de «somme réduite» pour désigner les fonctions Φ et Θ .

Plusieurs auteurs ont proposé encore d'autres notions pour caractériser la distribution des revenus. Cependant toutes ces grandeurs typiques peuvent être définies à l'aide des fonctions introduites ci-dessus. Ainsi, on pourrait obtenir les *moyennes partielles*, par exemple, en divisant soit (4) par (2), soit (4') par (2'). Dans le même ordre d'idées il nous paraît utile de rappeler la notion d'*indice de concentration* introduite par Corrado Gini [17] et qui a récemment fait l'objet d'un travail en Suisse [4]. En partant de (4'), on peut définir la fonction

$$\theta(u) = \frac{\Theta(u)}{M}; \quad (4'')$$

puisque $\theta(0) = 1$, on peut désigner cette quantité, donnant la proportion de la somme des revenus supérieurs à u par rapport au revenu total de la population, comme somme complémentaire normée. L'indice de concentration $\iota(u)$ est alors déterminé par la relation:

$$\theta^{\iota(u)}(u) = H(u), \quad (5)$$

ou, ce qui revient au même, par

$$\iota(u) = \frac{\log H(u)}{\log \theta(u)}. \quad (5')$$

La signification concrète de ι ressort de (5) si l'on se rappelle qu'aussi bien θ que H varient de 0 à 1; en effet, si la distribution envisagée donne en un point u : $\theta = \frac{1}{10}$ et $H = \frac{1}{100}$, on peut écrire $\left(\frac{1}{10}\right)^2 = \frac{1}{100}$, c'est-à-dire $\iota = 2$. Ainsi, un dixième de la somme totale des revenus est détenu par un centième des personnes envisagées, ce centième appartenant au groupe des personnes à revenus élevés.

D'autres quantités sont encore utilisées pour la description d'une distribution; cependant, elles ne la déterminent en général pas entièrement. Signalons à ce titre *les quantiles* définis par les abscisses u de manière à ce que $F(u) = \frac{j}{i}$, où $j < i$ sont des nombres entiers (médiane si $i = 2$, quartiles si $i = 4$, déciles si $i = 10$, percentiles si $i = 100$). Une autre grandeur typique est *le mode ou la dominante*, abscisse du maximum de $f(u)$; on suppose ainsi l'existence d'un seul maximum; les distributions jouissant de cette propriété sont appelées unimodales.

3. Considérons maintenant un ensemble de personnes dont chacune d'elles est caractérisée simultanément par *deux variables, le revenu u et l'âge x* . Nous situons ainsi le problème à l'intérieur d'un plan économétrique (u, x) . Dans ce plan nous délimitons d'abord un domaine D obéissant aux inégalités: $0 \leq u < \infty$ et $x_0 \leq x \leq x_1$, où x_0 signifie la limite d'âge inférieure et x_1 la limite d'âge supérieure entrant en ligne de compte. A l'intérieur de D nous définissons *la fonction de fréquence $\varphi(u, x)$* , non négative, continue et dérivable par rapport aux deux variables. Nous choisissons d'emblée pour $\varphi(u, x)$ la forme canonique, c'est-à-dire nous exigeons que φ soit normé dans le plan (u, x) , de sorte que:

$$\int_0^{\infty} \int_{x_0}^{x_1} \varphi(u, x) du dx = 1. \quad (6)$$

Nous dirons que $\varphi(u, x)$ définit la distribution simultanée de la population considérée suivant les deux variables u et x .

De cette distribution à deux dimensions dont la fonction de fréquence est $\varphi(u, x)$ nous allons tirer, d'une part, par projection sur les deux axes *deux distributions marginales* [8] et, d'autre part, en considérant à tour de rôle chacune des deux variables comme paramètre, *deux familles de distributions conditionnelles*. Il s'agit ici du problème classique du calcul des distributions à deux variables dont nous jugeons également utile de rappeler les grandes lignes. Par la suite — et en cela réside le fond même de ce travail — nous aborderons par un autre côté le lien existant entre ces fonctions. Mais examinons d'abord comment s'effectuent les opérations classiques dont il vient d'être question.

Commençons par la projection de $\varphi(u, x)$ sur l'axe des ordonnées x qui est obtenue par l'intégration suivante en u :

$$\lambda(x) = \int_0^{\infty} \varphi(u, x) du. \quad (7)$$

La masse statistique complète est ainsi rangée selon l'âge x de sorte que $\lambda(x)$ donne *la structure d'âge* de la population envisagée, et ceci sans égard aux revenus. Cette fonction de fréquence marginale à une variable $\lambda(x)$ est normée sur l'axe des x puisque $\varphi(u, x)$ l'est dans le plan.

Nous sommes maintenant à même de définir *la distribution des revenus d'un ensemble d'âge x* . En considérant x pour un moment

comme une constante, il vient pour la fonction de fréquence de cette distribution conditionnelle :

$$f(u,x) = \frac{\varphi(u,x)}{\lambda(x)}. \quad (8)$$

En faisant varier x entre ses deux limites x_0 et x_1 , nous obtenons une famille de distributions conditionnelles de revenus relatives aux différents âges x . La nouvelle fonction à deux variables $f(u,x)$ définit ainsi des distributions de revenus sur des droites parallèles à l'axe des u . En intégrant (8) en u entre 0 et ∞ on voit que les fonctions $f(u,x)$ sont normées sur ces droites. De $f(u,x)$ on pourra déduire toutes les quantités que nous avons introduites sous le n° 2.

En projetant $\varphi(u,x)$ sur l'axe des u , nous définissons la fonction de fréquence de la *distribution globale des revenus* de la manière que voici :

$$\bar{f}(u) = \int_{x_0}^{x_1} \varphi(u,x) dx. \quad (9)$$

Cette distribution marginale ne dépend plus de l'âge x . En intégrant en u de 0 à ∞ , on retrouve (6), ce qui signifie non seulement que $\bar{f}(u)$ est normé sur l'axe des u , mais encore que toute la masse statistique complète est rangée selon la variable u . Ainsi, on a bien la distribution globale des revenus de l'ensemble complet des personnes considérées, quel que soit leur âge. Ici encore, nous pouvons appliquer la théorie des distributions à une variable, exposée sous n° 2. Nous reviendrons d'ailleurs sur cette question au § 4 ci-après. — Il va sans dire qu'on pourra encore obtenir, à part les fonctions $f(u,x)$, une deuxième famille de distributions conditionnelles dont les fonctions de fréquence s'obtiennent en divisant $\varphi(u,x)$ par $\bar{f}(u)$, ce qui nous donne la structure d'âge spécifique relative à un revenu u donné.

4. Comme nous l'avons relevé au n° 1, les problèmes financiers de la sécurité sociale introduisent, outre les variables u et x , la variable du temps t . Les distributions à deux dimensions envisagées au n° 3 varieront, en général, d'un plan économétrique à l'autre; en d'autres termes on sera amené à étudier ce qui se passe à l'intérieur de différents plans parallèles $(u,x)_t$ de l'espace (t,u,x) . Ainsi, la *distribution des revenus dans les plans* $(u,x)_t$ sera caractérisée par la fonction à trois

variables $\varphi_t(u, x)$ et qui donne la densité des personnes vivant à l'instant t , ayant le revenu u et l'âge x . On en tire, à l'aide des formules (7) à (9), écrites avec l'indice t , les trois distributions suivantes à une seule dimension, dont les fonctions de fréquence seront notées par :

$\lambda_t(x)$: définissant des distributions marginales donnant la structure d'âge de la population à l'instant t (fonctions de fréquence normées sur des droites parallèles à l'axe des x à l'intérieur du plan $u = 0$);

$f_t(u, x)$: déterminant des distributions conditionnelles qui définissent la répartition des revenus d'un ensemble d'âge x vivant à l'instant t (fonctions de fréquence normées sur des droites parallèles à l'axe des u);

$\bar{f}_t(u)$: définissant des distributions marginales donnant la répartition globale des revenus de la population vivant à l'instant t (fonctions de fréquence normées sur des droites parallèles à l'axe des u à l'intérieur du plan $x = 0$).

Il sera utile de consulter à propos de ces distributions le graphique n° 1 de l'annexe II.

Les observations statistiques ne donnant pour ainsi dire jamais la fonction $\varphi_t(u, x)$, nous ne pouvons pas procéder de la manière directe envisagée ci-dessus. Les données empiriques sont même insuffisantes pour déterminer toutes les fonctions nécessaires pour calculer $\varphi_t(u, x)$. Pour résoudre ce *problème fondamental*, il est dès lors indispensable d'introduire à côté des quelques fonctions fournies par l'observation statistique des hypothèses plausibles. D'ailleurs, même si l'observation statistique nous permettait de déterminer complètement la fonction $\varphi_t(u, x)$, elle ne serait valable que pour le passé. Or, le rôle essentiel de l'actuaire est de s'occuper de l'avenir; il est vrai qu'il tire son instrument de calcul numérique des observations statistiques, mais il doit souvent y apporter des modifications qui tiennent compte des tendances d'évolution déjà constatées. Comme on effectue en général les calculs démographiques séparément des estimations économétriques, il est plus indiqué et plus simple de procéder aux modifications mentionnées en faisant des hypothèses sur les éléments constitutifs de la fonction $\varphi_t(u, x)$. Ainsi, la méthode que nous exposerons par la suite gardera toujours, du moins dans ses principes, toute son utilité.

§ 2. L'échelle des revenus moyens et son inversion

5. Avant d'aborder le problème qui nous préoccupe, il est nécessaire d'introduire une fonction importante, l'une des trois fonctions fondamentales qui serviront précisément à calculer $\varphi_t(u, x)$, les deux autres découlant sans plus de ce qui a été dit au n° 4. Il s'agit de la loi de variation du revenu moyen avec l'âge x que nous appellerons « échelle des revenus moyens », conformément à la terminologie employée par les actuaires anglais, King [24], par exemple. Plaçons-nous à l'intérieur du plan vertical $t = 0$ de l'espace (t, u, x) et considérons les distributions de revenus le long de deux droites parallèles à l'axe des u , l'une correspondant à l'âge x et l'autre à la limite d'âge inférieure x_0 . En omettant l'indice $t = 0$, les deux fonctions de fréquence correspondantes s'écrivent $f(u, x)$ et $f(u, x_0)$. Calculons les moyennes de ces deux distributions selon la formule (3) avec $a = 0$ et $k = 1$, indices qu'on négligera, comme convenu. Il vient pour la moyenne de la distribution concernant l'âge x :

$$M(x) = \int_0^{\infty} u f(u, x) du. \quad (10)$$

Le revenu moyen relatif à l'âge de référence x_0 se calculant d'une manière analogue, nous pouvons rapporter les moyennes $M(x)$ à $M(x_0)$ et définir ainsi la fonction:

$$s(x) = \frac{M(x)}{M(x_0)}. \quad (11)$$

Cette fonction donne précisément *l'échelle des revenus moyens*. Pour le moment, elle n'est définie que dans le plan $t = 0$; mais nous verrons par la suite que grâce aux hypothèses introduites au § 3 ci-après, elle ne dépend pas de t , de sorte qu'il s'agit bien d'une fonction de x seul, proportionnelle à $M(x)$. En vertu des hypothèses admises la fonction $s(x)$ est continue et dérivable. Notons encore la relation évidente:

$$s(x_0) = 1. \quad (11')$$

Si les données empiriques concernant les fonctions de fréquence $f(u, x)$ manquent, on dispose, en revanche, de quelques observations quant aux moyennes des revenus par âge et, partant, quant à $s(x)$. L'étude de la variation de $s(x)$ donne comme image classique de cette fonction

une courbe croissante de 1 à s' entre x_0 et x' , ce dernier âge étant compris entre x_0 et x_1 , puis légèrement décroissante de s' à s_1 entre x' et x_1 ; sa concavité est tournée vers le bas. Il est cependant des cas où $s(x)$ croît d'une manière monotone sur tout l'intervalle x_0, x_1 , le maximum $s' = s_1$ n'étant alors atteint qu'au point x_1 .

En partant de $s(x)$, on pourrait définir d'autres grandeurs, en particulier *l'intensité* $\sigma(x)$ de la variation des revenus moyens; elle serait alors calculée selon une méthode analogue à celle utilisée pour déterminer le taux instantané de mortalité à partir de la fonction de survie $l(x)$. On peut établir ainsi un parallélisme analytique presque complet entre la fonction économétrique $s(x)$ et la fonction biométrique $l(x)$.

6. Si la fonction $s(x)$ est monotone entre x_0 et x_1 , il existe *une fonction inverse à détermination univoque*:

$$x = \psi(s). \quad (12)$$

Cependant, si $s(x)$ présente en $x' < x_1$ un maximum s' , la fonction inverse $\psi(s)$ est à *détermination double*. Pour fixer les idées admettons que l'on ait $1 < s_1 < s'$. Notons la première détermination de $\psi(s)$ qui correspond à la branche croissante de $s(x)$ par $\psi_1(s)$, et la seconde qui correspond à la branche décroissante par $\psi_2(s)$. De la sorte, nous pouvons écrire les relations suivantes, dont nous nous servirons à plusieurs reprises par la suite:

$$\left. \begin{aligned} \psi_1(1) &= x_0 \\ \psi_1(s') &= \psi_2(s') = x' \\ \psi_2(s_1) &= x_1; \end{aligned} \right\} \quad (12')$$

relations qui se réduiront, dans le cas particulier monotone où $x' = x_1$, à:

$$\left. \begin{aligned} \psi(1) &= x_0 \\ \psi(s_1) &= x_1. \end{aligned} \right\} \quad (12'')$$

7. Nous avons fait remarquer plus haut que la notion de l'échelle $s(x)$ a déjà été utilisée par des actuaires anglais, en particulier par King [24]. Ce sont des problèmes actuariels ayant trait aux caisses de pensions qui ont incité cet auteur à introduire «*the salary scale*». Le mémoire de King se borne à définir une échelle de salaire tirée des statistiques d'une caisse de pensions pour une année donnée, en comparant les moyennes de salaires suivant l'âge des membres de la caisse.

Il suppose par la suite que les salaires moyens relatifs à un âge x se développent dans le temps suivant l'échelle observée; il passe ainsi d'une notion «verticale» à une notion «diagonale» (cf. n° 1). Comme King admet une proportionnalité rigoureuse entre les prestations de la caisse et les salaires, il lui suffit, comme nous le montrerons au chapitre III, de travailler avec les moyennes sans faire appel à des distributions de salaires autour de ces moyennes. L'essentiel de l'étude de King est d'ailleurs consacré à l'introduction du «salary scale» dans les formules actuarielles classiques. L'auteur de ce mémoire relève qu'un développement très rapide d'une entreprise peut faire changer la composition du personnel de telle sorte que les salaires moyens relatifs à un âge donné s'en trouvent affectés, généralement dans le sens d'une baisse, du fait que le recrutement du personnel nouveau s'effectue avant tout pour les occupations subalternes. C'est donc un renouvellement non homogène qui trouble dans ce cas la structure des salaires de l'entreprise. Dans le cadre de la sécurité sociale, il est rare d'observer des phénomènes semblables. Chez King, la fonction $s(x)$ se réfère essentiellement à une *notion collective*, de même que notre définition initiale selon (11) établie à l'aide des revenus moyens, et non individuels, d'un groupe d'âge x .

Ces applications aux caisses de pensions mises à part, on peut dire que la notion de l'échelle $s(x)$ a été quelque peu négligée par les actuaires. En Suisse, on a envisagé des cas particuliers linéaires dans les travaux [22] et [29] relatifs à l'équilibre financier de l'assurance-vieillesse et survivants (AVS), ainsi que dans l'étude ayant trait à la situation des caisses de pensions considérée dans le cadre de l'AVS [30]. On trouvera dans ces ouvrages, comme d'ailleurs dans un exposé de l'Office fédéral de l'industrie, des arts et métiers et du travail sur les ouvriers accidentés de la Caisse nationale suisse en cas d'accidents [9], des *indications numériques et graphiques*. L'allure numérique de $s(x)$ est évidemment différente suivant le sexe, l'état civil, le genre des revenus considérés (salaire, revenu mixte, etc.) et suivant les groupes économiques (ouvriers, employés, etc.). Un premier dépouillement des comptes individuels de l'AVS a permis de déceler pour le revenu du travail de la population masculine suisse les valeurs suivantes, ceci en posant $s(20) = 1$:

$$s(30) = 2,29; \quad s(40) = 2,65; \quad s(50) = 2,71; \quad s(60) = 2,43.$$

§ 3. Le calcul de la distribution des revenus dans l'espace (t, u, x)
à partir de trois fonctions fondamentales

8. Nous avons déjà montré, au n° 4, qu'aussi bien au point de vue statistique qu'au point de vue actuariel, la fonction $\varphi_t(u, x)$ définissant la répartition des revenus dans tous les plans économétriques $(u, x)_t$ n'est en général pas donnée à l'avance. Il faut la calculer à partir d'éléments plus simples. Montrons comment on peut résoudre ce problème en supposant connues *les trois fonctions fondamentales* suivantes :

- 1° $f_0(u, x_0) = g(u)$, qui détermine la distribution des revenus de l'ensemble d'âge x_0 vivant au moment $t = 0$, c'est-à-dire la distribution des revenus sur la droite $x = x_0$, parallèle à l'axe des u , du plan $t = 0$; nous supposons $g(u)$ normé sur cette droite.
- 2° $s(x)$, l'échelle des revenus moyens qui définit la loi de variation des revenus moyens avec l'âge et qui a été introduite au n° 5; on sait que $s(x_0) = 1$.
- 3° $\lambda_t(x)$, fonction à deux variables donnant, à l'instant t , la structure d'âge de la population envisagée; nous admettons qu'elle soit normée sur toute droite parallèle à l'axe des x du plan biométrique (t, x) .

On peut résoudre le problème en deux étapes successives. Dans la première, nous déduisons, à l'intérieur du plan $t = 0$, toutes les distributions situées sur des droites parallèles à l'axe des u et, dans la deuxième étape, nous calculerons, à partir de ces distributions celles se rattachant à tous les plans parallèles au plan initial $t = 0$.

9. La première étape de la résolution consiste donc à calculer les fonctions de fréquence des *distributions sur les droites parallèles à l'axe des u* , et ceci à l'intérieur du plan $t = 0$. Nous associons, pour un instant, à la droite $x = x_0$ la variable de revenus u' et à une droite parallèle quelconque concernant l'âge x la variable u . Nous admettons le changement de variable suivant :

$$u = s(x) u'; \tag{13}$$

en appliquant (13) à la fonction de fréquence $g(u')$, sa transformée s'écrit, sous sa forme normée: $\frac{1}{s(x)} g\left[\frac{u}{s(x)}\right]$. Nous supposons maintenant que cette transformée coïncide précisément avec la fonction

relative à la droite parallèle de l'âge x ; cela constitue notre *hypothèse A*, en vertu de laquelle nous pouvons donc écrire:

$$f_0(u, x) = \frac{1}{s(x)} g \left[\frac{u}{s(x)} \right]. \quad (14)$$

Relevons d'abord que *toutes les fonctions de fréquence conditionnelles $f_0(u, x)$ s'expriment en u par la même fonction g* . En effet, (14) montre que c'est toujours la même opération g qui s'applique à la variable u , quel que soit x , cette dernière variable n'ayant, pour une droite donnée, que la signification d'un paramètre. Certains auteurs, par exemple Haller [19], considèrent les fonctions de fréquence déduites de l'une d'entre elles par une substitution linéaire et homogène, dans le genre de (13), comme appartenant à des distributions «semblables au sens étroit».

Cette appellation nous mène à l'interprétation géométrique des relations (13) et (14). L'adjectif «semblable» ne paraît pas convenir ici puisqu'il pourrait laisser sous-entendre que nous sommes en présence d'une transformation par homothétie, ce qui n'est manifestement pas le cas. Dans le plan (u, f) , les relations (13) et (14) définissent simplement une *dilatation*, différente suivant les deux axes, de la courbe de fréquence $g(u)$. Nous avons donc plutôt un cas élémentaire de transformation par affinité. Notons encore que le produit (13) \times (14) donne pour tous les points u correspondant à un u' donné une aire de rectangle constante et égale à $u' g(u')$; il est ainsi facile de construire une courbe $f_0(u, x)$ quelconque à partir de la courbe $g(u')$.

Calculons encore la moyenne de la distribution définie par (14): il vient, en tenant compte de la transformation (13):

$$M(x) = \frac{1}{s(x)} \int_0^{\infty} u g \left[\frac{u}{s(x)} \right] du = s(x) \int_0^{\infty} u' g(u') du', \quad \text{de sorte que:}$$

$$M(x) = s(x) M(x_0) \quad (14')$$

ce qui prouve que *la distribution définie par (14) admet précisément $s(x)$ comme loi de variation du revenu moyen*. Ce résultat n'a d'ailleurs rien d'étonnant lorsqu'on se rapporte à l'interprétation individuelle de la transformation (13) sur laquelle nous reviendrons brièvement au n° 11 ci-après.

10. La deuxième étape conduisant à la solution de notre problème consiste à passer des distributions dans le plan $t = 0$ à celles des plans parallèles se rattachant à un instant t quelconque. Pour opérer ce passage, nous admettons que l'on puisse écrire :

$$f_t(u, x) = f_0(u, x) = f(u, x), \quad (15)$$

ce qui revient à introduire *l'hypothèse B*, selon laquelle toutes les distributions de revenus sur les droites concernant un âge x donné restent les mêmes, quel que soit l'instant t envisagé. Leurs fonctions de fréquence, indépendantes de t , se calculent donc directement à partir des deux fonctions fondamentales g et s , selon la formule (14). En passant du plan $t = 0$ à un plan parallèle, nous effectuons donc une simple *translation* des courbes de fréquence; ces dernières sont ainsi normées sur toutes les droites parallèles à l'axe des u .

Nous pouvons immédiatement trouver le lien entre les fonctions $\varphi_0(u, x)$ et $\varphi_t(u, x)$ concernant *les distributions dans les plans parallèles au plan $t = 0$* ; en vertu de (7) et (8) notés avec l'indice t , (15) permet d'écrire :

$$\varphi_t(u, x) = \frac{\lambda_t(x)}{\lambda_0(x)} \varphi_0(u, x). \quad (16)$$

Reste à montrer que $\varphi_t(u, x)$ se calcule directement à partir des trois fonctions fondamentales. De (16) on tire d'abord :

$$\varphi_t(u, x) = \lambda_t(x) f(u, x); \quad (17)$$

en substituant à $f(u, x)$ son expression selon (14), nous trouvons finalement :

$$\varphi_t(u, x) = \frac{\lambda_t(x)}{s(x)} g \left[\frac{u}{s(x)} \right]. \quad (17')$$

Ainsi, le calcul de $\varphi_t(u, x)$ à partir des trois fonctions fondamentales $\lambda_t(x)$, $s(x)$ et $g(u')$ s'effectue d'une manière simple pour tout point de l'espace bio-économétrique. En intégrant (17) en u et en x sur tout le domaine de définition, on voit immédiatement que la fonction φ_t est normée dans le plan $(u, x)_t$, λ_t et f étant normés sur des droites, comme nous l'avons admis au n° 8.

Connaissant dès lors la distribution des revenus dans tout l'espace, nous sommes maintenant à même d'en déduire *les distributions globales*

relatives aux personnes vivant à l'instant t . Il suffit pour cela d'appliquer la formule de projection (9) écrite avec l'indice t ; ainsi, il vient, en tenant compte de (17):

$$\bar{f}_t(u) = \int_{x_0}^{x_1} \varphi_t(u, x) dx = \int_{x_0}^{x_1} \lambda_t(x) f(u, x) dx, \quad (18)$$

où l'on peut, le cas échéant, remplacer $f(u, x)$ par son expression (14).

11. *Résumant les résultats acquis grâce à l'introduction de nos deux hypothèses A (dilatation) et B (translation), nous pouvons affirmer (cf. graphique n° 1):*

- 1° La fonction $\varphi_t(u, x)$, donnant la distribution des revenus dans un plan parallèle quelconque au plan $t = 0$, se calcule en chaque point de l'espace par (17') à l'aide des trois fonctions fondamentales $\lambda_t(x)$, $s(x)$ et $g(u')$. Elle ne dépend de t que par l'intermédiaire de la fonction biométrique $\lambda_t(x)$ seule.
- 2° La fonction $f_t(u, x)$, déterminant la distribution des revenus sur toutes les droites parallèles à l'axe des u , se ramène à une fonction de fréquence $f(u, x)$, indépendante de t , valable pour tous les ensembles d'âge x ; elle se calcule par (14) à l'aide des deux fonctions fondamentales $s(x)$ et $g(u')$ et s'exprime, par rapport à la variable u , par la même fonction g , ceci quel que soit x . Les courbes de fréquence pour un âge x s'obtiennent par dilatation à deux dimensions à partir de la courbe de fréquence de l'âge x_0 .
- 3° La fonction $\bar{f}_t(u)$, définissant la distribution globale des revenus de la population vivant à l'instant t , se calcule selon (18), et ne dépend de t que par l'intermédiaire de la fonction biométrique $\lambda_t(x)$ seule.

Par la formule (14'), nous avons mis en évidence que la distribution définie par $f_0(u, x)$ admettait $s(x)$ comme loi de variation des revenus moyens avec l'âge. Du fait de la relation (15), qui n'est autre que l'expression de l'hypothèse B, nous pouvons maintenant affirmer que toutes les distributions de revenus sur des droites parallèles à l'axe des u admettent, quel que soit t , $s(x)$ comme échelle des revenus moyens. L'échelle $s(x)$ est donc bien une fonction de x seul. En particulier, ce qui précède est aussi vrai pour des effectifs liés biométriquement,

c'est-à-dire qui évoluent, dans le plan (t, x) sur une diagonale (cf. n° 1). Considérons, pour fixer les idées, un effectif $L_0(x_0)$ qui deviendra, t années plus tard, l'effectif $L_t(x_0 + t)$, les deux se trouvant sur la diagonale $x = x_0 + t$; c'est d'ailleurs dans ces conditions qu'a été établi le graphique n° 1 de l'annexe II. Le premier de ces effectifs se distribuera en u selon $f(u, x_0)$ et le second selon $f(u, x)$. En reprenant le calcul aboutissant à (14'), on aura encore: $M(x) = s(x) M(x_0)$, de sorte que *ces effectifs liés biométriquement admettent aussi la loi $s(x)$* . De cette manière, la transformation (13) peut être interprétée à titre individuel; elle signifie alors que le revenu de chaque individu d'âge x_0 se dilate au cours de sa carrière selon (13). Ainsi, la notion $s(x)$, définie au § 2 comme notion «verticale» et valable pour $t = 0$, se transforme en notion «diagonale», d'une manière d'ailleurs analogue à ce qui a été fait par King, comme nous l'avons déjà relevé au n° 7.

Reste à savoir si nos *hypothèses A et B sont plausibles*. Nous venons de voir que l'hypothèse *A* revient à admettre une dilatation $s(x)$ uniforme pour tous les revenus au cours des carrières individuelles. L'hypothèse *B* n'est pas moins plausible; en effet, lorsqu'on admet une unité monétaire intrinsèque à pouvoir d'achat constant, ce qui est habituellement le cas dans une même série de calculs actuariels, le niveau moyen des salaires pour un âge donné ne subit, en règle générale, que des fluctuations relativement faibles au cours des années. Les variations des revenus moyens dues à un changement brusque dans le renouvellement, tel qu'il a été signalé par King (cf. n° 7) dans le cas particulier d'une entreprise industrielle, jouent, pour l'ensemble de la population, un rôle beaucoup moins marqué. Il serait d'ailleurs bien difficile de remplacer l'hypothèse *B* par une autre, plus plausible encore; il faudrait alors faire œuvre de prophète et prédire l'évolution économique elle-même.

La formule (17) montre que les *calculs démographiques* de l'actuariat peuvent être effectués indépendamment des *calculs économétriques*. Ainsi, on est à même de calculer l'élément démographique ou biométrique $\lambda_t(x)$ selon toutes les finesses de l'actuariat classique en mettant en jeu les intensités de sortie et de renouvellement. Lorsque les effectifs $L_t(x) = L_t \lambda_t(x)$ sont ainsi déterminés, on leur appliquera les fonctions économétriques $f(u, x)$. On voit par la même formule (17) de quelle manière on tend vers *l'état stationnaire*. En vertu de l'hypothèse *B*, la composante économétrique relative à un âge x est

stationnaire dès l'instant $t = 0$. En revanche, la composante biométrique relative au même âge x ne tend, en général, vers l'état stationnaire, caractérisé par la structure d'âge de l'ordre de survie $l(x)$, que lentement et par oscillations successives. De ce fait, l'état stationnaire économétrique concernant l'ensemble de la population, et non plus pour un seul âge x , n'est réalisé qu'en même temps que l'état stationnaire biométrique. Ainsi, par exemple, la somme des revenus de la population qui est calculée selon (4) à l'aide de $\bar{f}_t(u)$, déterminé lui-même par (18), variera avec t en même temps que $\lambda_t(x)$.

§ 4. Le calcul de la distribution globale des revenus à l'époque t

12. Nous venons de voir que les distributions de revenus dans deux plans économétriques $(u, x)_t$ et $(u, x)_{t'}$, ne diffèrent entre elles que par les deux fonctions biométriques $\lambda_t(x)$ et $\lambda_{t'}(x)$. Or, comme nous nous proposons de dégager l'aspect économétrique du problème, nous pouvons concentrer notre attention sur un seul de ces plans; il est en effet facile de passer ensuite, à l'aide de (16), de l'étude faite dans un des plans, à celle entreprise dans un autre. Nous supposons, cependant, que $\varphi_t(u, x)$ n'est pas donné d'emblée, mais reconstruit à l'aide des fonctions fondamentales λ_t , s et g comme le veut la méthode exposée au § précédent; φ_t se calcule donc selon les formules (17) ou (17'). Pour simplifier l'écriture, nous omettrons dans la suite l'indice t et nous noterons, en particulier, $\lambda(x)$ à la place de $\lambda_t(x)$. En plus de cette fonction, nous supposons la fonction de fréquence conditionnelle $f(u, x)$ également connue; nous la noterons directement sous cette forme, sans cependant oublier qu'elle est calculée selon la formule (14). Nous sommes donc en possession de toutes les données permettant de calculer, outre $\varphi(u, x)$, la fonction de fréquence $\bar{f}(u)$, pour laquelle nous avons déjà établi la formule (18). Le calcul de φ ne posant plus de problèmes spéciaux nous examinerons dans ce § *le calcul de la distribution marginale sur l'axe des u* , à partir des distributions conditionnelles sur les droites parallèles à cet axe; nous étudierons d'une manière plus approfondie, non seulement les fonctions de fréquence, mais encore, les autres fonctions statistiques ayant trait à la distribution globale des revenus, telles que les fonctions cumulatives, les sommes de revenus et les moments.

Jusqu'ici, nous avons toujours supposé que la variable des revenus u prenait toutes les valeurs comprises entre 0 et ∞ en admettant cependant (cf. n° 2) pour la fonction de fréquence certaines valeurs nulles. Parmi les fonctions de fréquence, il en est dont l'intervalle de définition ne commence qu'avec un point d'abscisse a positive, et d'autres dont l'intervalle est borné supérieurement par un point d'abscisse b , également positive. Cette remarque vaut en particulier pour les distributions dites tronquées [8]. Aussi, pour ne pas limiter l'étude du problème par le choix des bornes particulières 0 et ∞ , nous supposons que l'intervalle de définition de la fonction de fréquence fondamentale $f(u', x_0)$ soit donné par le segment $a_0 \leq u' \leq b_0$, où a_0 et b_0 sont des nombres positifs; f est alors normé sur ce segment, puisqu'à l'extérieur de celui-ci la masse statistique est indistinctement nulle. Il est évident que l'intervalle de définition d'une fonction de fréquence $f(u, x)$ quelconque subit alors la dilatation (13), de sorte que $f(u, x)$ sera défini dans un intervalle $a_0 s(x) \leq u \leq b_0 s(x)$. En faisant varier x , les bornes inférieures définiront dans le plan (u, x) une courbe (a) et les bornes supérieures une courbe (b), dont les équations sont respectivement:

$$\xi = a_0 s(x), \quad (19)$$

$$\zeta = b_0 s(x). \quad (19')$$

Le domaine de définition D de la fonction $\varphi(u, x)$ prend alors la forme indiquée au graphique n° 2 de l'annexe II. Il vient en particulier, si l'on se reporte au n° 5:

$$\left. \begin{aligned} a_1 &= a_0 s(x_1), & b_1 &= b_0 s(x_1), \\ a' &= a_0 s(x'), & b' &= b_0 s(x'). \end{aligned} \right\} \quad (19'')$$

Dans la suite, ce sont avant tout les relations inverses, introduites au

n° 6 avec $s = \frac{\xi}{a_0} = \frac{\zeta}{b_0}$, qui seront utilisées:

$$x = \psi_i \left(\frac{\xi}{a_0} \right) = \psi_i \left(\frac{\zeta}{b_0} \right), \quad (20)$$

où $i = 1$ ou 2 , suivant la détermination entrant en ligne de compte; ainsi l'on aura les valeurs particulières ci-après qui découlent aussi de (12'):

$$\left. \begin{aligned} x_0 &= \psi_1 \left(\frac{a_0}{a_0} \right) = \psi_1 \left(\frac{b_0}{b_0} \right) \\ x' &= \psi_{1,2} \left(\frac{a'}{a_0} \right) = \psi_{1,2} \left(\frac{b'}{b_0} \right) \\ x_1 &= \psi_2 \left(\frac{a_1}{a_0} \right) = \psi_2 \left(\frac{b_1}{b_0} \right). \end{aligned} \right\} \quad (20')$$

Rappelons que dans le cas où $s(x)$ croît d'une manière monotone, x_1 coïncide avec x' , de sorte que $\psi(s)$ est à détermination univoque.

En admettant pour $f(u, x_0)$ des distributions classiques (voir par exemple la liste des distributions établie par Haller [19]), le domaine de définition D ainsi établi se simplifie évidemment; souvent on a une des trois possibilités: 1° $a_0 = 0, b_0 = \infty$; 2° $a_0 > 0, b_0 = \infty$ et 3° $a_0 = 0, 0 < b_0 < \infty$. Ainsi, la courbe (a) peut se confondre avec l'axe des x et la courbe (b) être repoussée à l'infini. Il est cependant toujours possible de reporter à un point a_0 situé à gauche de l'origine une fonction de fréquence dont l'intervalle de définition commence à 0; il suffit pour cela de passer à la nouvelle variable u' par: $u = u' - a_0$. De la sorte, le cas 2° ci-dessus peut être aisément réalisé. Il faut souligner que le cas général $0 < a_0 < b_0 < \infty$ correspond beaucoup mieux à la nature de la variable u , puisque le revenu ne devrait pas descendre au-dessous d'un certain minimum vital a_0 , d'une part, et ne saurait dépasser un certain maximum, si grand fût-il, d'autre part. C'est là d'ailleurs un des postulats avancés par M. Fréchet [14] et sur lequel nous reviendrons au chap. II.

13. Connaissant la configuration exacte du domaine de définition de $\varphi(u, x) = \lambda(x) f(u, x)$, nous pouvons calculer explicitement la fonction de fréquence $\bar{f}(u)$ donnant la distribution globale des revenus. Nous dégagerons ainsi le lien entre les fonctions de fréquence conditionnelles $f(u, x)$ et la marginale résultante $\bar{f}(u)$. Il ressort du graphique n° 2 qu'il faut distinguer sur l'axe des u cinq intervalles qui ont été désignés par (1), (1'), (2), (3) et (3'). L'intégrale de la formule (18), prise entre les limites x_0 et x_1 doit être remplacée, selon l'intervalle envisagé, par les intégrales ci-après; on notera, par exemple, une intégrale

$\int_{x_0}^{x_1} \lambda(x) f(u, x) dx$ par le symbole $[x_0, x_1]$ et on se rappellera les relations

(20) avec ξ respectivement $\zeta = u$. En désignant par $\bar{f}_j(u)$ la fonction de fréquence sur l'intervalle (j) , on aura en admettant de plus $a' < b_0$:

$$\left. \begin{aligned} a_0 \leq u < a_1: \bar{f}_1(u) &= \left[x_0, \psi_1\left(\frac{u}{a_0}\right) \right] \\ a_1 \leq u < a': \bar{f}_{1'}(u) &= \left[x_0, \psi_1\left(\frac{u}{a_0}\right) \right] + \left[\psi_2\left(\frac{u}{a_0}\right), x_1 \right] \\ a' \leq u < b_0: \bar{f}_2(u) &= [x_0, x_1] \\ b_0 \leq u < b_1: \bar{f}_3(u) &= \left[\psi_1\left(\frac{u}{b_0}\right), x_1 \right] \\ b_1 \leq u < b': \bar{f}_{3'}(u) &= \left[\psi_1\left(\frac{u}{b_0}\right), \psi_2\left(\frac{u}{b_0}\right) \right]. \end{aligned} \right\} \quad (21)$$

Ces formules permettent donc de déterminer la fonction de fréquence marginale \bar{f} à partir des fonctions de fréquence conditionnelles, en pondérant ces dernières avec la fonction de fréquence donnant la structure d'âge; nous pouvons alors désigner $\lambda(x)$ comme *fonction pondératrice*. Lorsque $s(x)$ croît d'une manière monotone avec x , a_1 se confond avec a' et b_1 avec b' , de sorte que les cinq intervalles se réduisent à trois: (1), (2) et (3). Dans ce cas on peut calculer sur l'intervalle (1) par exemple, la fonction de fréquence de l'ensemble mixte par la formule explicite ci-après:

$$\bar{f}_1(u) = \int_{x_0}^{\psi\left(\frac{u}{a_0}\right)} \lambda(x) f(u, x) dx. \quad (22)$$

Cette formule nous servira encore au § suivant. Sur les autres intervalles, cette fonction se calculera d'ailleurs d'une manière tout à fait analogue, même dans le cas non monotone, ainsi que nous l'avons montré symboliquement par les formules (21).

14. L'analyse des cinq expressions (21) trouvées pour $\bar{f}(u)$ montre que nous sommes bien en présence de cinq fonctions différentes de u . Ainsi, la courbe de fréquence de la distribution globale des revenus se compose de cinq branches différentes. Les formules font ressortir que les ordonnées aux points de jonction coïncident. Quel est maintenant la nature du contact en ces points? En général, ce contact est de l'ordre 0, c'est-à-dire qu'il y a, comme nous venons de le faire

remarquer, simple raccord non-tangentiel entre les différentes branches. Nous pouvons cependant examiner *les conditions de raccord tangentiel entre les cinq branches de la courbe de fréquence*, c'est-à-dire les conditions de contact du 1^{er} ordre. Nous attachons une certaine importance à ce problème, vu que quelques auteurs ont essayé de trouver, dans le cas particulier des distributions de Pareto, un raccord satisfaisant entre une courbe de Pareto pour les revenus supérieurs et une autre courbe pour les revenus inférieurs. Ces essais n'ont guère été fructueux; M. Fréchet [15] a proposé, dans ce cas particulier, une formule qui réalise un simple raccord, sans cependant être tangentiel. Nous allons montrer que les conditions de raccord tangentiel sont simples et souvent réalisables dans les applications. Pour obtenir ces conditions pour les quatre points de jonction a_1 , a' , b_0 et b_1 , il suffit de calculer les dérivées premières des cinq branches à partir de leurs équations (21) et de les évaluer deux à deux aux quatre points de contact. En ce qui concerne les calculs, il convient d'appliquer les règles de dérivation d'une intégrale définie, dont aussi bien les limites que la fonction sous le signe de somme dépendent du paramètre u .

Considérons, par exemple, le contact au point $u = a_1$. D'après (21) les deux branches concernant les intervalles (1) et (1') ont respectivement les deux équations explicites que voici:

$$\bar{f}_1(u) = \int_{x_0}^{\psi_1} \lambda(x) f(u, x) dx \quad \text{et}$$

$$\bar{f}_{1'}(u) = \int_{x_0}^{\psi_1} \lambda(x) f(u, x) dx + \int_{\psi_2}^{x_1} \lambda(x) f(u, x) dx,$$

où les limites ψ_1 et ψ_2 sont fonctions du rapport $\frac{u}{a_0}$. Pour les dérivées premières il vient:

$$\bar{f}'_1(u) = \int_{x_0}^{\psi_1} \lambda(x) \frac{\partial f(u, x)}{\partial u} dx + \lambda(\psi_1) f(u, \psi_1) \frac{\psi'_1}{a_0}, \quad \text{et}$$

$$\bar{f}'_{1'}(u) = \int_{x_0}^{\psi_1} \lambda(x) \frac{\partial f(u, x)}{\partial u} dx + \lambda(\psi_1) f(u, \psi_1) \frac{\psi'_1}{a_0}$$

$$+ \int_{\psi_2}^{x_1} \lambda(x) \frac{\partial f(u, x)}{\partial u} dx - \lambda(\psi_2) f(u, \psi_2) \frac{\psi'_2}{a_0}.$$

Si l'on écrit les valeurs de ces dérivées en $u = a_1$ et si l'on égale les résultats, on voit que les deux premiers termes de \bar{f}'_1 sont identiques aux deux termes de \bar{f}'_1 ; de plus, on sait d'après (20') que $\psi_2\left(\frac{a_1}{a_0}\right) = x_1$ de sorte qu'il suffit d'annuler le produit $\lambda(x_1) f(a_1, x_1) \psi_2\left(\frac{a_1}{a_0}\right)$ pour obtenir l'égalité des dérivées premières en a_1 . En procédant d'une manière analogue pour les autres points de contact, nous obtenons dès lors les quatre conditions suivantes pour les raccords tangentiels en :

$$\left. \begin{aligned} a_1: \lambda(x_1) f(a_1, x_1) \psi_2\left(\frac{a_1}{a_0}\right) &= 0 \\ a': \lambda(x') f(a', x') \left[\psi_1\left(\frac{a'}{a_0}\right) - \psi_2\left(\frac{a'}{a_0}\right) \right] &= 0 \\ b_0: \lambda(x_0) f(b_0, x_0) \psi_1\left(\frac{b_0}{b_0}\right) &= 0 \\ b_1: \lambda(x_1) f(b_1, x_1) \psi_2\left(\frac{b_1}{b_0}\right) &= 0. \end{aligned} \right\} \quad (23)$$

Pour obtenir les raccords tangentiels, il suffit d'annuler dans les quatre conditions ci-dessus un seul des trois facteurs. On examinera dans chaque cas concret, si la jonction tangentielle peut être réalisée en un seul ou simultanément en plusieurs, voire en tous les points envisagés. En particulier, on regardera de près ce qui se passe au point a' où $\psi_1 - \psi_2$ peut amener une indétermination. Cette étude se simplifie évidemment dans le cas monotone de $s(x)$, les quatre conditions se réduisant alors à deux, c'est-à-dire à celles concernant a_1 et b_0 . Nous verrons au chapitre II quelques exemples concrets.

15. Nous avons démontré, au n° 9, que les fonctions $f(u, x)$ s'exprimaient toutes, grâce à la dilatation effectuée selon l'hypothèse A , par une même fonction en u . Il est dès lors naturel de se demander si la courbe de fréquence marginale $\bar{f}(u)$ peut également appartenir à la même famille de fonctions en u que les fonctions de fréquence conditionnelles $f(u, x)$. Nous allons montrer que tel est effectivement le cas sur

l'intervalle (2) lorsque la fonction fondamentale $g(u')$ se présente sous la forme suivante:

$$g\left(\frac{u}{s}\right) = A \left(\frac{u}{s}\right)^B. \quad (24)$$

De cette façon, il viendra pour la fonction de fréquence à deux dimensions $\varphi(u, x)$, en tenant compte de (17) et (14):

$$\varphi(u, x) = A \lambda(x) s^{-B-1}(x) u^B. \quad (25)$$

En effectuant la projection sur l'axe des u , nous obtenons dès lors pour la marginale sur l'intervalle (2):

$$\bar{f}_2(u) = A u^B \int_{x_0}^{x_1} \lambda(x) s^{-B-1}(x) dx;$$

l'intégrale à limites fixes x_0 et x_1 se réduit à une constante K , de sorte que \bar{f}_2 est de la forme:

$$\bar{f}_2(u) = C u^B, \quad (26)$$

et appartient donc à la même famille de courbes que g et, partant, que $f(u, x)$. En revanche, sur les autres intervalles ($j \neq 2$), les intégrales de (21) dépendent au moins d'une limite variable avec u ; ainsi on arrive à une expression de la forme:

$$\bar{f}_j(u) = A u^B h_j(u); \quad (26')$$

sur quatre intervalles donc, la courbe de fréquence marginale n'appartient en général plus à la même famille de courbes que les conditionnelles. Dans le cas d'un $s(x)$ monotone, il ne demeure que deux intervalles où \bar{f} ne s'exprime plus à l'aide de la même fonction que $f(u, x)$. — Nous avons insisté ici sur cette forme particulière (24) de la fonction fondamentale g , vu que nous examinerons, au chapitre II, deux distributions dont les fonctions de fréquence prennent précisément la forme dont il vient d'être question; il s'agit des distributions hyperbolique et parabolique.

16. Nous avons rappelé au n° 2, outre la fonction de fréquence, les principales fonctions statistiques caractérisant les distributions à une seule dimension. Indiquons ainsi les liens qui existent entre ces autres fonctions de la distribution marginale sur l'axe des u , d'une part, et celles des distributions conditionnelles parallèles à cet axe, d'autre part. Qu'en est-il d'abord des fonctions cumulatives?

Introduisons, conformément aux formules (2) et (2'), les définitions que voici; soient

$$F(u, x) = \int_{\xi}^u f(\eta, x) d\eta, \quad (27)$$

$$H(u, x) = \int_u^{\zeta} f(\eta, x) d\eta, \quad (27')$$

la fonction cumulative pour les distributions conditionnelles et sa complémentaire, les limites ξ et ζ étant définies selon (19) et (19'). Notons

que pour $u > b_0$: $F(u, x) = 1$, pour autant que $x < \psi_1\left(\frac{u}{b_0}\right)$ et $x > \psi_2\left(\frac{u}{b_0}\right)$. On a, d'une manière analogue pour certaines valeurs de x :

$H(u, x) = 1$ à condition que $u < a'$. — Signalons encore les lois de transformation de $F(u, x)$ et $H(u, x)$ par rapport à la substitution (13). En partant, pour l'âge x_0 , de $F(u', x_0)$ on trouvera sans plus:

$$F(u, x) = F\left[\frac{u}{s(x)}, x_0\right], \quad (28)$$

et une relation analogue pour $H(u, x)$. Les fonctions cumulatives des distributions conditionnelles se déduisent donc d'une manière très simple de la cumulative relative à l'âge x_0 . Géométriquement, dans le plan (u, F) , (28) revient à une dilatation dans une seule direction, celle de l'axe des u . En dérivant (28) par rapport à u , on retrouve évidemment (14).

Pour les fonctions cumulatives de la distribution marginale, il vient:

$$\bar{F}(u) = \int_{a_0}^u \bar{f}(\eta) d\eta, \quad (29)$$

$$\bar{H}(u) = \int_u^{b'} \bar{f}(\eta) d\eta. \quad (29')$$

Si nous remplaçons dans (29) $\bar{f}(\eta)$ par ses expressions (21), et si nous indiquons par $x(u)$ le chemin d'intégration $\parallel x$, de longueur variable avec le u considéré, la cumulative s'obtient par l'intégrale double que voici:

$$\bar{F}(u) = \int_{a_0}^u d\eta \int_{x(u)} \lambda(x) f(\eta, x) dx = \int_{x_0}^{x_1} \lambda(x) dx \int_{\xi}^{\zeta, u} f(\eta, x) d\eta.$$

Pour comprendre l'alternative ζ ou u de la limite supérieure de l'intégrale en η reportons-nous à la configuration exacte du domaine d'intégration reproduit au graphique n° 2. Choisissons un point u dans l'intervalle (3') et menons par cette abscisse une parallèle à l'axe des x . Cette parallèle est décomposée par la courbe (b) en segments désignés par e (extérieurs) et i (intérieurs). Pour les âges x relatifs aux segments e il suffit de prendre la limite ζ vu qu'il n'y a pas de masse statistique à droite de la courbe (b); ainsi les intégrales correspondantes égalent l'unité, les fonctions f étant normées. En revanche, pour les x concernant le segment i la limite sera donnée par le u considéré; les intégrales en η se réduisent donc grâce à (27) à $F(u, x)$. Si l'on décompose l'intégrale en x selon les segments i et e , il vient:

$$\bar{F}(u) = \int_{i(u)} \lambda(x) F(u, x) dx + \int_{e(u)} \lambda(x) dx. \quad (30)$$

Les segments i et e sont évidemment fonctions de u . Si $u < b_0$, il n'y a que des segments i de sorte que le deuxième terme de (30) disparaît. Ainsi, et vu qu'à gauche de (a): $F(u, x) = 0$, l'intégrale le long de $i(u)$ est limitée, dans les intervalles (1), (1') et (2), de la même manière que dans les trois premières formules (21). En revanche, dans les intervalles (3) et (3'), il vient s'ajouter aux intégrales le long de $i(u)$, limitées comme dans les deux dernières formules (21), une, respectivement deux intégrales le long de $e(u)$. — Un raisonnement analogue permet d'écrire la relation suivante pour la cumulative complémentaire, en remarquant toutefois que c'est la courbe (a) qui détermine cette fois les segments i et e :

$$\bar{H}(u) = \int_{i(u)} \lambda(x) H(u, x) dx + \int_{e(u)} \lambda(x) dx, \quad (30')$$

où, cependant, le deuxième terme disparaît dès que $u > a'$. — Dans le cas banal où le domaine D est rectangulaire (30) et (30') se réduisent à la première intégrale, limitée par x_0 et x_1 . — Les fonctions \bar{F} et \bar{H} auront ainsi dans le cas général, suivant l'intervalle envisagé, une autre expression et partant donneront lieu à cinq branches de courbes différentes. On vérifie aisément, pour chacun des cinq intervalles (j), les relations $\bar{F}'_j(u) = \bar{f}_j(u)$ et $\bar{F}''_j(u) = \bar{f}'_j(u)$; ainsi les contacts aux points de jonction des cinq branches de la cumulative marginale ne sont pas seulement tangentiels, c'est-à-dire d'ordre 1, mais encore d'ordre 2, pourvu que les conditions (23) soient remplies.

17. Nous pouvons raisonner d'une manière semblable sur *les sommes réduites des revenus*, telles qu'elles ont été définies par les formules (4) et (4'). On aboutira ainsi au lien entre $\Phi(u, x)$, relatif aux distributions conditionnelles, et $\bar{\Phi}(u)$, relatif à la marginale, ainsi qu'à celui entre $\Theta(u, x)$ et $\bar{\Theta}(u)$:

$$\bar{\Phi}(u) = \int_{i(u)} \lambda(x) \Phi(u, x) dx + \int_{e(u)} \lambda(x) M(x) dx, \quad (31)$$

$$\bar{\Theta}(u) = \int_{i(u)} \lambda(x) \Theta(u, x) dx + \int_{e(u)} \lambda(x) M(x) dx, \quad (31')$$

où $M(x)$ représente la moyenne d'une distribution conditionnelle parallèle à l'axe des u . — Indiquons encore la loi de transformation des fonctions $\Phi(u, x)$ par rapport à la substitution (13):

$$\Phi(u, x) = s(x) \Phi \left[\frac{u}{s(x)}, x_0 \right]. \quad (32)$$

18. Il reste à étudier *les relations entre les moments* de la distribution sur l'axe des u et ceux des distributions sur les droites parallèles. Nous nous bornerons à exposer le cas des moments par rapport à l'abscisse $u = 0$, les moments par rapport à une abscisse quelconque étant alors obtenus par les formules (3') et (3''). Soient $M_k(x)$ les moments d'ordre k pour un âge x quelconque; nous aurons en appliquant la substitution (13):

$$M_k(x) = \int_{\xi}^{\zeta} u^k f(u, x) du = s^k(x) \int_{a_0}^{b_0} u'^k g(u') du',$$

de sorte que:

$$M_k(x) = s^k(x) M_k(x_0); \quad (33)$$

cette relation donne la loi de transformation des moments à partir de ceux correspondant à l'âge de référence x_0 . La formule (14') n'est qu'un cas particulier de cette relation générale. — Désignant maintenant par \bar{M}_k les moments de la distribution marginale sur l'axe des u , nous obtenons:

$$\bar{M}_k = \int_{a_0}^{b'} u^k \bar{f}(u) du = \int_{a_0}^{b'} u^k du \int_{x(u)} \lambda(x) f(u, x) dx = \int_{x_0}^{x_1} \lambda(x) dx \int_{\xi}^{\zeta} u^k f(u, x) du;$$

ainsi, il vient:

$$\bar{M}_k = \int_{x_0}^{x_1} \lambda(x) M_k(x) dx = M_k(x_0) \int_{x_0}^{x_1} \lambda(x) s^k(x) dx. \quad (34)$$

De la sorte, les moments de la distribution marginale apparaissent comme les moyennes pondérées des moments des distributions conditionnelles. Comme nous l'avons vu, cette propriété est d'ailleurs commune aux autres fonctions statistiques \bar{f} , \bar{F} et $\bar{\Phi}$ définies à l'intérieur du plan $(u, x)_t$, ce qui justifie bien l'appellation de fonction pondératrice pour $\lambda(x)$.

§ 5. Les équations intégrales d'un problème connexe

19. Posons-nous la question s'il est possible de calculer la fonction de fréquence à deux dimensions $\varphi(u, x)$ à partir d'un couple quelconque de fonctions choisies parmi $f(u, x)$, $\lambda(x)$ et $\bar{f}(u)$. Seul le premier des *trois problèmes possibles* a été envisagé jusqu'ici, c'est-à-dire la détermination de φ à l'aide des conditionnelles $f(u, x)$ et de la marginale $\lambda(x)$; la solution est alors donnée par (17). La deuxième possibilité où l'on se donne les deux marginales $\lambda(x)$ et $\bar{f}(u)$ mène, en général, à un problème indéterminé; on sait, en effet, que celui-ci n'a une solution déterminée que dans le cas de distribution indépendante des variables x et u et où $\varphi(u, x)$ est par définition le produit des deux marginales λ et \bar{f} . Or ceci est en général exclu dans notre méthode, vu la dépendance des distributions postulées par l'hypothèse A (cf. n° 9); on vérifie aisément que pour réaliser l'indépendance il faut $s(x) = 1$. Reste le troisième cas où l'on se donne les deux fonctions $f(u, x)$ et $\bar{f}(u)$. Comme le montre (17), il suffirait de déterminer d'abord $\lambda(x)$ pour connaître ensuite $\varphi(u, x)$. Nous allons montrer que $\lambda(x)$ peut précisément être calculé à partir de $f(u, x)$ et $\bar{f}(u)$ à l'aide d'une équation intégrale.

N'oublions cependant pas qu'il s'agit là d'un problème d'intérêt plutôt théorique. En effet, on peut admettre que dans les applications la structure d'âge $\lambda(x)$, donc l'élément démographique, est toujours donnée avant les distributions des revenus \bar{f} et surtout $f(u, x)$. Au point de vue théorique, en revanche, on peut se donner pour \bar{f} une certaine fonction en u (par exemple une semi-normale) et pour $f(u, x)$ une autre (par exemple une hyperbolique), cette dernière dépendant en même temps d'un paramètre x . On recherchera alors la fonction $\lambda(x)$ qui amène la *compatibilité statistique des deux fonctions connues* \bar{f} et $f(u, x)$. Vu la nature théorique de la question, nous nous bornerons à envisager quelques cas simples et, en particulier, nous admettrons

pour $s(x)$ une fonction monotone. Ainsi, le domaine de définition de $\varphi(u, x)$ se simplifie, puisque a_1 se confond avec a' et b_1 avec b' .

20. Ainsi que nous l'avons dit au n° 12, le domaine D dépend, non seulement de $s(x)$, mais aussi de l'intervalle de définition de la fonction fondamentale $f(u, x_0)$. Envisageons d'abord le cas où l'on a : $0 \leq u < \infty$. Les relations (21) se réduisent alors à :

$$\bar{f}(u) = \int_{x_0}^{x_1} \lambda(x) f(u, x) dx. \quad (35)$$

En se donnant $\bar{f}(u)$ ainsi que $f(u, x)$, qui sera pris comme noyau, la pondératrice $\lambda(x)$ sera la fonction inconnue de l'équation à limites fixes de première espèce (type de Fredholm). Ce genre d'équations est, parmi les types classiques, le plus difficile à résoudre. On trouvera quelques indications quant aux méthodes de résolution par exemple chez Goursat [18] ou chez Paul Lévy [26]. Il faut évidemment écarter les solutions négatives sur $x_0 \leq x \leq x_1$, celles-ci signifiant l'incompatibilité statistique des fonctions $\bar{f}(u)$ et $f(u, x)$. Il n'est donc pas certain que toutes les solutions admises au point de vue analytique aient un sens statistique. Notons encore que si l'on s'impose pour $\bar{f}(u)$ ainsi que pour $f(u, x)$ la forme normée, la solution $\lambda(x)$ sera normée également.

21. Considérons maintenant le cas le plus fréquemment rencontré dans les applications où l'on a pour la fonction $f(u, x_0)$: $a_0 \leq u < \infty$. Pour l'intervalle (1) nous sommes alors en présence de l'équation (22) et pour l'intervalle (2) c'est la formule (35) qu'il sied d'envisager.

Opérons sur (22) le changement de variable : $x = \psi\left(\frac{\xi}{a_0}\right)$. Les limites d'intégration en ξ seront alors, pour $x = x_0$: $\xi = a_0$ et pour $x = \psi\left(\frac{u}{a_0}\right)$: $\xi = u$. L'on obtient ainsi, en posant au préalable :

$${}^* \lambda(\xi) = \lambda \left[\psi \left(\frac{\xi}{a_0} \right) \right] \psi' \left(\frac{\xi}{a_0} \right) \frac{1}{a_0} \quad \text{et} \quad {}^* f(u, \xi) = f \left[u, \psi \left(\frac{\xi}{a_0} \right) \right]$$

la relation que voici :

$$\bar{f}_1(u) = \int_{a_0}^u {}^* \lambda(\xi) {}^* f(u, \xi) d\xi. \quad (36)$$

L'équation (22) a, de la sorte, pu être ramenée au type classique d'équation de Volterra de première espèce, dont la solution pourra être trouvée en la transformant en une équation de deuxième espèce, relativement facile à résoudre par la méthode du noyau résolvant. Si l'on se donne pour $\bar{f}_1(u)$ et $\bar{f}_2(u)$ la même courbe d'ensemble, l'équation (35) devient le cas limite de (22) et se trouve résolue du même coup. Autrement il faut calculer \bar{f}_2 à l'aide des λ obtenus à partir de \bar{f}_1 .

Un cas élémentaire de solution se présente lorsque le noyau $f(u, x)$ est dégénéré. Ceci arrive en particulier si l'on admet pour la fonction fondamentale g la forme (24), puisqu'on a alors :

$$f(u, x) = A u^B s^{-B-1}(x). \quad (37)$$

Substituons cette expression dans (22) et effectuons le changement de variable $x = \psi\left(\frac{\xi}{a_0}\right)$; l'on obtiendra ainsi l'équation de Volterra à noyau dégénéré :

$$\bar{f}_1(u) = A a_0^{B+1} u^B \int_{a_0}^u \lambda^*(\xi) \xi^{-B-1} d\xi. \quad (37')$$

En multipliant des deux côtés par u^{-B} et en dérivant par rapport à u , la solution s'exprime comme suit :

$$\lambda^*(u) = \frac{u \bar{f}_1'(u) - B \bar{f}_1(u)}{A a_0^{B+1}}, \quad (37'')$$

où l'on passera de u à x par le changement de variable (19). Si le numérateur est positif, les fonctions \bar{f}_1 et $f(u, x)$ sont compatibles au point de vue statistique. — Lorsque (24) respectivement (37) a lieu, la deuxième branche $\bar{f}_2(u)$ doit être une même fonction en u que $f(u, x)$ et diffère donc, en général, de $\bar{f}_1(u)$. La solution $\lambda(x)$ trouvée par (37'') doit alors servir à calculer $\bar{f}_2(u)$ à l'aide de (35), puisqu'alors on ne saurait se donner $\bar{f}_2(u)$ arbitrairement.

Chapitre II

Formules de répartition des revenus

§ 1. Postulats, méthodes déductives et vue d'ensemble

22. L'étude mathématique de la théorie de la répartition des revenus date d'un demi-siècle environ. A l'origine de cette nouvelle orientation des recherches, on trouve le nom de Vilfredo Pareto [31]. En effet, ce dernier a proposé une formule analytique simple pour exprimer la fonction cumulative complémentaire de la distribution des revenus. Appliquée à des données d'observation nombreuses provenant de plusieurs pays, cette formule s'avéra excellente pour interpréter la répartition des revenus, pour le moins celle des revenus moyens et élevés. Depuis lors, de nombreux auteurs ont essayé d'améliorer cette expression analytique, dans le sens d'une adaptation plus parfaite aux observations statistiques, surtout en ce qui concerne les petits revenus pour lesquels la formule de Pareto ne peut guère s'appliquer. Ce n'est que tout récemment que ces recherches ont été orientées dans une direction nouvelle. C'est à Maurice Fréchet [14] que revient le mérite d'avoir donné un fondement probabiliste à cette théorie, esquissée très brièvement ci-après au n° 23. *Les recherches effectuées dans les deux directions* dont il vient d'être question n'ont cependant pour objet que la distribution globale des revenus; en revanche, nous, nous plaçons dans les ensembles d'âge x les formules analytiques donnant la répartition des revenus, ainsi que le veut notre théorie mathématique exposée au chap. I. Nous nous proposons donc, dans la suite, d'étudier quelques formules concrètes que nous adopterons ensuite pour l'interprétation analytique des distributions conditionnelles; à l'aide de ces dernières, nous déterminerons la répartition globale des revenus. Il est alors évident que l'allure de celle-ci doit obéir aux exigences que les observations ont permis de découvrir jusqu'ici.

C'est encore Maurice Fréchet [l. c.] qui a dégagé le plus nettement ces exigences; nous pouvons les considérer comme de véritables *postulats* auxquels doit obéir la distribution globale des revenus. Résumons-les de la manière suivante:

- 1° Le champ de variation de la variable u doit satisfaire à l'inégalité $a \leq u \leq b$, a étant le revenu minimum vital et b le plus grand revenu individuel observé; du fait que ce dernier est généralement inconnu, on suppose souvent $b \equiv \infty$.
- 2° La fonction de fréquence $f(u)$ doit s'annuler en a et en b .
- 3° La distribution des revenus est unimodale; en d'autres termes, $f(u)$ n'a qu'un seul maximum.

Le premier de ces postulats résulte du fait incontestable que, d'une part, «un individu ne peut vivre de rien» [14] et que, d'autre part, les revenus d'une personne ne sauraient dépasser un montant fini, choisi suffisamment grand. Le second postulat, moins évident que le précédent, découle des observations statistiques, de même que le troisième. — Ces postulats ne s'appliquent qu'à la distribution globale des revenus. Les composantes pour les âges x peuvent fort bien ne pas obéir aux points 2° et 3°, mais l'intervention de la pondératrice $\lambda(x)$ peut conduire à une distribution globale satisfaisant ces deux conditions.

23. De nombreux auteurs ont essayé d'expliquer, par des *méthodes déductives*, la théorie de la répartition des revenus. On peut classer ces explications déductives en quatre groupes:

- 1° Explication probabiliste, chez Fréchet [14].
- 2° Explication à l'aide des particularités ou facultés personnelles, chez Fréchet [l. c.], Rhodes [34] et Ammon [2].
- 3° Explication par l'hypothèse du tamisage [20].
- 4° Explication par l'hypothèse de l'effet proportionnel, chez Gibrat [16].

Essayons de résumer l'essentiel de ces méthodes. La seule *explication probabiliste* que nous ayons rencontrée se trouve chez Fréchet [l. c.]; nous avons ici en vue celle que cet auteur déduit des jeux de hasard modèles. Il envisage le tirage de noms d'une urne contenant les noms de L personnes. En remettant chaque fois le nom tiré dans l'urne, on sait que la probabilité p_m d'extraire le nom m fois au cours de n tirages est donnée par le terme $\binom{n}{m} \left(\frac{1}{L}\right)^m \left(1 - \frac{1}{L}\right)^{n-m}$ de la distribution binomiale. En prenant n suffisamment grand, la distribution

binomiale selon m tendra vers la distribution normale $N(z;0,1)$, de moyenne 0, de précision 1 et où $z = \frac{mL - n}{\sqrt{2n(L-1)}}$. Le revenu u attribué à une personne sera alors une fonction $u(m)$ du nombre m de tirages amenant le nom de la personne. Si le revenu u est proportionnel au nombre m des tirages favorables au nom considéré, c'est-à-dire si $u(m) = cm$, la distribution selon u est elle-même normale; cette propriété subsiste même si $u(m) = cm + k$. Or, on constate que la distribution normale ne convient que très approximativement au problème de la répartition des revenus. Ceci n'a rien d'étonnant puisqu'une personne dont les moyens sont déjà importants pourra, comparée à une personne ne disposant que de faibles ressources, accroître plus facilement ses revenus; il est alors naturel de supposer que le revenu u augmente sensiblement plus vite que le nombre m des circonstances favorables; un changement de variable linéaire $u = cm + k$ ne saurait donc convenir. Pour cette raison, Fréchet admet une croissance en progression arithmétique pour m , tandis que u augmentera selon une progression géométrique, de sorte que l'on appliquera à la distribution N le changement de variable suivant $u - u_0 = e^{Am+B}$. On trouvera ainsi la distribution logarithmique normale que nous envisagerons au § 3. En choisissant convenablement d'autres fonctions $u(m)$, on peut retrouver les différentes formules proposées pour interpréter la répartition des revenus, en particulier la loi de Pareto. Ainsi, contrairement à ce qu'affirme Pareto lui-même [31], toutes les distributions de revenus peuvent être considérées comme un des effets du hasard.

Dans le même travail, Fréchet propose une explication «plus rationnelle» reposant sur *les particularités ou facultés personnelles*. Pour ceci, il envisage alternativement deux hypothèses, soit une hypothèse A , selon laquelle «le revenu de chacun est parfaitement déterminé quand on connaît les particularités propres à l'individu considéré», et une hypothèse B admettant «que la connaissance des particularités relatives à cet individu ne déterminent que les probabilités qu'a cet individu de dépasser des revenus respectifs donnés». Ces hypothèses, soumises au calcul, conduisent à des formules très générales, difficiles à appliquer, puisqu'il est rare de pouvoir obtenir, simplement par des observations statistiques, les éléments dont elles se composent. Fréchet fournit cependant un exemple «synthétique» d'application, dont les

résultats vérifient approximativement la loi de Pareto. On trouve déjà des explications de la même nature, cependant sans calculs précis, dans un travail ancien d'Ammon [2] et, plus récemment, chez Rhodes [34]. Ce dernier, en particulier, fait remarquer que ce sont certains caractères essentiels d'une personne qui la situent à une abscisse déterminée de la distribution des revenus. Notre théorie, par laquelle nous associons à un groupe de caractères (âge, sexe, état civil) une distribution propre, ne fait qu'appliquer l'affirmation de Rhodes en deux étapes: la première consiste à placer l'individu à l'intérieur d'une distribution répondant à ses caractères fondamentaux (par exemple: une jeune femme célibataire n'a généralement pas les mêmes possibilités économiques qu'un homme marié dans la force de l'âge), la seconde, à lui assigner, dans le cadre de cette première distribution, une abscisse conforme à ses aptitudes mentales ou physiques particulières.

L'explication reposant sur *l'hypothèse du tamisage* [20] suppose que les personnes disposant d'un revenu $\geq u_i$ sont «tamisées» et que seules celles qui traversent le tamis accèdent au revenu u_{i+1} . Cette méthode comporte des hypothèses sur la finesse des tamisages successifs. Enfin, l'explication par *l'hypothèse de l'effet proportionnel* de Gibrat [16] admet que l'accroissement, du , du revenu est proportionnel à u , c'est-à-dire que $du = u dz$; elle peut être ramenée visiblement à l'hypothèse de Fréchet selon laquelle la progression en m est arithmétique, tandis que celle en u est géométrique. Si cette dernière hypothèse est, dans ses grandes lignes, conforme à l'explication probabiliste de Fréchet, celle du tamisage semble pouvoir se ramener à l'explication fondée sur les facultés personnelles, ces dernières étant déterminantes pour permettre le passage à travers les tamis.

24. Si les distributions classiques entrant en ligne de compte pour représenter la répartition globale des revenus sont déjà nombreuses, nous en aurons encore davantage à disposition pour rendre l'image de la répartition propre aux groupes d'âge x , du fait qu'on peut imposer à ces dernières moins de conditions restrictives, quant à leur forme. Si l'on veut étudier toutes *les distributions pouvant être prises en considération* pour les problèmes envisagés ici, il faudra analyser, par exemple, les courbes de Pearson, auxquelles Odhnoff [28] a consacré un travail récent, ou encore le système déduit par Risser [35] d'une équation différentielle, plus simple que celle de Pearson.

De même, on pourra consulter la liste établie par Haller [19] comprenant 70 distributions, pour la plupart déjà connues. Il ne saurait être question de passer en revue, dans le présent travail, toutes ces distributions qui, d'ailleurs, sont souvent peu maniables tant au point de vue analytique que numérique, ni même de reprendre toutes celles qui ont été proposées par différents auteurs pour décrire d'une manière plus ou moins appropriée la répartition globale des revenus. Citons ici, en particulier, les noms de March, de Vinci, d'Amoroso, de Benini [20] et de Krämer [25]. Là encore, Maurice Fréchet a résumé dans un mémoire, antérieur à celui déjà cité [15], les principales formules de répartition des revenus. Il y étudie, avant tout, les formules de Pareto et de la distribution logarithmique normale, désignée par le nom de ceux qui l'ont introduite ou améliorée: McAlister, Edgeworth, Kapteyn et Gibrat. En outre, Fréchet y propose, pour les petits revenus, de corriger la distribution de Pareto, sur laquelle nous aurons l'occasion de revenir plus tard. Mentionnons encore que Fréchet attache, comparativement à d'autres auteurs, une plus grande importance à la notion de «somme des revenus supérieurs à une certaine abscisse u ».

Il nous a paru opportun de limiter notre étude à *l'application de la méthode générale* exposée dans le chapitre I à quelques distributions importantes et, avant tout, à celle de Pareto qui, à notre avis, joui dans ce domaine de la même situation privilégiée de départ que la distribution normale dans les problèmes stochastiques. Si tous les auteurs modernes sont unanimes à souligner, d'une part, la remarquable coïncidence de la distribution de Pareto avec les observations statistiques pour les revenus dépassant un certain niveau, ils relèvent tous, d'autre part, son insuffisance pour les petits revenus et essaient de la corriger dans ce dernier secteur. L'application de notre méthode permettra précisément d'obtenir une distribution globale corrigeant d'une manière naturelle et souple ce grave défaut de la distribution de Pareto. Cette nouvelle distribution corrigée sera comparée aux distributions semi-normale et logarithmique normale; toutes les trois obéissent aux postulats énoncés au n° 22 et dépendent de trois paramètres, dont l'un d'entre eux a une signification commune et donne le point d'attache a_0 sur l'abscisse. De plus, l'étude de la fonction de Pareto conduira à un type voisin de cette dernière, limité à droite par un point fini b_0 , dont on tirera parti dans certains problèmes particuliers relatifs à la sécurité sociale. Relevons encore que toutes les

distributions envisagées dans ce chapitre peuvent être déduites analytiquement soit de la distribution normale, soit de la semi-normale.

Dans les problèmes pratiques, on choisira parmi toutes les distributions celle qui ajustera le plus exactement le matériel statistique donné. Telle distribution, bien qu'interprétant fidèlement la distribution des revenus complets ne saurait convenir, par exemple, à celle des salaires ou des revenus afférents à tel groupe économique déterminé. Pour l'ajustement, on utilisera avec fruit la méthode des moments. Ce n'est qu'en comparant le résultat de l'ajustement aux observations statistiques qu'on pourra vérifier la valeur de l'instrument mathématique choisi et l'exactitude des hypothèses dont il est fait mention au n° 23. Parmi le *matériel statistique disponible en Suisse*, citons les publications relatives à l'impôt pour la défense nationale [1], ainsi qu'une enquête spéciale sur les salaires des militaires effectuée par l'Administration des fonds centraux de compensation à Genève [13]; c'est d'ailleurs de cette enquête que l'Office fédéral des assurances sociales a déduit les distributions théoriques utilisées dans les calculs relatifs à l'AVS [29]. Des observations plus complètes concernant le revenu du travail pourront être désormais tirées des comptes individuels de cotisations établis pour l'AVS. Mentionnons encore l'enquête de Brüscheiler [5] relative aux salaires des employés de commerce.

§ 2. Les distributions hyperbolique (Pareto) P et semi-hyperbolique \bar{P}

25. Admettons maintenant que la fonction fondamentale g qui définit dans un plan $(u, x)_t$ quelconque la distribution des revenus sur la droite $x = x_0$ soit donnée par la *fonction de fréquence de Pareto*. Dès lors, il est nécessaire de rappeler brièvement les principales fonctions statistiques relatives à cette distribution classique. A cette fin, nous reprenons d'abord les notations figurant au n° 2. La fonction de fréquence s'écrit, sous sa forme normée, de la manière suivante:

$$f(u) = \alpha a^\alpha u^{-\alpha-1}. \quad (38)$$

Résumons rapidement les principales *propriétés analytiques* de cette fonction de fréquence, en précisant d'abord que ses deux paramètres a et α sont positifs ($\alpha > 0$ résulte de l'existence de l'intégrale des aires; nous verrons qu'au point de vue statistique il faut que $\alpha > 1$). Le champ de variation répond à l'inégalité $a \leq u < \infty$, d'où ressort la

signification du paramètre a . À droite de a , f décroît toujours de $f(a) = \frac{\alpha}{a}$ à $f(\infty) = 0$, sa concavité étant constamment tournée vers le haut. Cramér [8] la classe ainsi parmi les distributions tronquées (voir graphique n° 3). Nous voyons d'emblée que la distribution de Pareto ne satisfait pas les postulats 2° et 3° énoncés au n° 22. Nous verrons plus loin comment on peut se servir, en revanche, de cette distribution pour en construire d'autres répondant entièrement aux trois postulats en question. Notons encore qu'on peut déduire $f(u)$, pour $a = 1$, d'une distribution semi-normale particulière [10], par la substitution $z = \ln u$, et que cette fonction de fréquence découle également de l'équation différentielle de Pearson. — Il est parfois utile de noter la distribution définie par (38) par le symbole $P(u; a, \alpha)$ qui met ainsi en évidence les deux paramètres.

La fonction cumulative complémentaire $H(u)$ mérite une attention spéciale; elle s'obtient à partir de (38) par (2') et s'écrit:

$$H(u) = \left(\frac{a}{u}\right)^\alpha. \quad (39)$$

Elle est du type hyperbolique, comme d'ailleurs la fonction de fréquence elle-même, ce qui justifie l'appellation de P . (39) fournit, en prenant les logarithmes, la droite logarithmique de Pareto bien connue des économistes; on écrira cette dernière de la manière suivante, en posant: $\log u = y$, $\log H = z$ et $\alpha \log a = d$

$$z = d - \alpha y. \quad (39')$$

De cette formule, se dégage la signification du paramètre $-\alpha$ comme coefficient angulaire de la droite logarithmique (voir graphique n° 5). C'est d'ailleurs en appliquant la méthode des moindres carrés à cette droite que les économistes ajustent presque toujours le matériel statistique.

Les moments par rapport à l'origine s'expriment d'une manière plus simple que ceux par rapport à a ou par rapport à la moyenne; (3) donnera pour $a = 0$:

$$M_k = \frac{\alpha}{\alpha - k} a^k. \quad (40)$$

Pour que M_k existe, il faut $\alpha > k$.

A partir de (4'), on calculera encore *la somme réduite des revenus supérieurs à u*; il vient ainsi:

$$\Theta(u) = \frac{\alpha}{\alpha - 1} a^\alpha u^{-\alpha+1}. \quad (41)$$

Pour avoir une somme finie de revenus, il est donc bien nécessaire que $\alpha > 1$, ce qui revient à exiger, au point de vue statistique, au moins l'existence du moment d'ordre 1, c'est-à-dire de la moyenne. Notons que l'on a $\Theta(a) = M$. De (41), on peut d'ailleurs tirer une nouvelle droite logarithmique. Fréchet [14] en propose encore une autre, liant $\log \Theta$ à $\log H$, relation facile à déduire en éliminant u de (39) et (41).

L'indice de concentration ι défini par la formule (5) se présente ici sous une forme particulièrement simple. On voit que:

$$\theta(u) = \frac{\Theta(u)}{M} = \left(\frac{a}{u}\right)^{\alpha-1}, \quad (41')$$

de sorte que l'on obtiendra pour l'indice introduit par Gini [17]:

$$\iota = \frac{\alpha}{\alpha - 1}; \quad (42)$$

ainsi, pour la distribution de Pareto cet indice est indépendant de l'abscisse u considérée. C'est d'ailleurs cette invariance, constatée empiriquement pour beaucoup de distributions de revenus, qui a imposé l'indice en question à l'attention des statisticiens.

26. Comme nous l'avons laissé entendre, nous supposons que la fonction fondamentale g , associée à l'âge x_0 , se présente sous la forme (38). Conformément au n° 9, nous pouvons poser:

$$g(u') = \alpha a_0^\alpha u'^{-\alpha-1}, \quad (43)$$

cette fonction étant définie sur l'intervalle $a_0 \leq u' < \infty$. Opérons le changement de variable (13): $u = s(x) u'$; en ayant soin de garder la forme normée, on aura pour la transformée de (43):

$$\frac{1}{s(x)} g\left[\frac{u}{s(x)}\right] = \frac{\alpha}{s(x)} a_0^\alpha \left[\frac{u}{s(x)}\right]^{-\alpha-1} = \alpha a_0^\alpha s^\alpha(x) u^{-\alpha-1};$$

si nous posons maintenant :

$$a_x = a_0 s(x), \quad (43')$$

nous pouvons écrire, grâce aux deux hypothèses *A* (dilatation) et *B* (translation) :

$$f(u, x) = \alpha a_x^\alpha u^{-\alpha-1}. \quad (43'')$$

Par (13), on passera donc d'une distribution $P(u; a_0, \alpha)$ à une autre $P(u; a_x, \alpha)$. Interprétant ce résultat, nous pouvons affirmer, que *les revenus de chaque ensemble d'âge x se répartissent selon une distribution de Pareto, de paramètres α (invariant) et $a_x = a_0 s(x)$* . Cet énoncé ne nous surprend pas, puisqu'il ne s'agit que d'un cas particulier des formules (14) et (15). Nous pouvons également examiner si les moyennes des distributions sur les droites parallèles à l'axe des u suivent l'échelle $s(x)$; pour ceci, il suffit d'écrire la relation (40), pour $k = 1$, tantôt avec a_x ce qui donne $M(x)$, et tantôt avec a_0 ce qui détermine $M(x_0)$; nous voyons bien que de la division $M(x)$ par $M(x_0)$ résulte la loi $s(x)$.

Grâce à notre hypothèse *B*, nous connaissons également *la distribution des revenus dans n'importe quel plan parallèle au plan initial $t = 0$* ; en effet, nous savons par la formule (17) que

$$\varphi_t(u, x) = \alpha \lambda_t(x) a_x^\alpha u^{-\alpha-1}, \quad (44)$$

de sorte que nous pouvons ainsi aborder, dans un plan économétrique quelconque, la projection de φ_t sur l'axe des u de ce plan.

27. En nous reportant aux remarques du n° 12, nous savons que l'intervalle de définition $a_0 \leq u' < \infty$ pour la distribution relative à l'âge x_0 crée pour la distribution à deux dimensions $\varphi(u, x)$ un *domaine de définition* limité à gauche par une courbe (*a*), d'équation $a_x = \xi = a_0 s(x)$; la limite de droite donnée dans le cas général par la courbe (*b*) est repoussée à l'infini (cf. graphique n° 2 de l'annexe II). De cette manière, nous avons à envisager *pour la fonction de fréquence de la distribution globale des revenus trois intervalles* sur l'axe des u : (1), (1') et (2) dont le nombre se réduit même à deux, dans le cas où $s(x)$ croît d'une manière monotone. Pour la projection, il sied d'appliquer les trois premières formules (21), dans lesquelles on substituera à $\varphi(u, x)$ son expression selon (44). Nous aurons dès lors pour les trois branches de courbe et selon les intervalles :

$$a_0 \leq u < a_1 : \bar{f}_1(u) = \alpha a_0^\alpha u^{-\alpha-1} \int_{x_0}^{\psi_1} \lambda(x) s^\alpha(x) dx, \quad (45)$$

$$a_1 \leq u < a' : \bar{f}_1(u) = \alpha a_0^\alpha u^{-\alpha-1} \left[\int_{x_0}^{\psi_1} \lambda(x) s^\alpha(x) dx + \int_{\psi_2}^{x_1} \lambda(x) s^\alpha(x) dx \right], \quad (45')$$

$$a' \leq u < \infty : \bar{f}_2(u) = \alpha a_0^\alpha u^{-\alpha-1} \int_{x_0}^{x_1} \lambda(x) s^\alpha(x) dx; \quad (45'')$$

les limites variables ψ_1 et ψ_2 sont à calculer au point $\frac{u}{a_0}$. Les équations

(45) et (45') sont ainsi de la forme (26'), et (45'') se présente comme cas particulier de (26). Ceci ne doit guère nous étonner puisque (43)

montre que la fonction fondamentale $g\left(\frac{u}{s}\right)$ de Pareto est bien de la forme $A\left(\frac{u}{s}\right)^B$. De ce fait, la branche $\bar{f}_2(u)$ de la marginale représente

elle-même une fonction de fréquence d'une distribution de Pareto $P(u; \bar{a}, \alpha)$ où l'on a pour la constante \bar{a} :

$$\bar{a}^\alpha = a_0^\alpha \int_{x_0}^{x_1} \lambda(x) s^\alpha(x) dx. \quad (45''')$$

Quoique l'allure de la courbe d'ensemble dépende avant tout du choix des deux fonctions fondamentales $\lambda(x)$ et $s(x)$, on voit cependant facilement que, même si ces fonctions ne sont pas précisées, la fonction de fréquence \bar{f} s'annule au point a_0 , où elle est croissante, pour atteindre au point a' la valeur $\alpha \bar{a}^\alpha a'^{-\alpha-1}$ et décroître ensuite d'une manière monotone comme le veut la loi de Pareto. Cette distribution marginale est entièrement déterminée si l'on connaît les deux paramètres α et a_0 ainsi que les deux fonctions $\lambda(x)$ et $s(x)$. Nous sommes donc en présence d'une *nouvelle distribution* $\bar{P}(u; \lambda, s, a_0, \alpha)$ que nous proposons d'appeler *semi-hyperbolique*, du fait que la branche droite de la fonction de fréquence constitue une courbe hyperbolique de Pareto. En général, a_1 et a' sont les abscisses de points singuliers de la courbe d'ensemble puisque celle-ci y présentera deux tangentes différentes. Cependant, les deux premières conditions (23) nous enseignent comment on peut opérer en ces points le *raccord tangentiel*; $f(a_1, x_1)$ ainsi que $f(a', x')$ étant positifs dans le cas considéré ici, il faut essayer d'obtenir le contact du 1^{er} ordre à l'aide des fonctions λ et ψ' . En a_1 la jonction sera

tangentielle si $x_1 = \omega$, l'on a ainsi $\lambda(\omega) \equiv 0$. En a' , la question est plus délicate puisqu'en général $\lambda(x') > 0$; il faudra alors examiner $\psi_1'(s') - \psi_2'(s')$, ce qui peut donner lieu à une indétermination que l'on essaiera de lever.

Les autres fonctions statistiques liées à \bar{P} s'obtiennent directement de nos formules générales. Ainsi la cumulative complémentaire \bar{H} se calcule selon (30') où $H(u, x) = a_0^\alpha s''(x) u^{-\alpha}$; la somme réduite $\bar{\Theta}$ des revenus $> u$ découle de (31') avec $\Theta(u, x) = \frac{\alpha}{\alpha-1} a_0^\alpha s''(x) u^{-\alpha+1}$ et $M(x) = \frac{\alpha}{\alpha-1} a_0 s(x)$; dans les expressions donnant $\bar{H}(u)$ et $\bar{\Theta}(u)$, on mettra en évidence pour $u > a'$ le paramètre \bar{a} défini par (45'''). Enfin, les moments sont donnés par (34) en se rappelant que

$$M_k(x) = \frac{\alpha}{\alpha-k} a_0^k s^k(x).$$

28. En choisissant pour $\lambda(x)$ et $s(x)$ des fonctions convenables, on pourra effectuer les intégrations indiquées par les formules (45). Nous allons considérer un cas particulièrement simple où $\lambda(x)$ et $s(x)$ sont des fonctions linéaires. Supposons que $\lambda(x)$ soit de la forme $A - Bx$ et déterminons A et B de façon que, d'une part, $\lambda(x_1) = 0$ et, d'autre part, $\lambda(x)$ soit normé entre x_0 et x_1 ; ainsi il vient:

$$\lambda(x) = 2 \frac{x_1 - x}{(x_1 - x_0)^2}. \quad (46)$$

Pour $s(x)$ nous admettons l'expression que voici:

$$s(x) = 1 + C(x - x_0), \quad (47)$$

de sorte que $s(x_0) = 1$. En se souvenant que $s(x_1) = \frac{a_1}{a_0}$ on trouvera encore une expression utile pour la suite des calculs:

$$C(x_1 - x_0) = \frac{a_1 - a_0}{a_0}; \quad (47')$$

(47) donne lieu à la fonction inverse suivante:

$$x = \psi(s) = x_0 + \frac{s-1}{C}. \quad (48)$$

Le domaine d'intégration décrit au début du n° 27 se simplifie encore; d'abord x' et x_1 coïncident (cas monotone) et la courbe (a) se réduit à la droite $\xi = a_0[1 + C(x - x_0)]$. Pour le calcul des intégrales (45) et (45'') on effectuera le changement de variable indiqué soit par (19), soit par (20), c'est-à-dire $\xi = a_0 s(x)$ ou $x = \psi\left(\frac{\xi}{a_0}\right)$.

Tous calculs faits, on trouvera pour la fonction de fréquence marginale une courbe d'ensemble à deux branches, dont voici les équations:

$$\bar{f}_1(u) = \frac{2\alpha}{(a_1 - a_0)^2} \left[\frac{a_1}{\alpha + 1} - \frac{u}{\alpha + 2} - a_0^{\alpha+1} \left(\frac{a_1}{\alpha + 1} - \frac{a_0}{\alpha + 2} \right) u^{-\alpha-1} \right], \quad (49)$$

$$\bar{f}_2(u) = \frac{2\alpha}{(a_1 - a_0)^2} \left[\frac{a_1(a_1^{\alpha+1} - a_0^{\alpha+1})}{\alpha + 1} - \frac{a_1^{\alpha+2} - a_0^{\alpha+2}}{\alpha + 2} \right] u^{-\alpha-1}. \quad (49')$$

Alors que $\bar{f}_2(u)$ est bien une courbe de Pareto, valable pour $u \geq a_1$, la branche $\bar{f}_1(u)$ se présente sous la forme d'une différence entre une droite décroissante et une courbe de Pareto; la droite est une corde de cette courbe, leurs ordonnées coïncidant en a_0 . L'étude des dérivées première et seconde montre que la fonction \bar{f}_1 a sa concavité tournée vers le bas et qu'elle croît en a_0 et décroît en a_1 ; la première branche a donc un maximum entre a_0 et a_1 . Le sens de concavité de la courbe d'ensemble change d'ailleurs au point a_1 . Nous savons d'avance qu'en ce point les deux courbes se raccordent tangentiellement, puisque nous avons choisi $\lambda(x_1) = 0$; on le vérifiera d'ailleurs directement à partir de (49) et (49'). Cette fonction de fréquence définit une *distribution semi-hyperbolique particulière* $\bar{P}(u; a_0, a_1, \alpha)$; en effet, trois paramètres suffisent pour la déterminer complètement, ainsi qu'il ressort des formules ci-dessus. L'allure de la fonction de fréquence d'une telle distribution est donnée au graphique n° 3 de l'annexe II.

Pour les applications il est utile de connaître encore *les autres fonctions statistiques*. Nous les déduisons directement de (30'), (31') et (34) en procédant comme indiqué à la fin du n° 27 et en effectuant toujours le même changement de variable $x = \psi\left(\frac{\xi}{a_0}\right)$. Nous aurons ainsi successivement, soit pour l'intervalle (1), soit pour (2):

Cumulative complémentaire:

$$\bar{H}_1(u) = \frac{2}{(a_1 - a_0)^2} \left[\frac{\alpha}{\alpha + 2} \frac{u^2}{2} - a_1 \frac{\alpha}{\alpha + 1} u + \frac{a_1^2}{2} - a_0^{\alpha+1} \left(\frac{a_1}{\alpha + 1} - \frac{a_0}{\alpha + 2} \right) u^{-\alpha} \right], \quad (50)$$

$$\bar{H}_2(u) = \frac{2}{(a_1 - a_0)^2} \left[\frac{a_1(a_1^{\alpha+1} - a_0^{\alpha+1})}{\alpha + 1} - \frac{a_1^{\alpha+2} - a_0^{\alpha+2}}{\alpha + 2} \right] u^{-\alpha}. \quad (50')$$

Somme complémentaire réduite:

$$\bar{\Theta}_1(u) = \frac{2\alpha}{(\alpha - 1)(a_1 - a_0)^2} \left[\frac{a_1^3}{6} - a_1 \frac{\alpha - 1}{\alpha + 1} \frac{u^2}{2} + \frac{\alpha - 1}{\alpha + 2} \frac{u^3}{3} - a_0^{\alpha+1} \left(\frac{a_1}{\alpha + 1} - \frac{a_0}{\alpha + 2} \right) u^{-\alpha+1} \right], \quad (51)$$

$$\bar{\Theta}_2(u) = \frac{2\alpha}{(\alpha - 1)(a_1 - a_0)^2} \left[\frac{a_1(a_1^{\alpha+1} - a_0^{\alpha+1})}{\alpha + 1} - \frac{a_1^{\alpha+2} - a_0^{\alpha+2}}{\alpha + 2} \right] u^{-\alpha+1}. \quad (51')$$

Moments d'ordre k :

$$\bar{M}_k = \frac{2\alpha}{(\alpha - k)(a_1 - a_0)^2} \left[\frac{a_1(a_1^{k+1} - a_0^{k+1})}{k + 1} - \frac{a_1^{k+2} - a_0^{k+2}}{k + 2} \right]. \quad (52)$$

En particulier, les trois premiers moments peuvent se mettre sous la forme des relations suivantes qui sont utiles si l'on veut déterminer les trois paramètres a_0 , a_1 et α à partir des données empiriques:

$$\left. \begin{aligned} \bar{M} &= \frac{\alpha}{\alpha - 1} \frac{2a_0 + a_1}{3} \\ \bar{M}_2 &= \frac{\alpha}{\alpha - 2} \frac{3a_0^2 + 2a_0a_1 + a_1^2}{6} \\ \bar{M}_3 &= \frac{\alpha}{\alpha - 3} \frac{4a_0^3 + 3a_0^2a_1 + 2a_0a_1^2 + a_1^3}{10} \end{aligned} \right\} \quad (52')$$

Ces expressions rappellent d'ailleurs celle de la formule (40) où l'on aurait remplacé a , a^2 ou a^3 par les moyennes indiquées dans les deuxièmes fractions de (52'). L'existence des moments de \bar{P} est liée à la condition $\alpha > k$, comme cela est le cas pour P .

29. Aujourd'hui encore on se sert, pour représenter la distribution des revenus de la population entière, de la loi de Pareto, en ne l'appliquant cependant qu'à des revenus qui dépassent un certain niveau. Ainsi, M. Winkler, par exemple, a présenté très récemment un travail sur cette loi au congrès de l'Institut international de statistique qui eut lieu en 1949 à Berne [37]. Plusieurs auteurs ont essayé d'apporter à la loi de Pareto des *corrections artificielles* pour les petits revenus. Dans ce but, Rhodes [34] également soucieux de garder pour la branche de droite (queue) la distribution de Pareto, a imaginé un procédé pour substituer à cette dernière une branche (tête) d'une autre courbe. Cependant, le résultat de ce procédé ne saurait être considéré comme satisfaisant, du moment qu'au point de jonction il y a une discontinuité et que la loi des aires n'est pas rigoureusement observée par la courbe d'ensemble. — Signalons encore un travail de Cockfield [7] qui démontre à l'aide d'observations fiscales anglaises récentes que la distribution de Pareto s'adapte fort bien, du moins à partir des revenus dépassant £ 200. Pour les revenus inférieurs, il « casse » d'une manière empirique la droite logarithmique. — Maurice Fréchet [15] a exposé, antérieurement à Rhodes, une autre méthode destinée à corriger la distribution de Pareto pour les petits revenus. A cette fin, il part de la distribution qui représente la première loi de Laplace à fonction de fréquence $\frac{1}{2} e^{-|z|}$; en lui appliquant la transformation: $z = a \ln(u - u_0) + b$, il trouve une courbe à deux branches (une pour $z < 0$, l'autre pour $z > 0$), dont celle de droite est encore une courbe classique de Pareto, valable à partir de la médiane m , celle de gauche étant du type « parabolique » que nous introduisons au § 4 ci-après. Fréchet montre la continuité de la courbe d'ensemble, même au point de jonction m des deux branches; en revanche, elles n'ont pas la même tangente en ce point. Toutes ces améliorations sont cependant réalisées par des artifices. Il suffit, pour obtenir une *correction naturelle* de supposer que les ensembles d'âge x se distribuent selon P , ce qui conduit pour l'ensemble de la population à la distribution semi-hyperbolique. Dans ce cas les branches de la courbe de fréquence se raccordent toujours sans aucune discontinuité et même dans de nombreux cas peuvent se joindre tangentiellement.

Si la distribution $P(u; a, \alpha)$ de Pareto donne la répartition des revenus en *première approximation*, notre distribution semi-hyperbolique $\bar{P}(u; a_0, a_1, \alpha)$ décrite au n° 28 résout le problème en *deuxième*

approximation. En effet, on aurait pour l'ensemble de la population également une distribution de Pareto si l'on admettait dans notre méthode que $s(x)$ se réduit à une constante = 1. En introduisant l'hypothèse de linéarité pour $\lambda(x)$ et $s(x)$, on s'approche déjà un peu plus de la réalité. On trouverait une *approximation encore meilleure*, si l'on choisissait, par exemple, pour $\lambda(x)$ une structure d'âge observée et pour $s(x)$ un polynome convenable. La courbe de fréquence proposée en deuxième approximation répond déjà aux trois postulats du n° 22, ce qui n'est pas encore le cas pour la première approximation. Relevons encore que la semi-hyperbolique permet également de calculer par une intégration simple la somme des revenus dépassant une certaine abscisse u ; ainsi, cette distribution satisfait une exigence fondamentale énoncée par Fréchet.

Nous avons constaté au n° 26 *l'invariance du paramètre α* par rapport à la substitution $u = u' s(x)$ où l'on pourrait aussi attribuer à x d'autres significations que l'âge. Ce fait explique peut-être le phénomène remarquable que les droites logarithmiques concernant tous les pays observés par Pareto étaient presque parallèles. Cette constatation ne doit cependant être valable que pour des époques pendant lesquelles la situation économique reste stable. En effet, la dépréciation continue de la monnaie semble provoquer une augmentation de α ; tel est du moins le résultat des observations statistiques faites par Billeter [4] et selon lesquelles l'indice de concentration
$$\iota = \frac{\alpha}{\alpha - 1}$$
 a diminué au cours des dernières décennies, surtout lors des dévaluations marquées faisant suite aux guerres.

Nous pouvons évidemment considérer les distributions semi-hyperboliques comme représentant à leur tour la répartition de revenus des ensembles d'âge x . On prendra alors, par exemple, comme fonction fondamentale la distribution $\bar{P}(u'; a_0, a_1, \alpha)$. En lui appliquant la substitution $u = u' s(x)$, a_0 et a_1 se transformeront encore selon le modèle (43'), tandis que α reste de nouveau invariant. Si l'on étudie la nouvelle distribution globale correspondante, on trouvera aisément que la branche droite pour $u \geq a_2$, où $a_2 = s_1 a_1 = s_1^2 a_0$, sera encore une courbe de Pareto. On se trouve en présence d'une *itération* qu'on pourra continuer indéfiniment et pour laquelle il subsistera toujours une courbe de Pareto à droite du point $a_n = s_1^n a_0$.

§ 3. Les distributions semi-normale G et logarithmique normale J

30. Dans ce § nous allons étudier les lois de transformation et les propriétés projectives de deux distributions qui se sont avérées également utiles dans la théorie de la répartition des revenus. Considérons d'abord *la fonction de fréquence de la distribution semi-normale* $G(u; a, \gamma, \varepsilon)$, qui dépend de trois paramètres comme le montre son expression sous forme normée:

$$f(u) = \frac{\gamma^\varepsilon}{\Gamma(\varepsilon)} (u - a)^{\varepsilon-1} e^{-\gamma(u-a)}. \quad (53)$$

Le paramètre $a > 0$ définit l'abscisse initiale de l'intervalle de définition: $a \leq u < \infty$; nous verrons ci-après le sens des deux autres paramètres, γ et ε , ainsi que les propriétés essentielles de la fonction eulérienne de deuxième espèce Γ . La relation (53) se déduit, par une substitution linéaire, de l'équation de la courbe de Pearson du type III. C'est Steffensen [36] qui a introduit la dénomination de «semi-normale», en prenant toutefois comme variable $y^2 = u - a$. La formule (53) est une généralisation de la forme que nous trouvons chez Féraud [10] pour $a = 0$. Il est d'ailleurs bien connu que $G(u; 0, 1, \frac{r}{2})$ est la distribution du $\chi^2 = u$. Steffensen montre que G tend, pour $\varepsilon = \infty$, vers une distribution normale N . Avant de rappeler les propriétés analytiques et statistiques de G , notons que leur étude introduit les fonctions eulériennes de deuxième espèce, à savoir:

la fonction Γ complète:
$$\Gamma(\varepsilon) = \int_0^\infty \zeta^{\varepsilon-1} e^{-\zeta} d\zeta,$$

la fonction Γ incomplète:
$$\Gamma(z, \varepsilon) = \int_0^z \zeta^{\varepsilon-1} e^{-\zeta} d\zeta.$$

Pour la première, on utilisera ici la relation:

$$\Gamma(\varepsilon + k) = \varepsilon(\varepsilon + 1) \dots (\varepsilon + k - 1) \Gamma(\varepsilon),$$

où k est un entier positif. Cette formule de récurrence permet également de ramener le calcul de $\Gamma(\varepsilon + k)$ à celui de $\Gamma(1 \leq \varepsilon \leq 2)$ qu'on rencontre dans les tables de Pearson [33].

Les valeurs numériques de la seconde de ces fonctions peuvent être déduites de l'intégrale: $I(z, \varepsilon - 1) = \frac{\Gamma(z, \varepsilon)}{\Gamma(\varepsilon)}$ qu'on trouvera dans un autre recueil de tables de Pearson [32], dans les colonnes à paramètre $p = \varepsilon - 1$ et à l'argument $u^* = \frac{z}{\sqrt{\varepsilon}}$.

La formule (53), où γ et ε sont positifs, comporte deux familles principales de courbes, suivant la valeur de ε :

- 1° $\varepsilon \leq 1$ courbe toujours décroissante, rappelant celle de Pareto; deux cas sont à distinguer: $\varepsilon < 1$ (asymptotique aux deux droites: $f = 0$ et $u = a$) et $\varepsilon = 1$ (asymptotique à $f = 0$, mais valeur finie en $u = a$).
- 2° $\varepsilon > 1$ courbe unimodale, s'annulant en $u = a$ et $u = \infty$, avec mode $u_0 = \frac{\varepsilon - 1}{\gamma} + a$ dont les deux points d'inflexion sont symétriques par rapport à u_0 et distants de $\frac{\sqrt{\varepsilon - 1}}{\gamma}$. Trois cas sont à distinguer suivant les tangentes à la courbe en $u = a$: $\varepsilon < 2$ (tangente verticale), $\varepsilon = 2$ (tangente bissectrice) et $\varepsilon > 2$ (tangente horizontale).

Considérons maintenant les autres fonctions statistiques introduites au n° 2. Pour la fonction cumulative, il vient:

$$F(u) = \frac{\Gamma[\gamma(u - a), \varepsilon]}{\Gamma(\varepsilon)} = I[\gamma(u - a), \varepsilon - 1]; \quad (54)$$

le calcul de $F(u)$ se ramène ainsi à une simple lecture des tables de Pearson [32] en prenant comme argument $u^* = \frac{\gamma(u - a)}{\sqrt{\varepsilon}} = \frac{u - a}{\mu_2}$; on voit que Pearson a pris comme unité d'argument la dispersion μ_2 (voir ci-après).

Les moments par rapport à a prennent ici la forme simple que voici:

$${}_aM_k = \frac{1}{\gamma^k} \frac{\Gamma(\varepsilon + k)}{\Gamma(\varepsilon)} = \frac{\varepsilon(\varepsilon + 1) \dots (\varepsilon + k - 1)}{\gamma^k}. \quad (55)$$

Relevons en particulier les expressions obtenues, pour certaines grandeurs typiques, à partir des moments pris par rapport à la moyenne ${}_aM = \frac{\varepsilon}{\gamma}$:

Dispersion (standard deviation): $\mu_2 = \sqrt{\mathfrak{M}_2} = \frac{\sqrt{\varepsilon}}{\gamma},$

Dissymétrie (skewness): $\mu_3 = \frac{\mathfrak{M}_3}{\mu_2^3} = \frac{2}{\sqrt{\varepsilon}},$

Excès (kurtosis): $\mu_4 = \frac{\mathfrak{M}_4}{\mu_2^4} - 3 = \frac{6}{\varepsilon}.$

On en déduit les deux nombres de Pearson indiquant le type de courbe de son système:

$$\beta_1 = \mu_3^2 = \frac{4}{\varepsilon},$$

$$\beta_2 = \mu_4 + 3 = \frac{6}{\varepsilon} + 3.$$

Pour la somme réduite $\Phi(u)$ des revenus $\leq u$, on aura une expression qui se ramène également aux fonctions eulériennes:

$$\Phi(u) = {}_aF(u) + \frac{\Gamma[\gamma(u-a), \varepsilon + 1]}{\gamma \Gamma(\varepsilon)}. \quad (56)$$

Nous pouvons maintenant examiner les lois de transformation de G . Nous admettons d'abord pour l'âge x_0 une distribution $G(u'; a_0, \gamma_0, \varepsilon)$. On voit aisément qu'elle se transforme par la substitution $u = s(x) u'$ en une distribution $G(u; a_x, \gamma_x, \varepsilon)$; les nouveaux paramètres se déduisent de la manière suivante de ceux concernant la fonction fondamentale $f(u', x_0)$, calculée d'après (53):

$$a_x = a_0 s(x), \quad (57)$$

$$\gamma_x = \frac{\gamma_0}{s(x)}; \quad (57')$$

le paramètre ε est invariant. Les distributions conditionnelles parallèles à l'axe des u sont définies par une famille de courbes semi-normales, ce qui est également conforme à notre théorie générale exposée au n° 9. Même si $a_0 = 0$, les courbes de fréquence seraient dilatées en vertu de la relation (57').

Les propriétés projectives de G sont moins simples. La fonction de fréquence à deux dimensions $\varphi(u, x) = \lambda(x) f(u, x)$ se trouve maintenant définie dans un domaine D , limité à gauche par la courbe (a) d'équation $\xi = a_x$, et ouvert vers la droite. Pour la fonction de fréquence concernant la distribution globale on obtient, comme dans le cas Pareto, une courbe à trois branches, sauf dans le cas particulier où $a_0 = 0$, qui ne comporte qu'une seule branche. Cependant, les intégrales de la formule (21) ne peuvent plus se ramener aux fonctions classiques, même pour le cas linéaire de $s(x)$ et $\lambda(x)$. En particulier, ces marginales ne s'expriment plus par les fonctions I incomplètes. Pour les applications, on en est réduit à des évaluations numériques, faciles à effectuer en pondérant les conditionnelles semi-normales avec $\lambda(x)$. Les marginales n'appartiennent pas à la famille des semi-normales, même sur l'intervalle (2). On n'obtient une marginale semi-normale que dans le cas banal où $s(x) = 1$. En revanche, il y a en général raccord tangentiel en a_1 ; vu que $f(a_x, x) \equiv 0$, les conditions (23) sont ainsi remplies en ce point.

31. Nous allons considérer encore une autre distribution dont la fonction de fréquence est définie dans l'intervalle $a_x \leq u < \infty$. Il s'agit de la *distribution logarithmique normale* $J(u; a, c, \varkappa)$, appelée aussi distribution de Galton, de McAlister, d'Edgeworth, de Kapteyn ou de Gibrat. Nous écrivons pour sa fonction de fréquence:

$$f(u) = \frac{1}{\varkappa(u-a)\sqrt{2\pi}} e^{-\frac{1}{2\varkappa^2} \left(\ln \frac{u-a}{c-a}\right)^2}. \quad (58)$$

Les paramètres a , c et \varkappa sont tous positifs. Le changement de variable:

$$z = \frac{1}{\varkappa} \ln \frac{u-a}{c-a}, \quad (59)$$

donc

$$dz = \frac{du}{\varkappa(u-a)}, \quad (59')$$

ramène J à une distribution normale réduite $N\left(z; 0, \frac{1}{\sqrt{2}}\right)$ et permet ainsi d'effectuer les calculs numériques à l'aide de tables donnant les fonctions de fréquence et les fonctions cumulatives de N . Sous cette dernière forme, la précision de N est donc $= \frac{1}{\sqrt{2}}$. Si dans le change-

ment de variable envisagé ci-dessus, on avait négligé le facteur $\frac{1}{z}$, on aurait été d'abord ramené à $N\left(z; 0, \frac{1}{z\sqrt{2}}\right)$ qui fournit la signification du paramètre z comme dispersion de la distribution N correspondant à J . Si l'on dispose de l'autre série classique de tables relatives à $N(z^*; 0, 1)$, on posera $z^* = \frac{1}{z\sqrt{2}} \ln \frac{u-a}{c-a}$.

Les propriétés analytiques de $f(u)$ ne sont pas très simples. Fréchet [14] fait remarquer que $f(u)$ est unimodale et que son maximum se trouve entre a et c . On trouvera au sujet de ces propriétés des indications assez complètes chez Gibrat [16] qui désigne d'ailleurs l'équation (59'), où du est proportionnel à u , comme «loi de l'effet proportionnel». Cette transformation de Gibrat a été fort remarquée; Rhodes [34] l'utilise, par exemple, pour transformer une courbe de Pareto et faire de la courbe ainsi modifiée la «tête» d'une courbe discontinue à deux branches, le tronçon de droite restant une courbe de Pareto et celui de gauche prenant l'allure d'une courbe unimodale (cf. n° 29). D'ailleurs la transformation (59') est la même que celle utilisée par Fréchet [l. c.] dans une de ses explications probabilistes résumées au n° 23 et qui mène effectivement à la distribution J envisagée ici. On trouvera pour cette dernière des exemples numériques dans un ouvrage allemand récent [3].

Les seules fonctions statistiques qui s'expriment simplement sont, outre la fonction de fréquence, la cumulative $F(u)$, ainsi que les moments; on a d'abord:

$$F(u) = \frac{1}{2} [1 + \Psi(z)], \quad (60)$$

z étant défini par (59). On trouve cette expression toute calculée, par exemple, dans les tables de Pearson [33], sous la désignation $\frac{1}{2}(1 + \alpha)$; Ψ représente ici l'intégrale donnant la probabilité, selon N , d'une erreur absolue $< z$. Quant à l'expression des moments par rapport à a , on aura:

$${}_aM_k = (c-a)^k e^{\frac{k^2 z^2}{2}}; \quad (61)$$

elle résulte du calcul d'une intégrale de la forme

$$\int_{-\infty}^{+\infty} e^{-mz^2 + 2nz} dz = e^{\frac{n^2}{m}} \sqrt{\frac{\pi}{m}}.$$

En revanche, le calcul de la somme réduite des revenus inférieurs ou supérieurs à u aboutit à une expression dont l'intégration ne peut s'effectuer par les procédés analytiques usuels. Vu l'importance pratique de cette notion, Fréchet [14] a tenu à mettre en évidence cet inconvénient. Rappelons qu'aucune des autres distributions envisagées jusqu'ici ne présente le désavantage en question; cette constatation est heureuse, puisque, dans le calcul des rentes moyennes (voir chap. III), l'introduction de cette fonction rend de précieux services.

Quant aux *lois de transformation et les propriétés projectives* de cette distribution, on verra facilement que la fonction de fréquence de $J(u'; a_0, c_0, \varkappa)$ relative à l'âge x_0 se transformera, pour l'âge x , en la fonction de fréquence de $J(u; a_x, c_x, \varkappa)$. Le paramètre \varkappa est invariant, tandis que les deux autres s'obtiennent ainsi qu'il suit:

$$a_x = a_0 s(x), \quad (62)$$

$$c_x = c_0 s(x). \quad (62')$$

Pour les propriétés projectives, il faut s'en tenir aux trois premières formules (21) qui donneront pour la distribution globale des revenus une courbe dont deux des trois branches se raccordent en général tangentiellement en a_1 . Tout ce que nous avons dit au sujet des propriétés projectives pour la distribution semi-normale G peut d'ailleurs être transcrit ici.

32. Pour représenter la distribution effective des revenus, c'est, après la distribution de Pareto, la logarithmique normale J dont on s'est servi le plus fréquemment jusqu'à ce jour. En revanche, la distribution semi-normale G n'a pas été utilisée, du moins à notre connaissance, dans les problèmes de ce genre. Pour notre part, nous avons eu l'occasion de l'employer pour résoudre le problème de la rente moyenne ordinaire de l'AVS [22]. On ne saurait donner, a priori, la préférence à l'une ou à l'autre des distributions traitées. Leur choix devra toujours être dicté par la structure du matériel statistique donné. *La comparaison des trois distributions unimodales \bar{P} , G et J* est dès lors naturelle. A cet effet, nous avons réuni sur le graphique n° 3 de l'annexe II leur fonction de fréquence respective. Il ne s'agit là que d'exemples, l'allure de ces courbes variant dans une forte mesure avec la valeur de leurs paramètres. Nous pensons que l'une

ou l'autre des deux courbes de fréquence de la distribution semi-hyperbolique $\bar{P}(u; a_0, a_1, \alpha)$ reproduites au graphique n° 3 correspond en deuxième approximation (cf. n° 29) à la répartition actuelle des revenus complets en Suisse. Nous avons également reporté sur le graphique la première approximation donnée par une courbe de Pareto P , de même paramètre $\alpha = 2$ et de même moyenne $M = 6000$ francs que les semi-hyperboliques \bar{P} . Ces trois distributions conduisent à la même somme totale annuelle de revenus personnels, évaluée actuellement à 16 milliards de francs suisses environ [6]. Les distributions G et J ont également la même moyenne, mais des abscisses initiales a (minimum vital), ainsi que des dispersions différentes. Signalons à ce propos que les valeurs relativement basses du paramètre α qu'on tire des statistiques de revenus provoque une dispersion très grande, sinon infinie, ainsi qu'il ressort des formules (40) et (52), pour $k = 2$. Il s'agit là, semble-t-il, d'une caractéristique de la distribution des *revenus complets*, provenant de quelques revenus isolés très élevés. En revanche, ce phénomène ne s'observe pas dans la distribution particulière aux *revenus du travail*. Pour cette raison, il convient de choisir pour cette dernière soit une distribution \bar{P} à paramètre α nettement supérieur à 2, soit une distribution G ou J . — L'étude du graphique n° 5 montre que la droite logarithmique concernant la cumulative de \bar{P} se transforme sans heurt en courbe au point a_1 , ce qui est la conséquence du contact d'ordre 2 existant en ce point. On évite ainsi de manière naturelle la cassure de Cockfield [7] déjà relevée au n° 29.

§ 4. Les distributions parabolique Q et semi-parabolique \bar{Q}

33. Dans le présent chapitre, nous avons jusqu'alors uniquement considéré des distributions définies dans un intervalle $a \leq u < \infty$; ceci est d'ailleurs le cas usuel pour la répartition des revenus, comme le fait remarquer M. Fréchet [14], du moins lorsqu'on ignore le montant du revenu le plus élevé. Abordons maintenant le cas d'une *distribution définie sur un intervalle* $0 \leq u < b$. Par la formule (38), nous avons donné la fonction de fréquence de Pareto sous sa forme normée. La forme générale, contenant une constante arbitraire C , s'écrit :

$$f(u) = C u^{-a-1}. \quad (63)$$

Par définition, on obtient la forme normée en déterminant C par la loi des aires, en supposant $b > a$:

$$C \int_a^b u^{-\alpha-1} du = \frac{C}{\alpha} (a^{-\alpha} - b^{-\alpha}) = 1.$$

Il est curieux que l'on n'ait envisagé jusqu'aujourd'hui — du moins à notre connaissance — que le cas où α est positif. L'intégrale des aires n'a alors de sens que lorsque $a > 0$, cependant que b peut croître au-delà de toute limite finie. En revanche, si α est négatif, a peut s'annuler, mais b ne doit pas dépasser un nombre fini. Posons donc $\beta = -\alpha$ et prenons comme nouveau champ de variation $0 \leq u < b$; on obtient ainsi, sous forme normée, la fonction de fréquence suivante:

$$f(u) = \beta b^{-\beta} u^{\beta-1}. \quad (63')$$

Considérons rapidement *les propriétés analytiques de la fonction de fréquence $f(u)$* . La formule (63') comprend, contrairement à (38), plusieurs formes de courbes (cf. graphique n° 4):

- 1° $\beta < 1$ courbe décroissante de $f(0) = \infty$ à $f(b) = \frac{\beta}{b}$, concavité tournée vers le haut;
- 2° $\beta = 1$ droite horizontale de la distribution rectangulaire $f = \frac{1}{b}$;
- 3° $1 < \beta < 2$ courbe croissante de $f(0) = 0$ à $f(b) = \frac{\beta}{b}$, concavité tournée vers le bas;
- 4° $\beta = 2$ droite croissante par l'origine, tronquée par $f(b) = \frac{2}{\beta}$;
- 5° $\beta > 2$ courbe croissante de $f(0) = 0$ à $f(b) = \frac{\beta}{b}$, concavité tournée vers le haut.

Il est intéressant de noter la correspondance entre certaines droites du plan (u, f) et celles du plan logarithmique (y, z) considéré ci-après. L'étude de la moyenne permet de montrer que dans les problèmes pratiques on trouvera en général $\beta < 1$.

Mentionnons maintenant *la fonction cumulative* obtenue de (63') à l'aide de (2):

$$F(u) = \left(\frac{u}{b}\right)^\beta. \quad (64)$$

Il ne s'agit donc plus ici de la cumulative complémentaire comme c'était le cas dans la distribution de Pareto exposée au n° 25. En comparant $F(u)$ à $H(u)$ donné par (39), on comprend pourquoi nous appelons *distribution parabolique* $Q(u;b,\beta)$ cette nouvelle distribution définie par (63'). Plus exactement il faudrait parler de «distribution à cumulative parabolique», puisque la fonction de fréquence elle-même renferme ($\beta < 1$) un type de courbe hyperbolique. Cependant, la désignation abrégée ne saurait prêter à confusion. De (64) on tire une droite logarithmique:

$$z = \beta y - d, \quad (64')$$

où $y = \log u$, $z = \log F$ et $d = \beta \log b$. Cette droite, à coefficient angulaire $\beta > 0$ est donc croissante (cf. graphique n° 5 de l'annexe II).

Les autres fonctions statistiques sont analogues à celles de la distribution hyperbolique; en effet, on trouvera pour *les moments* pris par rapport à l'origine:

$$M_k = \frac{\beta}{\beta + k} b^k. \quad (65)$$

L'existence des moments de tout ordre est donc toujours assurée, ceci contrairement au type hyperbolique. On aura encore pour *la somme réduite des revenus* $< u$ l'expression simple que voici:

$$\Phi(u) = \frac{\beta}{\beta + 1} b^{-\beta} u^{\beta+1}. \quad (66)$$

Ici également, on pourrait mettre en évidence l'existence d'autres droites logarithmiques, liant d'une part $\log \Phi$ à $\log u$ et d'autre part $\log F$ à $\log \Phi$. — Il est curieux de noter que la branche gauche de la courbe corrigée de Pareto déduite par Fréchet [15] de la première loi de Laplace (cf. n° 29) est une courbe parabolique selon (63').

34. *Les lois de transformation* de $Q(u;b,\beta)$ sont symétriques à celles de $P(u;a,\alpha)$. En effet, si l'on opère le changement de variable (13): $u = s(x) u'$ sur la fonction de fréquence de $Q(u';b_0,\beta)$ valable pour x_0 , on trouvera pour l'âge x une distribution $Q(u;b_x,\beta)$; le paramètre β est invariant, tandis que l'on obtient b_x de la manière suivante:

$$b_x = b_0 s(x). \quad (67)$$

En nous reportant aux remarques du n^o 12, on constate que le domaine D (cf. graphique n^o 2), à l'intérieur duquel est définie la fonction de fréquence à deux dimensions $\varphi(u, x)$, voit sa limite (a) se confondre avec l'axe des x , tandis qu'il est limité à droite par la courbe (b). La projection pondérée des conditionnelles $Q(u; b_x, \beta)$ sur l'axe des u donne lieu à une *marginal* $\bar{Q}(u; \lambda, s, b_0, \beta)$ appelée *distribution semi-parabolique*. En effet, dans l'intervalle (2): $0 \leq u \leq b_0$ la fonction de fréquence \bar{f} est encore une fonction de fréquence parabolique du fait que (63') est bien de la forme (24). Dans le cas d'un $s(x)$ monotone, la courbe, dont les trois branches sont définies par les trois dernières formules (21), se réduit à une fonction de fréquence à deux branches. Le raccord tangentiel peut être obtenu en b_0 et b_1 , ainsi qu'il ressort des formules (23), par intermédiaire des fonctions λ et ψ , mais non par les valeurs de $f(u, x)$ sur la courbe (b). Ainsi, par exemple, la jonction sera en général tangentielle en b_1 si $x_1 = \omega$ puisqu'alors $\lambda(\omega) \equiv 0$. Pour les fonctions statistiques usuelles on appliquera les formules générales (30), (31) et (34).

35. Traitons encore un cas particulier qui peut rendre service pour déterminer, par exemple, le cercle de bénéficiaires de pensions de vieillesse dans un régime où l'attribution des prestations est subordonnée à la condition de ne pas dépasser certaines limites de revenus. Il s'agit donc de personnes âgées, l'âge x_0 signifiant alors la limite d'âge inférieure permettant l'octroi de pensions et x_1 se confondant avec l'âge terme ω . Admettons en première approximation une structure d'âge décroissant linéairement selon la formule (46). La nature des choses veut ici que l'échelle $s(x)$ des revenus moyens soit décroissante; nous la supposons également linéaire comme suit:

$$s(x) = 1 - C(x - x_0); \quad (68)$$

ainsi le domaine de définition (cf. graphique n^o 2) sera limité du côté droit par une droite décroissante entre les abscisses b_1 et b_0 . Par projection, on obtiendra ainsi une *distribution semi-parabolique particulière* $\bar{Q}(u; b_1, b_0, \beta)$ dont la fonction de fréquence est déterminée à l'aide de deux fonctions définies, l'une sur un intervalle (1): $0 \leq u < b_1$ et l'autre sur un intervalle (2): $b_1 \leq u \leq b_0$. Pour calculer les intégrales de projection, on passera par le changement de variable $x = \psi\left(\frac{u}{b_0}\right)$.

Tous calculs faits on obtient :

$$\bar{f}_1(u) = \frac{2\beta}{(b_0 - b_1)^2} \left[\frac{b_0^{2-\beta} - b_1^{2-\beta}}{2-\beta} - b_1 \frac{b_0^{1-\beta} - b_1^{1-\beta}}{1-\beta} \right] u^{\beta-1}, \quad (69)$$

$$\bar{f}_2(u) = \frac{2\beta}{(b_0 - b_1)^2} \left[b_0^{1-\beta} \left(\frac{b_0}{2-\beta} - \frac{b_1}{1-\beta} \right) u^{\beta-1} - \frac{u}{2-\beta} + \frac{b_1}{1-\beta} \right]. \quad (69')$$

La première branche appartient elle-même à la famille des distributions paraboliques, tandis que la deuxième est la différence entre une courbe parabolique et une droite croissante. Si nous admettons $\beta < 1$, nous obtenons une courbe d'ensemble constamment décroissante, avec point d'inflexion et raccord tangentiel en b_1 , comme le montre le graphique n° 4. Les expressions pour les fonctions statistiques \bar{F} , $\bar{\Phi}$ et \bar{M}_k sont analogues à celles de la distribution semi-hyperbolique \bar{P} examinée au n° 28. Donnons, par exemple, les trois premiers moments :

$$\left. \begin{aligned} \bar{M} &= \frac{\beta}{\beta + 1} \frac{2b_0 + b_1}{3} \\ \bar{M}_2 &= \frac{\beta}{\beta + 2} \frac{3b_0^2 + 2b_0b_1 + b_1^2}{6} \\ \bar{M}_3 &= \frac{\beta}{\beta + 3} \frac{4b_0^3 + 3b_0^2b_1 + 2b_0b_1^2 + b_1^3}{10} \end{aligned} \right\} \quad (70)$$

Si $\beta > 1$, on peut obtenir pour \bar{Q} des courbes unimodales pouvant, le cas échéant, être utilisées pour ajuster la répartition des revenus d'autres ensembles partiels de la population.

Chapitre III

L'actuariat économétrique de la sécurité sociale

§ 1. L'équilibre financier dans l'espace bio-économétrique (t, u, x)

36. Chaque régime de sécurité sociale est caractérisé par un certain nombre de *dispositions matérielles* définies dans la législation nationale. Il convient de les étudier de près avant d'aborder le problème de leurs répercussions financières. D'une manière schématique, ces dispositions légales peuvent être groupées ainsi: Risques couverts (événements assurés) — Personnes englobées — Droit aux prestations — Montant des prestations — Ressources — Organisation.

Les régimes modernes de sécurité sociale se fondent sur l'universalité des *risques couverts*. Selon leur nature on peut classer ces derniers en quatre groupes:

- 1^o Décès — invalidité — vieillesse
- 2^o Maladie — accidents (professionnels ou non)
- 3^o Chômage dû à des causes soit économiques, soit militaires
- 4^o Maternité — allocations familiales.

Quant aux *personnes assurées*, on tend de plus en plus à englober dans un régime national la population entière, ou tout au moins des classes importantes de celle-ci (les salariés par exemple). Pour avoir *droit aux prestations* il faut évidemment que l'événement assuré se soit réalisé; de plus, on fait souvent dépendre ce droit de certaines conditions accessoires de besoin, d'âge, d'état civil, de résidence, de nationalité, etc. Le *montant des prestations* dépend en général du risque couvert, de la durée des cotisations ainsi que du montant de celles-ci. Pour assurer l'équilibre financier, les régimes de sécurité sociale font en principe appel aux *ressources* suivantes: cotisations personnelles, cotisations patronales, subsides des pouvoirs publics, prélèvement sur des fonds et intérêts de ces derniers. Les cotisations personnelles sont généralement fixées selon les revenus, ce qui est souvent le cas également pour les prestations, soit directement, soit indirectement lorsque ces dernières dépendent du montant des

cotisations. Ainsi apparaît déjà le lien existant entre ces questions financières de sécurité sociale et les problèmes concernant la répartition des revenus.

37. L'équilibre financier d'un régime peut être obtenu par différentes méthodes; on se rendra compte de leur variété en consultant, par exemple, l'ouvrage de Féraud sur l'organisation financière des assurances sociales dans divers pays [12], ou encore les considérations plus générales que cet auteur a publiées dans une autre de ses études [11]. Malgré cette diversité, on constate que tous les mécanismes financiers envisagés tendent vers un but commun, à savoir le maintien, à longue échéance, de l'équilibre entre recettes et dépenses. Il est dès lors naturel d'essayer d'exprimer ce but commun par une équation commune à toutes ces méthodes. Pour écrire cette dernière nous choisissons d'abord les *symboles suivants pour les ressources*:

${}^cL_t(u, x) du dx$ le nombre des cotisants vivant à l'instant t , disposant d'un revenu u à $u + du$ et âgés de x à $x + dx$. Chacun paiera une cotisation de

$w_i(u, x) dt$ entre les instants t et $t + dt$; le montant de cette cotisation peut dépendre des variables u et x .

On peut admettre que la cotisation w contienne également la cotisation patronale; lorsque les pouvoirs publics participent financièrement au régime, leurs subsides réduisent d'autant la somme des cotisations personnelles et patronales.

En ce qui concerne les dépenses et pour fixer les idées, il est préférable d'envisager isolément *un seul risque, caractérisé par un indice m* . Considérons d'abord le risque vieillesse à l'intérieur d'un régime fondé à l'époque $t = 0$ et à participation obligatoire dès l'âge x_0 . Ainsi, nous pouvons admettre un renouvellement des effectifs limité aux deux droites $t = 0$ et $x = x_0$ du plan (t, x) . Si la pension de vieillesse est attribuée à un âge fixe x_1 , la seule connaissance des deux quantités d'observation t et x suffit pour déterminer *la durée n de cotisations*. En effet, l'âge d'entrée \check{x} est déterminé par $\check{x} = x - t$ avec la restriction $\check{x} \geq x_0$. Lorsqu'on veut faire correspondre à chaque n une échelle de prestations, il est indiqué de calculer la durée à une année près; en désignant par E le nombre entier d'années contenu dans une période donnée, on aura pour un pensionné d'âge x vivant à l'instant t l'échelle de prestation d'indice $n = E(x_1 - \check{x})$. En ce qui concerne

le revenu servant à calculer la prestation, de terme annuel r , il convient de faire remarquer qu'il ne coïncide pas avec le revenu du pensionné observé à l'instant t et à l'âge x . En effet, la pension de vieillesse est en général calculée sur la base d'un *revenu spécifique* \hat{u} , par exemple sur le salaire gagné à l'âge x_1 ou encore sur la moyenne d'une tranche de salaires précédant l'âge x_1 . Pour les risques autres que la vieillesse, les quantités n et \hat{u} peuvent également jouer un rôle analogue. Remarquons cependant que dans le cas des pensions de survivants et d'invalidité les variables d'observation t et x ne déterminent plus d'une manière univoque la durée n . Pour y arriver il faudrait alors introduire soit une nouvelle variable continue n , soit des hypothèses simplificatrices (par exemple: âges moyens \bar{x}_1 de réalisation du risque, allocations uniques). En adoptant ici l'une de ces dernières, nous pouvons choisir *les symboles suivants quant aux prestations*:

${}^m L_t(\hat{u}, x) d\hat{u} dx$ le nombre des bénéficiaires concernant le risque d'indice m , vivant à l'instant t , dont la prestation a été calculée à l'aide d'un revenu spécifique \hat{u} à $\hat{u} + d\hat{u}$ et qui sont âgés de x à $x + dx$. Chacun touchera une prestation de

${}_n r(\hat{u}) dt$ entre les instants t et $t + dt$; ce montant dépendra en plus du risque envisagé, du revenu spécifique \hat{u} ainsi que de la durée n de cotisations supposée déterminée par t et x .

Il est clair qu'un mécanisme financier se distingue d'un autre par une répartition différente, dans l'espace (t, u, x) , des ressources et des dépenses entre les assurés. Or, une équation commune à tous les mécanismes doit être valable quelle que soit cette répartition; pour prétendre à cette généralité notre équation doit donc s'étendre à l'espace entier. Ainsi, *l'équation générale d'équilibre financier dans l'espace (t, u, x)* peut s'exprimer, pour un risque donné, par la formule ci-après. On aura pris soin de définir au préalable une loi de capitalisation $(1 + i)^t = e^{\delta t}$ s'appliquant à l'unité monétaire, soit à l'aide du taux d'intérêt annuel i , soit avec le taux instantané équivalent δ ; en introduisant le facteur d'escompte habituel $v = (1 + i)^{-1} = e^{-\delta}$, on pourra finalement écrire:

$$\int_0^{\infty} e^{-\delta t} dt \int_{x_0}^{\omega} dx \left[\int_0^{\infty} {}^c L_t(u, x) w_t(u, x) du - \int_0^{\infty} {}^m L_t(\hat{u}, x) {}_n r(\hat{u}) d\hat{u} \right] = 0. \quad (71)$$

Suivant le m envisagé on précisera les limites d'intégration en x ; pour le risque vieillesse par exemple, on ira de x_0 à x_1 pour les cotisants cL et de x_1 à ω pour les pensionnés mL .

38. De l'équation (71) nous déduirons au numéro suivant tous les mécanismes financiers usuels. Avant de rejoindre ainsi les théories de l'actuariat classique, effectuons d'abord les intégrations en u et en \hat{u} , ceci par *l'introduction systématique des distributions de revenus*. Les symboles adoptés pour les effectifs de cotisants et de bénéficiaires peuvent être définis directement à partir des fonctions de fréquence introduites au n° 4. Nous désignons par cL_t l'ensemble des cotisants vivant à l'instant t et par mL_t l'ensemble des bénéficiaires concernant le risque m ; nous admettons en outre que chacun de ces ensembles ait ses propres distributions caractérisées, soit par l'indice c , soit par m . Ainsi l'on pourra écrire les deux couples de formules de définition que voici:

$${}^cL_t(x) = {}^cL_t {}^c\lambda_t(x), \quad (72)$$

$${}^mL_t(x) = {}^mL_t {}^m\lambda_t(x); \quad (72')$$

$${}^cL_t(u, x) = {}^cL_t {}^c\varphi_t(u, x) = {}^cL_t(x) {}^c f_t(u, x), \quad (73)$$

$${}^mL_t(\hat{u}, x) = {}^mL_t {}^m\varphi_t(\hat{u}, x) = {}^mL_t(x) {}^m f_t(\hat{u}, x). \quad (73')$$

Portons maintenant les substitutions indiquées par (73) et (73') dans l'équation (71) et effectuons l'intégration en u , en posant:

$$\bar{w}_t(x) = \int_0^{\infty} w_t(u, x) {}^c f_t(u, x) du, \quad (74)$$

$$\bar{r}_t(x) = \int_0^{\infty} r_t(\hat{u}) {}^m f_t(\hat{u}, x) d\hat{u}. \quad (74')$$

La première de ces quantités représente *la cotisation moyenne d'un ensemble* ${}^cL_t(x)$ *de cotisants*. Puisque nous avons admis que n est complètement déterminé par la seule connaissance de t et x , la deuxième de ces quantités définit bien *la prestation moyenne d'un ensemble* ${}^mL_t(x)$ *de bénéficiaires*. C'est précisément le calcul des valeurs moyennes (74) et (74') qui constitue un des problèmes essentiels de ce que nous appelons l'actuariat économétrique.

Deux questions fondamentales se posent. La première est de savoir comment on peut *déduire les distributions de fréquence ${}^m f$ de celles de fréquence ${}^c f$* , ces dernières coïncidant en général avec les distributions de revenus à fonctions de fréquence $f(u, x)$ envisagées au n° 10 (cf. hypothèse *B*). Un exemple simple est fourni par le cas des pensions de vieillesse calculées sur le revenu précédant immédiatement l'attribution des prestations à l'âge x_1 ; ici on supposera ${}^m f(\hat{u}, x) = {}^c f(u, x_1)$. Nous verrons un exemple plus compliqué au § 3 ci-après. La deuxième question fondamentale se rapporte à la nature de *la dépendance fonctionnelle entre w et u , respectivement entre r et \hat{u}* . Quelle qu'en soit la nature, l'application des formules (74) et (74') permettra toujours le calcul direct de \bar{w} et \bar{r} . Cependant, si ce lien est relativement simple, on n'aura pas besoin de faire intervenir chaque valeur des fonctions de fréquence pour effectuer ce calcul. On arrivera souvent à déduire de (74) et (74') des expressions analytiques qui ne feront appel qu'à quelques valeurs particulières des fonctions statistiques usuelles, par exemple M_k , F et Φ . Le cas classique est celui où w et r sont liés linéairement à la variable du revenu. En posant par exemple $r(\hat{u}) = \rho \hat{u} + k$, on trouvera $\bar{r} = \rho \hat{M} + k$ où \hat{M} est la moyenne de la distribution des revenus spécifiques. Encore plus simple est le cas de la proportionnalité où $k = 0$ ou encore celui où $\rho = 0$, c'est-à-dire $r = k$. Ce sont précisément ces formes élémentaires de dépendance linéaire qui ont été utilisées à l'origine dans les systèmes de sécurité sociale. Ainsi les actuaires pouvaient se passer des distributions de revenus, les moyennes étant alors suffisantes. La structure des systèmes modernes de sécurité sociale est plus complexe et l'introduction des distributions de revenus est devenue de ce fait inéluctable, comme nous allons le montrer aux paragraphes suivants.

39. Une fois le calcul des moyennes effectué, (71) se réduira alors à une *équation d'équilibre financier dans le plan biométrique (t, x)* :

$$\int_0^{\infty} e^{-\delta t} dt \int_{x_0}^{\omega} [{}^c L_t(x) \bar{w}_t(x) - {}^m L_t(x) \bar{r}_t(x)] dx = 0. \quad (75)$$

Nous avons obtenu ainsi une équation d'équilibre analogue à celle de Kaan [21]; la formule (75) en diffère cependant par les quatre points suivants: 1° Application de la méthode continue; 2° Introduction dans l'équation de départ des variables «d'observation» t et x au lieu des

variables «d'entrée» \bar{t} et \bar{x} ; 3° Suppression de la variable n (hypothèse simplificatrice de renouvellement); 4° Présentation des dépenses à l'aide de deux facteurs L et r à la place du produit A de ces facteurs.

De l'équation (75) on peut facilement déduire les mécanismes financiers classiques; vu les limites des variables adoptées, l'intégration s'effectue dans un domaine rectangulaire ouvert du côté droit. L'équilibre global d'un régime est évidemment garanti s'il l'est déjà séparément dans tous les domaines partiels du plan. Nous pouvons donc former, en suivant la méthode de Kaan, des *groupes financièrement autonomes dans le plan* (t, x) et supposer qu'à l'intérieur de chacun des groupes ainsi constitués la cotisation w est constante; de cette manière on peut admettre que les prestations sont données et que les cotisations sont les inconnues. Le choix des domaines partiels doit cependant garantir le fonctionnement pratique du système choisi, en particulier, il faut qu'il y ait à part les bénéficiaires toujours des cotisants. L'autonomie financière peut être envisagée théoriquement même le long de certaines lignes du plan et même en des points isolés. En choisissant un point, les deux variables t et x sont fixes; en choisissant une ligne, une seule des variables est indépendante. Ce n'est qu'en considérant des domaines à deux dimensions que les deux variables t et x sont simultanément libres. Ainsi on pourra construire une infinité de systèmes financiers, même en ne choisissant que des lignes droites pour définir les groupes autonomes. On arrive aux *systèmes financiers classiques* par la schématisation suivante où nous noterons l'expression de (75) écrite entre crochets par le symbole $[\sim]$:

1° *La prime naturelle* s'obtient en prenant aussi bien t que x fixes. Pour trouver $\bar{w}_t(x)$ il suffit d'annuler $[\sim]$.

2° *La prime instantanée de répartition* consiste à choisir t fixe, c'est-à-dire de former des groupes autonomes sur les droites $t = \text{constant}$. On trouve la prime de répartition \bar{w}_t en annulant dans (75) $\int_{x_0}^{\omega} [\sim] dx$.

3° *La prime individuelle (capitalisation individuelle)* se déduit de (75) en réalisant l'autonomie financière le long des lignes de vie dont chacune est caractérisée par un âge d'entrée $\bar{x} \geq x_0$ constant; un fois \bar{x} fixé, x ne varie plus qu'en fonction de t . Pour retrouver les formules classiques de l'actuariat, il suffit de rem-

placer dans (75) x par $t + \bar{x}$ et de poser $\int_0^{\omega - \bar{x}} e^{-\delta t} [\sim] dt = 0$;

lorsque $x - t < x_0$ on appliquera la formule concernant la ligne de vie $x = t + x_0$. Vu notre hypothèse de renouvellement, les L évolueront selon les ordres d'extinction adoptés. Si l'on suppose \bar{w} indépendant de t on trouvera ainsi la prime individuelle $\bar{w}(\bar{x})$.

4° *La prime moyenne générale* est la conséquence directe de l'équation (75) étendue à tout le plan et dans laquelle il suffit de supposer \bar{w} indépendant de t et de x .

Le facteur $e^{-\delta t}$ n'intervenant pas dans les deux premiers systèmes, ceux-ci appartiennent aux systèmes de répartition. La présence de ce facteur dans les équations relatives aux systèmes 3° et 4° détermine, en revanche, une part de capitalisation. — Dans la pratique actuarielle, on trouvera encore d'autres systèmes financiers en envisageant dans le plan (t, x) des domaines partiels simples. Ainsi l'on obtient *le système de répartition par périodes* en découpant le domaine complet par des droites parallèles à l'axe des x ; si $t = t_1$ et $t = t_2$ sont deux droites successives on définira une prime \bar{w}_{t_1, t_2} et en particulier la prime de répartition annuelle lorsque $t_2 = t_1 + 1$. D'une manière semblable on arrive à *la prime moyenne par groupe de générations* en décomposant le plan (t, x) en bandes limitées par des lignes de vie parallèles. On écrira (75) en \bar{x} et on obtiendra la prime moyenne $\bar{w}(\bar{x}_1, \bar{x}_2)$, où \bar{x}_1 et \bar{x}_2 caractérisent les deux lignes de vie limites. Dès que $x - t \leq x_0$ on aura la prime $\bar{w}(x_0)$ pour les générations entrant dans le régime après $t = 0$, prime identique à la prime individuelle pour $\bar{x} = x_0$. — Kaan a encore défini d'autres systèmes financiers en formant d'autres groupes autonomes à l'aide de la variable n (durée de cotisations) qu'il a choisie alternativement comme élément fixe ou variable. Ces systèmes sont dans le cas général d'un intérêt purement théorique, mais coïncident ici avec les systèmes exposés ci-dessus, vu que nous avons limité le renouvellement sur les deux droites $t = 0$ et $x = x_0$.

Pour les calculs financiers de la sécurité sociale on a de plus en plus recours à *la méthode des budgets annuels*. Elle découle également de l'équation (75) si l'on met en évidence la somme annuelle Ω_τ des recettes et la somme annuelle A_τ des dépenses d'une année débutant à l'instant τ ; il suffit de poser:

$$\Omega_\tau = \int_\tau^{\tau+1} e^{-\delta(t-\tau)} dt \int_{x_0}^{\omega} {}^c L_t(x) \bar{w}_t(x) dx, \quad (76)$$

et

$$A_\tau = \int_\tau^{\tau+1} e^{-\delta(t-\tau)} dt \int_{x_0}^{\omega} {}^m L_t(x) \bar{r}_t(x) dx. \quad (76')$$

On commence en général par calculer ces quantités et l'équation d'équilibre (75) s'écrira alors de la manière suivante:

$$\sum_0^\infty (\Omega_\tau - A_\tau) e^{-\delta\tau} = 0. \quad (76'')$$

On voit par (76) et (76') l'importance que revêt également dans la méthode des budgets annuels le calcul des valeurs moyennes des cotisations et des prestations. Nous examinerons dans les paragraphes suivants quelques cas concrets.

Lorsque le système financier adopté comprend une part de capitalisation, il se constituera une *réserve mathématique collective* que l'on calculera, au début d'une année commençant à l'instant t , selon la formule ci-après:

$$V_t = \sum_0^{t-1} (\Omega_\tau - A_\tau) e^{\delta(t-\tau)}. \quad (77)$$

Dans de tels systèmes, les intérêts de cette réserve devront couvrir à partir d'un certain instant t , mais avant tout à l'état stationnaire, l'excédent des dépenses sur les autres ressources.

40. La formation de groupes financièrement autonomes peut s'effectuer non seulement dans le plan (t, x) , mais déjà dans l'espace (t, u, x) . A l'intérieur de chacun de ces groupes s'étendant à une portion de l'espace on réalise une *solidarité financière* entre ses membres. En effet, ceux-ci paient tous la même cotisation bien que présentant des risques financiers fort divers. C'est dans le système de la prime individuelle que l'on réduit cette solidarité au strict minimum nécessaire à tout mécanisme d'assurance. En effet, on demande alors à chaque assuré la prime exacte correspondant au risque présumé, ce dernier étant déterminé avant tout par l'âge d'entrée ainsi que par le montant de l'assurance contractée. Il faut donc en particulier que le rapport $\frac{r}{w}$

soit ici constant par rapport au revenu u , ce qui sera le cas si r et w sont proportionnels au même u . De la sorte, le revenu moyen correspondant s'élimine de l'équation (75) et cette dernière se ramène à une relation entre taux de cotisations et taux de prestations; c'est ainsi que se présentent en général les contrats de l'assurance privée. Si l'on ne considère que les groupes autonomes dans le plan (t, x) , on réalise à l'intérieur de ceux-ci une *solidarité biométrique* caractérisée par une prime moyenne \bar{w} identique pour des risques biométriques différents. Or, en procédant de cette façon, on néglige tout l'aspect économétrique du problème qui découle de l'introduction de la variable u . La prise en considération de cette dernière n'engendre, il est vrai, aucun système financier nouveau; c'est la variable t qui les détermine. En revanche, si l'on forme des groupes autonomes dans l'espace (t, u, x) pour lesquels $u_1 \leq u \leq u_2$, on crée dans chacune de ces classes une *solidarité économétrique* qui peut évidemment couvrir toute l'étendue de la variable u . C'est en particulier le rapport $\frac{r(\hat{u})}{\hat{u}}$ qui peut

varier à l'intérieur du groupe considéré, de sorte que les avantages retirés ne sont pas toujours proportionnels aux cotisations lorsque ces dernières sont elles-mêmes proportionnelles aux revenus.

Le but essentiel du présent chapitre était de montrer que *le problème du calcul des prestations moyennes*, dont nous parlerons plus en détail aux paragraphes suivants, se pose d'une manière très générale et réclame une solution avant même d'envisager la question des systèmes financiers qui peut être étudiée sur le seul plan biométrique. En effet, les actuaires partent en général directement d'une équation analogue à (75) ou de l'une de ses formes particulières. De cette façon ils admettent tacitement que le calcul des moyennes est déjà effectué ou, plus souvent encore, que r et w sont proportionnels aux revenus. Ce n'est d'ailleurs que dans ce dernier cas qu'on peut déduire de (75) un système financier méritant vraiment l'épithète d'individuel. Sans cette proportionnalité on réaliserait, même à l'intérieur d'un système de capitalisation dite individuelle, une solidarité économétrique; les symboles des moyennes \bar{r} et \bar{w} subsistant dans les formules confirment la chose.

Nous avons aussi tenu à montrer que *la méthode de Kaan* pouvait se généraliser et devait même l'être afin de ne pas rester incomplète. Les avis sont très partagés au sujet des mérites de cette méthode.

Il est indéniable qu'elle réalise une synthèse ingénieuse de tous les systèmes financiers. Son application n'a cependant pas conduit à d'autres mécanismes pratiques que ceux qui étaient déjà connus des actuaires; le principal mérite de la méthode est peut-être d'avoir précisément démontré que l'on avait épuisé la multiplicité des systèmes pratiques. Lorsqu'on ne considère que les domaines limités d'une manière rectiligne, on peut dégager de l'équation générale, faisant intervenir cette fois comme variable supplémentaire la durée n de cotisation, une quinzaine de mécanismes financiers, dont la moitié d'entre eux sont pratiquement inapplicables dans un régime de pensions invalidité-vieillesse-survivants. La substance de la méthode de Kaan apparaît plus clairement si on la transpose sur le terrain géométrique, comme nous l'avons préconisé au n° 39. On arrive ainsi du même coup à la débarrasser d'un appareil de calculs par trop laborieux à notre avis.

§ 2. La rente transitoire moyenne de l'AVS

41. Lorsque la durée n de versement des cotisations est inférieure à une année, le régime suisse d'assurance-vieillesse et survivants (AVS) prévoit des rentes dites transitoires. Le montant de leurs termes annuels dépend, d'une part, du genre de rente (rentes de vieillesse simple et pour couples, de veuves, d'orphelins simples et doubles) et, d'autre part, de la région (urbaines, mi-urbaines et rurales) où vit le bénéficiaire. Associons l'indice m au genre de rente et l'indice q au critère régional. Le droit à ces rentes est conditionnel puisqu'en effet, les personnes dont le revenu annuel u dépasse *les limites légales de revenus* ${}^m_q u_1$ (par exemple $u_1 = 3200$ francs pour les rentes de vieillesse de couples dans les régions urbaines) sont exclues du bénéfice des prestations. Aux termes actuels de la loi il faut prendre en compte le revenu u entier de la personne considérée. Il est question de reviser ces dispositions en augmentant les limites u_1 de 25 %, d'une part, et en ne prenant en compte que les trois quarts du revenu u , d'autre part. Pour cette raison, considérons ici le cas général où l'on ne tient compte pour l'attribution des rentes que d'une proportion $v \leq 1$ des revenus personnels.

La loi énumère encore *les termes annuels maxima* ${}^m_q r_0$ *des prestations* (par exemple $r_0 = 1200$ francs pour les rentes de vieillesse pour

couples, régions urbaines), mais prescrit que ces montants sont réduits dans la mesure où, avec le revenu annuel u (après revision: avec la part νu), ils dépassent les limites u_1 . Les dispositions légales font donc dépendre du revenu u , du moins sur certains intervalles, le terme annuel ${}^m_q r(u)$ de la rente. Ainsi, la loi impose indirectement les deux inégalités fondamentales suivantes où, pour simplifier les notations, nous négligeons les indices m et q :

$$r(u) \leq r_0, \quad (78)$$

et

$$\nu u + r(u) \leq u_1. \quad (78')$$

On déduit de ces inégalités, d'une part, le revenu au-dessous duquel on a toujours $r(u) = r_0$, revenu que nous appelons *limite de réduction* u_0 , et, d'autre part, le revenu au-dessus duquel il vient $r(u) = 0$; ce dernier représente *la limite effective de revenus* u'_1 . Ainsi nous pouvons écrire:

$$u_0 = \frac{u_1 - r_0}{\nu}, \quad (79)$$

et

$$u'_1 = \frac{u_1}{\nu}. \quad (79')$$

Les données numériques indiquées plus haut pour les rentes de couples urbaines conduisent, en supposant la revision légale effectuée, aux chiffres que voici: $u_1 = 1,25 \times 3200 = 4000$, $\nu = 0,75$, $u_0 = 3733$ et $u'_1 = 5333$.

Nous sommes maintenant en mesure d'exprimer la prestation annuelle en fonction de u sur tout l'intervalle de variation $0 \leq u \leq \infty$, comme suit:

$$\left. \begin{array}{l} 0 \leq u \leq u_0: \quad r(u) = r_0 \\ u_0 \leq u \leq u'_1: \quad r(u) = u_1 - \nu u = r_0 - \nu(u - u_0) \\ u'_1 \leq u < \infty: \quad r(u) = 0. \end{array} \right\} \quad (80)$$

La rente $r(u)$ est une fonction polygonale; en effet, on obtient graphiquement un polygone à trois segments, dont le dernier se confond avec l'axe des u . Afin de pouvoir apprécier l'effet de nivellement produit sur les revenus par les rentes transitoires, il est instructif de dessiner également le graphique représentant la fonction $u + r(u)$.

42. Pour évaluer les dépenses annuelles on appliquera la formule (76'). Il faudra d'abord calculer à l'instant t le nombre de bénéficiaires en partant de celui de toutes les personnes d'âge x ($0 < u < \infty$) pour lesquelles le risque envisagé s'est déjà réalisé. Si l'on désigne ce dernier nombre par $L_t(x)$, toujours en négligeant les indices m et q , le nombre de personnes dont le revenu est inférieur à u'_1 s'obtient en intégrant (73') de 0 à u'_1 ; ainsi il vient:

$$L_t^*(x) = F(u'_1, x) L_t(x), \quad (81)$$

où F est la fonction cumulative définie par (27); si possible, on prendra un F spécifique pour chaque couple de paramètres m et q .

Si l'on calcule la rente moyenne transitoire selon (74'), il ne faut pas perdre de vue que la fonction $f(\hat{u}, x)$ à appliquer ici donne la distribution des bénéficiaires d'âge x selon leur propre revenu, de sorte que $f_t(\hat{u}, x) = f(u, x)$. La rente moyenne variera donc ici avec l'âge observé; nous aurons ainsi:

$$\bar{r}(x) = \int_0^{\infty} r(u) f(u, x) du. \quad (82)$$

Décomposons l'intervalle d'intégration selon les trois intervalles de (80) et substituons dans (82) à $r(u)$ ses valeurs polygonales correspondantes. Si l'on introduit les fonctions cumulatives $F(u, x)$ ainsi que les sommes réduites $\Phi(u, x)$ l'on obtient finalement:

$$\bar{r}(x) = r_0 F(u_0, x) + u_1 [F(u'_1, x) - F(u_0, x)] - v [\Phi(u'_1, x) - \Phi(u_0, x)]. \quad (82')$$

En calculant $\bar{r}(x)$ d'après (82'), il faut alors, pour évaluer A_t , multiplier \bar{r} avec les effectifs $L_t(x)$. Si l'on veut se servir de $L_t^*(x)$ calculé selon (81), il faut évidemment prendre comme moyenne:

$$\bar{\bar{r}}(x) = \frac{\bar{r}(x)}{F(u'_1, x)}. \quad (82'')$$

Il est donc possible de calculer $\bar{r}(x)$ analytiquement, pourvu que les expressions définissant F et Φ soient intégrables. Tel est bien le cas, par exemple, lorsqu'on utilise pour les calculs numériques relatifs à chaque âge x une distribution parabolique $Q(u; b_x, \beta)$ avec $\beta < 1$ (cf. n° 34). On sait alors que la distribution globale de tous les rentiers transitoires caractérisés par un même couple d'indices m, q est une distribution semi-parabolique \bar{Q} . Lorsqu'on considère d'emblée la distribution globale, on peut d'ailleurs utiliser cette distribution \bar{Q}

pour effectuer les calculs en première approximation, comme nous l'avons suggéré au n° 35. On déterminera alors les trois paramètres b_1 , b_0 et β dans le cas particulier du n° 35 par trois conditions en s'imposant, par exemple, \bar{M} , $\bar{F}(u_0)$ et $\bar{F}(u'_1)$, quantités qui peuvent souvent être tirées des observations statistiques. — On pourrait également prendre en considération ici les distributions $P(u; a_x, \alpha)$ ou $G(u; a_x, \gamma_x, \varepsilon)$ avec $\varepsilon < 1$.

§ 3. La rente ordinaire moyenne de l'AVS

43. Dès que l'on a pour la durée de versement des cotisations: $n \geq 1$, la loi d'AVS suisse accorde des rentes dites ordinaires. Si l'événement assuré (vieillesse, décès) se réalise, ces prestations sont attribuées sans aucune condition supplémentaire. Il faut distinguer ici les mêmes *genres de rentes d'indice m* que pour les rentes transitoires (cf. n° 41). Le montant des rentes ne dépend pas seulement de m mais encore de *la durée n de cotisations*, définie à une année près comme au n° 37 par $n = E(x_1 - \bar{x})$. Pour la vieillesse, x_1 signifie la limite d'âge (= 65 ans) à laquelle sont attribuées les rentes; du fait qu'elles prennent toujours naissance soit un 1^{er} janvier, soit un 1^{er} juillet, la loi prévoit pratiquement une limite variable entre 65,0 et 65,5 ans. Pour les survivants, x_1 représente l'âge de décès — à condition que ce dernier survienne avant la limite vieillesse — du cotisant entrant en ligne de compte. L'âge \bar{x} (≥ 20 ans) est celui de l'assujettissement de la génération considérée. Ainsi, on a pour $1 \leq n \leq 19$ les 19 échelles de rentes partielles et pour $n \geq 20$ l'échelle des rentes complètes.

Le terme annuel d'une rente dépend encore de *la cotisation annuelle moyenne \hat{w}* que le cotisant aura réalisée depuis son âge d'entrée \bar{x} jusqu'à l'arrivée de l'événement assuré à l'âge x_1 . Nous calculerons ainsi \hat{w} comme moyenne arithmétique des cotisations w proportionnelles aux revenus du travail et payées pendant la carrière $\bar{x} \leq x \leq x_1$; rappelons que la loi fixe un taux de cotisation $\pi = 0,04$ comprenant la cotisation personnelle et patronale. Si l'on associe le revenu variable u à un âge x quelconque et u' à l'âge de référence $x_0 = 20$ ans, nous aurons pour les cotisations individuelles correspondantes:

$$w = \pi u \tag{83}$$

et

$$w' = \pi u'. \tag{83'}$$

Or, nous avons précisément admis, au n° 11, par l'interprétation biométrique de la relation (13) que les revenus individuels se transformaient par $u = s(x) u'$, de sorte que l'on a en vertu de (83) et (83') :

$$w = s(x) w'. \quad (83'')$$

Ainsi la moyenne arithmétique \hat{w} cherchée se calculera d'après la formule évidente :

$$\hat{w} = \frac{w'}{x_1 - \overset{*}{x}} \int_{\overset{*}{x}}^{x_1} s(x) dx = s(\overset{*}{x}, x_1) w', \quad (84)$$

où le symbole $s(\overset{*}{x}, x_1)$ apparaît comme moyenne arithmétique de l'échelle $s(x)$.

Connaissant les trois éléments déterminants m , n et \hat{w} , nous pouvons calculer maintenant un *terme annuel de la rente ordinaire* ${}_n r(\hat{w})$. Puisque la loi considère la rente de vieillesse simple complète ($20 \leq n \leq 45$) comme l'élément technique de base, nous partons de cette dernière. Les dispositions légales donnent pour le terme annuel de la rente envisagée les montants suivants selon les quatre intervalles en \hat{w} ci-après (chiffres en francs) :

$$\left. \begin{array}{l} 0 \leq \hat{w} \leq 30 : \quad {}_{20}r(\hat{w}) = 480 \\ 30 \leq \hat{w} \leq 150 : \quad {}_{20}r(\hat{w}) = 300 + 6\hat{w} \\ 150 \leq \hat{w} \leq 300 : \quad {}_{20}r(\hat{w}) = 900 + 2\hat{w} \\ 300 \leq \hat{w} < \infty : \quad {}_{20}r(\hat{w}) = 1500. \end{array} \right\} \quad (85)$$

Les rentes de vieillesse simples partielles ($1 \leq n \leq 19$) se calculent d'après les règles suivantes :

$$\left. \begin{array}{l} 0 \leq \hat{w} \leq 75 : \quad {}_n r(\hat{w}) = {}_{20}r(\hat{w}) \\ 75 \leq \hat{w} < \infty : \quad {}_n r(\hat{w}) = 750 + \frac{n}{20} [{}_{20}r(\hat{w}) - 750]. \end{array} \right\} \quad (85')$$

On déduit, en principe, les montants annuels des autres genres de rentes en multipliant le résultat obtenu à l'aide de (85) et (85') par les facteurs que voici : 1,6 pour la rente de couple ; 0,5 à 0,9 pour les rentes de veuves, suivant l'âge de la veuve au moment du décès du conjoint ; 0,3 pour la rente d'orphelin simple et 0,45 pour la rente

d'orphelin double, la réduction selon (85') ne s'appliquant cependant pas à ces deux derniers genres de rentes. — On voit immédiatement que ces rentes s'expriment dans chaque intervalle, et quels que soient les indices n ou m , par une fonction linéaire en \hat{w} . Toutes ces formules se réduisent donc au type suivant :

$$r(\hat{w}) = A_j + B_j \hat{w}, \quad (85'')$$

relation où les coefficients $A_j > 0$ et $B_j \geq 0$ dépendent également des paramètres m et n ; (85'') représente l'équation d'un segment de droite correspondant à un intervalle numéro j de l'axe des \hat{w} . On a cinq intervalles j pour les rentes partielles et quatre pour les rentes complètes. Les équations (85'') définissent ainsi, pour un couple d'indices m et n donné, un *contour polygonal*, concave vers le haut à la fin du premier intervalle et concave vers le bas aux jonctions des autres. Nous ne sommes donc plus en présence d'une linéarité uniforme. On consultera à ce sujet les graphiques de la publication [29].

44. Pour estimer les dépenses annuelles A_t selon (76'), nous supposons effectués les calculs démographiques menant aux effectifs ${}^m L_t(x)$. L'évaluation de la rente ordinaire moyenne \bar{r} se fait en principe d'après (74') en écrivant d'abord pour une rente de vieillesse (simple ou pour couple) :

$$\bar{r}_t(x) = \int_0^{\infty} {}_n r(\hat{w}) f_t(\hat{w}, x) d\hat{w}; \quad (86)$$

$r(\hat{w})$ est tiré de (85'') et la *distribution des pensionnés selon les cotisations annuelles moyennes* peut être déduite d'une distribution de revenus des cotisants comme nous l'avons annoncé au n° 38. En effet, il suffit de partir de la fonction fondamentale $f(u', x_0) = g(u')$ et d'effectuer les deux changements de variables successifs (84) et (83'), c'est-à-dire

$$\hat{w} = \pi s(x^*, x_1) u'. \quad (87)$$

Nous calculons la transformée de $g(u')$ et nous supposons qu'elle coïncide avec la fonction de fréquence $f_t(\hat{w}, x)$ relative à la distribution des cotisations annuelles moyennes réalisées entre $\hat{x} = x - t$ et x_1 . On introduit ainsi une nouvelle *hypothèse C* qui est analogue à l'hypothèse *A* du n° 9 et paraît aussi plausible que celle-ci du fait de l'interdépendance des revenus individuels en passant d'un âge x à un autre. Vu cette dépendance nous ne sommes plus en présence

du problème fondamental du calcul des probabilités concernant la distribution d'une somme de variables (ou de leur moyenne) chacune d'elles étant distribuée indépendamment l'une de l'autre. Il vient donc :

$$f_t(\hat{w}, x) = f(\hat{w}; \hat{x}, x_1) = \frac{1}{\pi s(\hat{x}, x_1)} g\left[\frac{\hat{w}}{\pi s(\hat{x}, x_1)}\right]. \quad (88)$$

Pour les rentes de survivants nous pouvons appliquer en principe le même procédé, x_1 signifiant alors l'âge de décès; ce dernier pouvant varier pour un t et x donnés, la première équation de (88) ne s'applique dans ce cas qu'en précisant pour f_t la durée n . Suivant les nécessités, on fera correspondre à chaque risque m sa propre fonction ${}^m s(x)$, surtout lorsque m dépend de l'état civil; n étant déterminé par \hat{x} et x_1 , on obtiendra également une distribution propre à chaque n . Ainsi, l'on aura bien calculé toutes les fonctions de fréquence dont on a besoin.

Le calcul de la *rente ordinaire moyenne* \bar{r} ne présentera dès lors aucune difficulté. Il suffit de substituer les expressions (85'') et (88) dans (86) et de décomposer l'intervalle d'intégration dans ses intervalles partiels j . En désignant par $\Delta_j F$ et $\Delta_j \Phi$ les accroissements des fonctions $F(\hat{w})$ et $\Phi(\hat{w})$ sur l'intervalle j , on aura :

$$\bar{r} = \sum_j [A_j \Delta_j F + B_j \Delta_j \Phi]. \quad (89)$$

Lorsque les expressions définissant F et Φ sont intégrables, on peut donc calculer analytiquement la rente moyenne. Ceci se vérifie en particulier pour la *distribution semi-normale* $G(u; a_x, \gamma_x, \varepsilon)$ dont nous avons préconisé [22] l'emploi pour les calculs concernant l'équilibre financier de l'AVS [29]. Ce cas particulier est d'ailleurs à l'origine de la méthode générale exposée au chap. I. Krämer [25] a proposé une autre distribution du type de Pearson, mais il suggère de calculer \bar{r} à l'aide d'une distribution globale unique, ce qui ne constitue qu'une première approximation pour la résolution du problème. Ce n'est que le dépouillement des comptes individuels qui indiquera en définitive à quelles fonctions il faut avoir recours pour ajuster les distributions observées.

45. Le calcul de la rente ordinaire moyenne soulève encore d'autres problèmes. Indiquons d'abord celui de la *comparaison de la rente moyenne* \bar{r} avec la *rente individuelle* $r(\bar{w})$, correspondant à la valeur moyenne \bar{w} des cotisations individuelles moyennes \hat{w} . Cette valeur

moyenne \bar{w} se calcule à l'aide de (88) et tombe à l'intérieur d'un intervalle \bar{j} déterminé de sorte que $r(\bar{w}) = A_{\bar{j}} + B_{\bar{j}}\bar{w}$. On peut calculer encore les valeurs moyennes partielles \bar{w}_j pour chaque intervalle j , de même que les quantités $\bar{r}_j = A_j + B_j\bar{w}_j$ telles que $\bar{r} = \sum_j \bar{r}_j \Delta_j F = \bar{A} + \bar{B}\bar{w}$, ceci en posant $\bar{A} = \sum_j A_j \Delta_j F$ et $\bar{B} = \sum_j B_j \frac{\bar{w}_j}{\bar{w}} \Delta_j F$. On trouvera ainsi, en éliminant \bar{w} des équations donnant $r(\bar{w})$ et \bar{r} :

$$\bar{r} = \frac{\bar{B}}{B_{\bar{j}}} [r(\bar{w}) - A_{\bar{j}}] + \bar{A}. \quad (90)$$

Une variation de \bar{w} indique une variation de f , donc de \bar{A} et \bar{B} , de sorte que (90) représente une courbe $\bar{r}(\bar{w})$. Lorsque \bar{w} déterminera un numéro \bar{j} d'intervalle suffisamment élevé, l'on aura toujours $\bar{r} < r(\bar{w})$, ce qui ressort directement de la comparaison de l'équation donnant \bar{r} avec l'identité: $r(\bar{w}) = r(\bar{w}) \sum_j \Delta_j F$ (concavité du polygone tournée vers le bas).

Il est naturel de se demander à quelle rente pourraient donner lieu les cotisations $w = s(x) w'$ payées entre \hat{x} et x_1 si l'on appliquait *le principe de l'équivalence actuarielle* qui réside dans l'équation de la prime individuelle (cf. n° 39), où r serait cette fois l'inconnue et w donné. Désignons d'un côté par a_x^* la valeur actuelle, pour un assuré d'âge \hat{x} , de tous les genres de rentes ordinaires de l'AVS auxquelles il pourrait avoir droit, en posant la rente de vieillesse simple = 1; d'un autre côté, appelons a_x^* la valeur actuelle des cotisations de cet assuré en supposant que la cotisation annuelle moyenne \hat{w} , calculée selon (84), soit = 1. A propos de ces valeurs actuelles nous renvoyons à la collection de tables: «Valeurs actuelles des cotisations et des rentes AVS» que vient de publier l'Office fédéral des assurances sociales. Les formules (94) et (95) ci-après donnent d'ailleurs une première idée à ce sujet (voir aussi la remarque suivant la première de ces formules). De cette manière, on pourra déterminer un taux de rente q_x^* comme suit:

$$a_x^* = q_x^* a_x^*; \quad (91)$$

q_x^* représente ainsi la rente de vieillesse simple acquise par le versement de cotisations dont la moyenne $\hat{w} = 1$. En réalisant une moyenne \hat{w} , l'assuré obtiendra la rente $r(\hat{w}, \hat{x})$ calculée en capitalisation individuelle:

$$r(\hat{w}, \hat{x}) = q_x^* \hat{w}. \quad (91')$$

Ceci représente une droite par l'origine qui coupera le polygone défini par (85'') en un point d'intersection I dont l'abscisse sera \hat{w}_I . A gauche de ce point, les différences ${}_n r(\hat{w}) - r(\hat{w}, \hat{x}^*)$ seront positives et, à droite, négatives. Ces différences mesurent la solidarité résidant dans le polygone (85''). Cette mesure tient compte à la fois de la solidarité économétrique ainsi que des deux formes de solidarité biométrique qui sont à la base de l'AVS et qui favorisent, d'une part, les personnes mariées et, d'autre part, les classes d'âge à $\hat{x} > x_0$. Nous désignons la droite (91') par *droite «biométrique»*; elle est en effet entièrement déterminée par le coefficient angulaire ϱ_x^* calculé à l'aide des seules données biométriques.

Nous pouvons construire encore une *droite «économétrique»*. Ce sera celle qui fournira pour un couple donné d'indices m, n la même rente moyenne \bar{r} que le polygone (85''). Pour l'obtenir il suffit de définir un taux $\bar{\varrho}$ à l'aide de la rente moyenne \bar{r} et de la valeur moyenne \bar{w} des cotisations individuelles moyennes \hat{w} , cette dernière étant calculée à l'aide de (88); on pose $\bar{\varrho} = \frac{\bar{r}}{\bar{w}}$ et on envisage la droite par l'origine:

$$\hat{r}^*(\hat{w}) = \bar{\varrho} \hat{w}, \quad (92)$$

qui donne bien la même rente moyenne \bar{r} que le polygone $r(\hat{w})$. En calculant les prestations selon cette droite, on accorde à un ensemble de bénéficiaires caractérisé par un groupe donné de critères biométriques (âge, sexe, état civil) en moyenne les mêmes avantages que le polygone (85''). La comparaison de ce dernier avec cette droite (92) met donc en évidence la seule solidarité économétrique. Vu que (91') permet de mesurer la combinaison des deux solidarités en question, on arrive à dégager la solidarité biométrique seule, en confrontant les deux droites (91') et (92).

§ 4. Problèmes économétriques relatifs aux caisses de pensions

46. Du fait de l'évolution économétrique récente, les problèmes relatifs aux caisses de pensions relèvent de plus en plus de méthodes particulières aux assurances sociales. *Les cotisations* sont en général proportionnelles aux salaires selon la relation $w = \pi u$ où π est le taux de cotisation uniforme. En revanche, pour *les pensions*, la proportionnalité au salaire spécifique servant de base au calcul des

prestations est de plus en plus corrigée en faveur des personnes à revenus faibles en leur garantissant une pension minimum r_0 ; en outre, on introduit un maximum r_1 de pensions par lequel on tronque les prestations correspondant aux revenus élevés. Le salaire spécifique utilisé pour la détermination du montant de la prestation est en général celui qui précède immédiatement l'attribution de la pension à l'âge x_1 . Associons le salaire u' à l'âge x_0 (limite d'âge inférieure pour l'admission) et u'' à l'âge x_1 , de sorte que $u'' = u' s(x_1)$. Si ϱ désigne le taux de pension — dépendant en général d'un couple d'indices m (risque: décès, invalidité, vieillesse) et n (durée de cotisation) — on aura les règles suivantes pour le calcul d'un terme annuel de la pension:

$$r(u'') = \varrho u'', \quad (93)$$

$$r_0 \leq r(u'') \leq r_1. \quad (93')$$

Pour examiner l'équilibre financier d'une caisse, l'actuaire établit un *bilan technique*. Lorsque les statuts de la caisse prévoient des primes de rachat pour couvrir les charges supplémentaires résultant des augmentations de salaires, le bilan est calculé sur la base des salaires u observés à l'époque t de l'établissement du bilan. Au contraire, si ces charges supplémentaires sont sensées être déjà couvertes par le taux de cotisation π , on peut tenir compte des augmentations de salaires en introduisant dans les calculs l'échelle $s(x)$ déduite du matériel statistique de la caisse. C'était là d'ailleurs le but des travaux de King [24] que nous avons signalés au n° 7. On s'écartera évidemment ainsi du principe de la prime individuelle pour se rapprocher de celui de la prime moyenne générale. Calculons d'abord, à l'époque t du bilan et pour un assuré d'âge x , la valeur actuelle $\bar{a}_{x:\overline{x_1-x}|}^*$ des cotisations futures, la cotisation initiale $w = w' s(x)$ augmentant d'une manière continue selon $s(\xi)$ pour $x < \xi < x_1$. Si $\bar{a}_{x:\overline{x_1-x}|}$ représente la valeur actuelle d'une cotisation annuelle de 1 franc, payable d'une manière continue, on obtient la valeur actuelle cherchée par:

$$\bar{a}_{x:\overline{x_1-x}|}^* = \pi u' \left[\bar{a}_{x:\overline{x_1-x}|} s(x) + \frac{1}{\bar{D}_x} \int_x^{x_1} \bar{D}_\xi \bar{a}_{\xi:\overline{x_1-\xi}|} ds(\xi) \right]. \quad (94)$$

Relevons qu'il suffit de diviser cette relation par $w' s(x, x_1)$ pour obtenir la valeur actuelle de cotisations dont la moyenne annuelle serait = 1. Calculons encore, à titre d'exemple, la valeur actuelle

${}_{x_1-x_1} \bar{a}_x^*$ d'une rente de vieillesse déterminée à l'aide du taux $\varrho_{x_1-x_1}$ sur la base du revenu u'' concernant l'âge x_1 ; il vient, en introduisant le symbole habituel ${}_{x_1-x_1} \bar{a}_x$ de la rente différée unité, payable de manière continue:

$${}_{x_1-x_1} \bar{a}_x^* = \varrho_{x_1-x_1} u' s(x_1) {}_{x_1-x_1} \bar{a}_x. \quad (95)$$

Si l'on veut partir directement du salaire u observé à l'époque du bilan, on remplacera dans les deux formules ci-dessus u' par $\frac{u}{s(x)}$.

Pour évaluer correctement la totalité des engagements, il faut évidemment tenir compte de la nature polygonale des prestations définies par (93) et (93'). Puisque $f(u'', x_1)$ donne la répartition présumée des futurs pensionnés selon le revenu déterminant u'' , il suffit de multiplier (95) par $L_t(x) f(u'', x_1)$ et d'intégrer en u'' ; on mettra ainsi en évidence la pension moyenne calculée à l'aide des valeurs polygonales sur les trois intervalles définis par (93) et (93'), ceci d'une manière analogue à celle de la formule (89). Pour les autres risques on procédera d'une façon semblable en tenant toutefois compte de la répartition f valable au moment de la réalisation du risque.

47. Il n'entre pas dans le cadre de ce travail d'étudier les relations des caisses de pensions (CP) avec l'AVS. Relevons simplement que du fait de l'augmentation du coût de la vie, les rentes ordinaires de l'AVS sont le plus souvent considérées par les CP comme allocations de vie chère venant s'ajouter aux pensions statutaires. Ce qui intéresse avant tout chaque assuré c'est de connaître le taux total de sa pension de retraite ainsi cumulée CP + AVS. Alors que le taux ϱ^{CP} est directement donné par les statuts, il convient de calculer correctement le taux ϱ^{AVS} en rapportant la rente AVS $r(\hat{w})$ au même salaire de référence que celui déterminant ϱ^{CP} , c'est-à-dire le salaire u'' ; on a donc:

$$\varrho^{AVS} = \frac{r(\hat{w})}{u''}. \quad (96)$$

Pour se rendre compte de la nature de ce taux, il faut remplacer $r(\hat{w})$ par son expression (85'') et exprimer u'' à l'aide de \hat{w} , ce qui est facile si l'on se rappelle que $u'' = s(x_1) u'$ et $\hat{w} = s(\hat{x}, x_1) \pi u'$; ainsi il vient:

$$u'' = \frac{\hat{w}}{\pi} \frac{s(x_1)}{s(\hat{x}, x_1)}. \quad (97)$$

La substitution dans (96) effectuée, on voit que ϱ^{AVS} est toujours décroissant avec u'' ; c'est là que réside d'ailleurs la solidarité économétrique de l'AVS qui préoccupe avant tout les actuaires étudiant le problème de l'adaptation des pensions aux rentes AVS. Afin de mettre en évidence la portée collective de l'AVS pour l'ensemble des membres d'une CP, l'actuaire établit souvent un bilan supplémentaire relatif aux seuls éléments AVS. Cette tâche sera facilitée par la nouvelle collection de tables « Valeurs actuelles des cotisations et des rentes AVS » déjà signalée au n° 45, collection remplaçant à ce point de vue la publication [30]. Pour l'étude détaillée des relations entre l'AVS et les CP nous renvoyons à cette dernière publication ainsi qu'à une autre [23] où nous avons eu l'occasion d'exposer le problème sous un aspect différent.

§ 5. Un problème économétrique concernant l'assurance-maladie

48. Plusieurs cantons et communes suisses ont décrété, sur la base de la loi fédérale sur l'assurance maladie du 13 juin 1911, cette assurance obligatoire pour toutes les personnes dont le revenu est inférieur à une certaine limite u_1 . Ainsi on peut essayer d'évaluer d'abord *le nombre total des personnes assurées* à l'aide de la distribution des revenus de la population entière considérée. La proportion des assurés sera donnée par la fonction cumulative de la distribution globale des revenus, fonction désignée par $\bar{F}_t(u)$ qui se calcule à l'aide de (30) et dont l'élément démographique $\lambda_t(x)$ seul varie avec le temps; la proportion des personnes assurées sera donc de $\bar{F}_t(u_1)$. On peut évidemment se fixer d'emblée la proportion $\bar{F}(u_1)$ et en déduire le u_1 correspondant. Ceci est facile si l'on prend la distribution semi-hyperbolique $\bar{P}(u; a_0, a_1, \alpha)$ en s'imposant d'avance $u_1 \geq a_1$; dès lors on a $\bar{F}(u_1) = 1 - \bar{H}(u_1)$ et $\bar{H}(u_1) = \bar{a}^\alpha u_1^{-\alpha}$, d'où u_1 .

49. La connaissance du nombre total des assurés ne suffit cependant pas pour évaluer correctement les charges financières, ces dernières dépendant avant tout du taux instantané de morbidité $\varkappa(x)$ qui varie sensiblement avec l'âge. Il faut donc déterminer *le nombre des assurés d'âge x* qui est fourni par la valeur particulière $F(u_1, x)$ de la cumulative conditionnelle définie par (27). Cette valeur varie avec x , mais demeure constante par rapport à t grâce à l'hypothèse B (cf. n° 10).

Si nous supposons, faute d'observations statistiques plus précises, que les assurés d'âge x se répartissent selon une distribution de Pareto $P(u; a_x, \alpha)$, le calcul de $F(u_1, x)$ se fait dès lors facilement. On aura, en effet: $F(u_1, x) = 1 - H(u_1, x)$ c'est-à-dire en se rappelant (39) et (43'):

$$F(u_1, x) = 1 - a_0^\alpha s^\alpha(x) u_1^{-\alpha}. \quad (98)$$

Cette formule donne l'équation d'une courbe en x , dont la concavité est tournée vers le haut, pour autant que celle de $s(x)$ soit tournée vers le bas. *Le nombre de jours de maladie* provenant des effectifs $L_t(x)$ au cours d'une année débutant à l'instant τ sera ainsi donné par:

$$K_\tau = \int_\tau^{\tau+1} dt \int_{x_0}^{x_1} L_t(x) F(u_1, x) \varkappa(x) dx, \quad (99)$$

où x_0 et x_1 sont les âges extrêmes à prendre en considération.

Conclusions

La méthode économétrique que nous avons développée au chapitre I nous a permis non seulement d'étudier plus à fond la théorie de la distribution des revenus mais encore de résoudre quelques problèmes posés par la technique mathématique des plus importantes branches de la sécurité sociale. *Le champ d'application* de cette méthode pourra cependant s'étendre facilement. Elle s'appliquera d'abord aux autres secteurs de la sécurité sociale, dès que les montants de prestations s'échelonnent d'une manière polygonale en fonction du revenu. Tel est par exemple le cas pour les assurances couvrant le risque chômage, de nature économique ou militaire. La méthode envisagée rendra sans doute également des services dans le domaine des impôts et des statistiques fiscales.

La généralisation de notre méthode pourra encore être poussée plus loin, tant au point de vue mathématique que statistique. En particulier, il serait intéressant de faire dépendre la fonction $s(x)$, l'une des trois fonctions fondamentales de cette théorie, des deux autres variables t et u .

Il devient évident que l'actuaire de la sécurité sociale aura de plus en plus à s'occuper de *problèmes économétriques*. C'est même en cela que son activité scientifique et pratique se distingue essentiellement de celle de l'actuaire de l'assurance privée. Nous espérons que la méthode exposée dans ce travail facilitera la résolution des problèmes posés par le développement de la sécurité sociale et que les idées suggérées ouvriront la voie à des recherches nouvelles.

Annexe I

Bibliographie

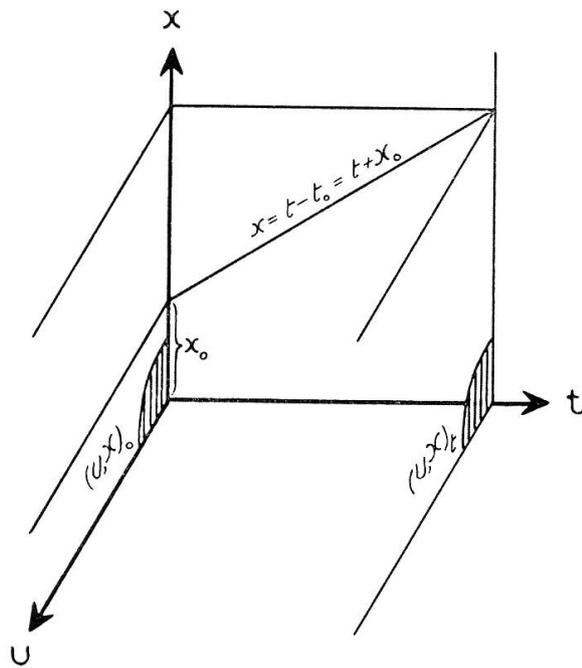
- [1] Administration fédérale des impôts: «Impôt fédéral pour la défense nationale, 3^e Période, Sacrifice pour la défense nationale 1945, Suisse, I^{re} Partie, Impôt pour la défense nationale» (Statistique de la Suisse/192^e fascicule, Série Md 53).
- [2] Ammon, O.: «Die Gesellschaftsordnung und ihre natürlichen Grundlagen» (Jena 1895).
- [3] Arbeitswissenschaftliches Institut: «Zwei Ausgleichungsmethoden für statistische Verteilungen» (Jahrbuch 1939, Berlin).
- [4] Billeter, E. P.: «Über die Messung der Einkommens-Konzentration» (Bern 1949).
- [5] Brüscheiler, C.: «Saläre und Arbeitsverhältnisse kaufmännischer Angestellter in der Schweiz» (Zürich 1938).
- [6] Bureau fédéral de statistique: «Le revenu national suisse 1938—1948» (Extrait de l'Annuaire statistique de la Suisse 1948, Berne 1948).
- [7] Cockfield, F. A.: «The distribution of incomes» (Economica, The London School of Economics and Political Science, London 1947/XI).
- [8] Cramér, H.: «Mathematical methods of statistics» (Uppsala 1945).
- [9] Département fédéral de l'économie publique: «La vie économique» (Berne 1946/VII).
- [10] Féraud, L.: «Les instruments mathématiques de la statistique» (Lausanne, Paris 1946).
- [11] Féraud, L.: «Introduction aux problèmes financiers de l'assurance sociale» (Extrait de la Revue internationale du Travail, vol. XXXVIII, n^o 1, Genève 1938/VII).
- [12] Féraud, L.: «Technique actuarielle et organisation financière des assurances sociales» (Bureau international du Travail, Etudes et Documents, Série M, n^o 17, Genève 1940).
- [13] Fonds centraux de compensation: «Enquête spéciale sur les conditions des militaires» (Genève 1943).
- [14] Fréchet, M.: «Nouveaux essais d'explication de la répartition des revenus» (Revue de l'Institut international de statistique, La Haye 1945).
- [15] Fréchet, M.: «Sur les formules de répartition des revenus» (Revue de l'Institut international de statistique, La Haye 1939).
- [16] Gibrat, R.: «Les inégalités économiques» (Paris 1931).
- [17] Gini, C.: «Memorie di Metodologia Statistica» Vol. 1: Variabilità e Concentrazione (Milano 1939).
- [18] Goursat, E.: «Cours d'analyse mathématique» (Paris 1927).

- [19] Haller, B.: «Verteilungsfunktionen und ihre Auszeichnung durch Funktionsgleichungen» (Bulletin de l'Association des Actuaires suisses, Vol. 45, 1^{er} fascicule, Berne 1945).
- [20] «Handwörterbuch der Staatswissenschaften» (Vol. III, Jena 1926).
- [21] Kaan, J.: «Die Finanzsysteme in der öffentlichen und in der privaten Versicherung» (Mitteilung des österreichisch-ungarischen Verbandes der Privat-Versicherungsanstalten, Wien 1909/XII).
- [22] Kaiser, E.: «Demographische und wirtschaftliche Rechnungsgrundlagen zur eidgenössischen Alters- und Hinterlassenenversicherung» (Bulletin de l'Association des Actuaires suisses, 1^{er} fascicule, Berne 1947).
- [23] Kaiser, E.: «Die städtischen Pensionskassen und die AHV» (Schweizerischer Städteverband, Zürich 1947).
- [24] King, G.: «On Staff Pension Funds» (Journal of the Institute of Actuaries, Vol. XXXIX, London 1905).
- [25] Krämer, M.: «Das Solidaritätsprinzip in der sozialen Altersversicherung» (Dissertation, Bern 1947).
- [26] Lévy, P.: «Leçons d'analyse fonctionnelle» (Paris 1922).
- [27] Linder, A.: «Statistische Methoden für Naturwissenschaftler, Mediziner und Ingenieure» (Basel 1945).
- [28] Odhnoff, W.: «Some studies of the characteristic functions and the semi-invariants of Pearson's frequency-functions» (Försäkrings-Matematiska Studier tillägnade Filip Lundberg, Stockholm 1946).
- [29] Office fédéral des assurances sociales: «L'équilibre financier de l'assurance vieillesse et survivants» (Berne 1947).
- [30] Office fédéral des assurances sociales: «La situation des institutions d'assurance existantes et futures dans l'assurance fédérale vieillesse et survivants» (Berne 1946).
- [31] Pareto, V.: «Cours d'économie politique» (Lausanne 1896).
- [32] Pearson, K.: «Tables of the incomplete Γ -Function» (Published by the Office of Biometrika, London, Re-issue 1946).
- [33] Pearson, K.: «Tables for Statisticians and Biometricians» (Issued by the Biometrika Office, University College, London 1930).
- [34] Rhodes, E. C.: «The distribution of incomes» (Economica, The London School of Economics and Political Science, Cambridge 1942/VIII).
- [35] Risser, R.: «Essai sur les courbes de distribution statistique» (Journal de la Société de Statistique de Paris, Paris 1948/VII—VIII).
- [36] Steffensen: «On the semi-normal distribution» (Uppsala 1937).
- [37] Winkler, W.: «The corrected Pareto law and its economic meaning» (XXVIth Session of the international statistical Institute, Bern 1949).

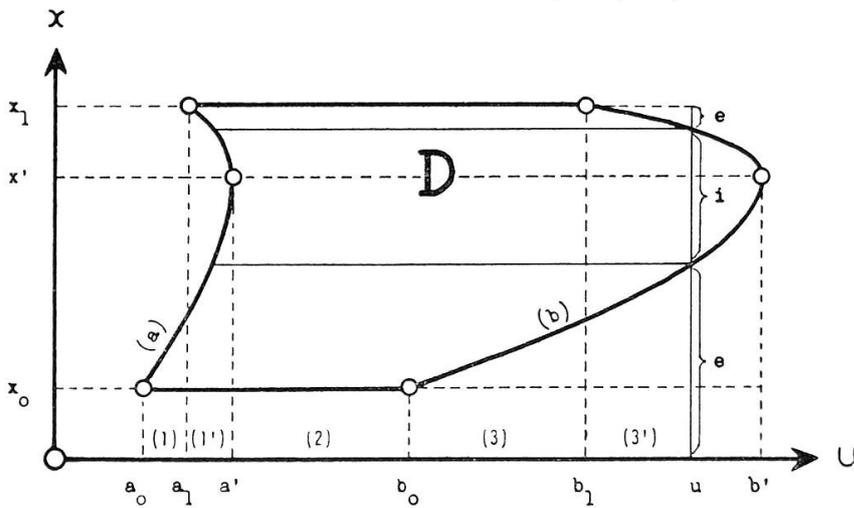
Annexe II

Représentations graphiques

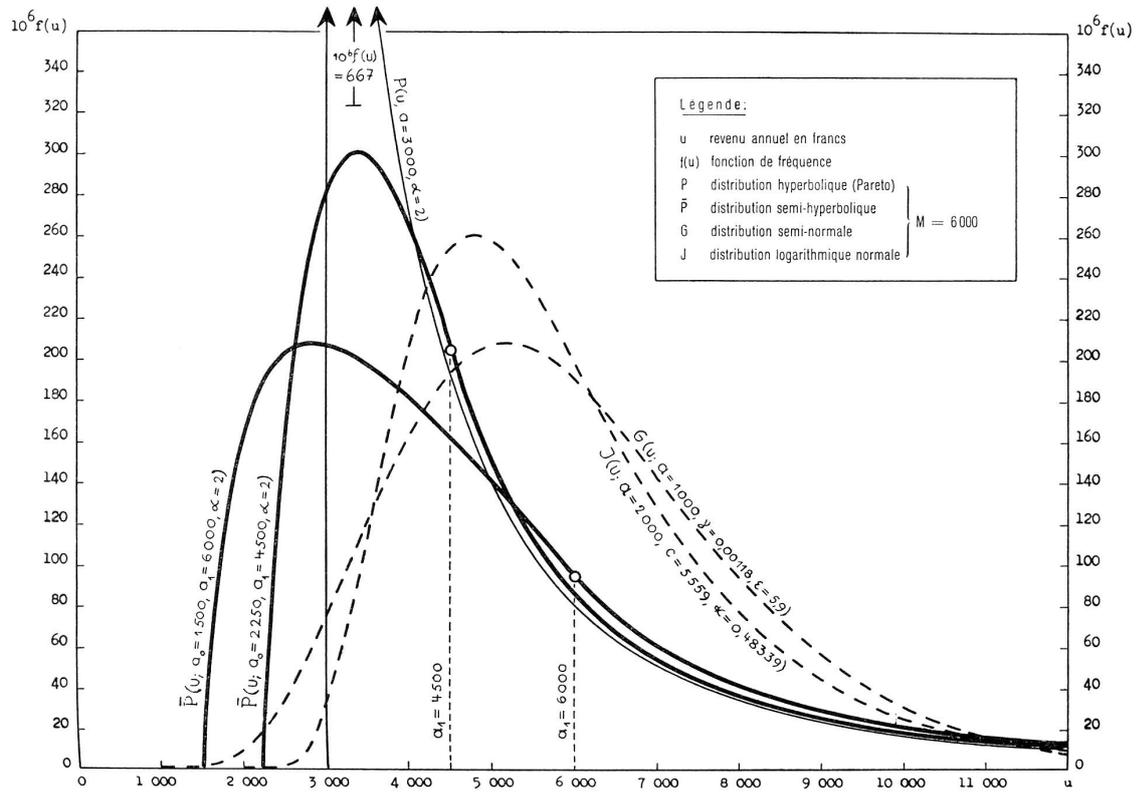
1. L'espace bio-économétrique (t, u, x)



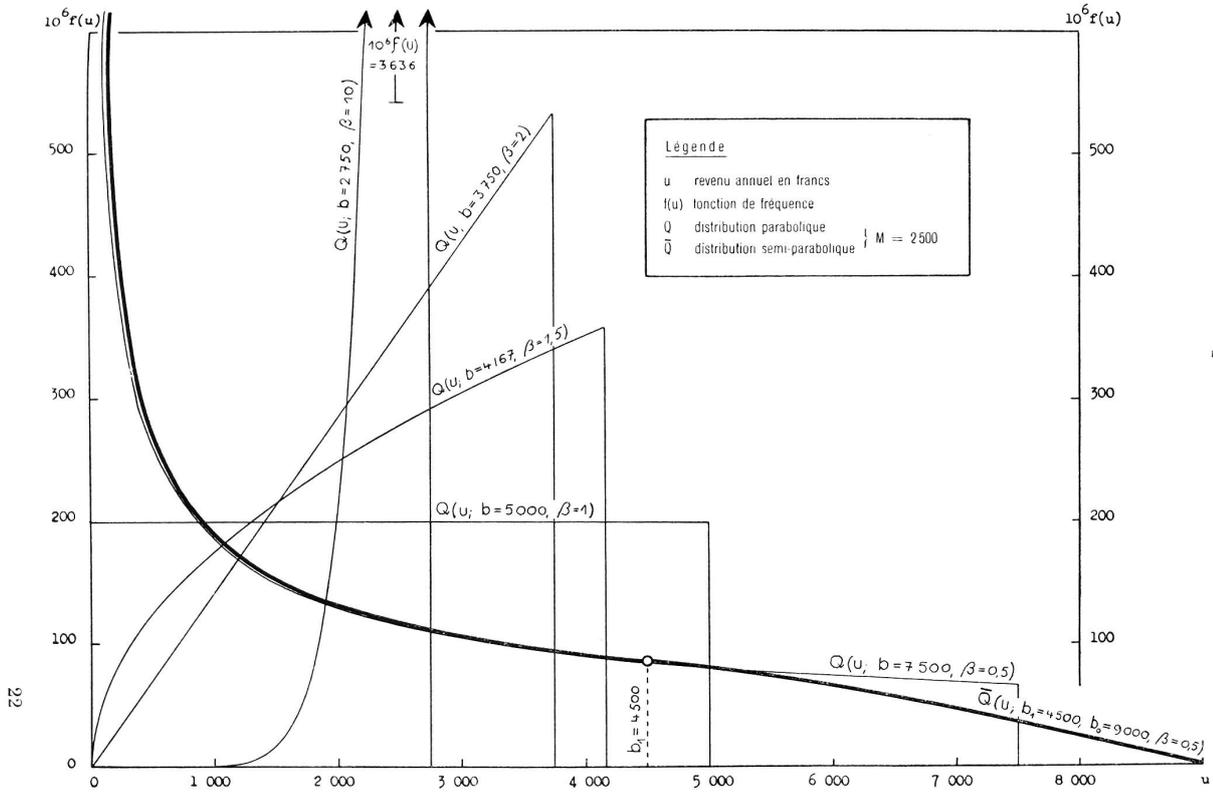
2. Le plan économétrique (u, x)



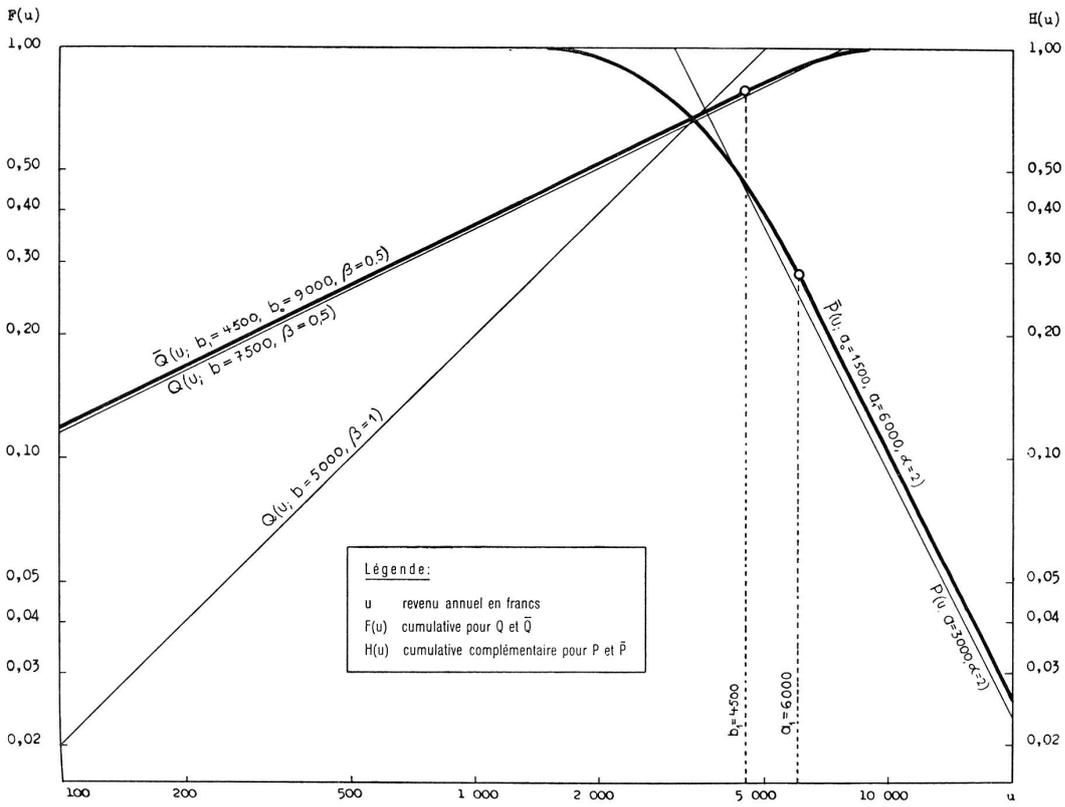
3. Les fonctions de fréquence des distributions P, P̄, G et J



4. Les fonctions de fréquence des distributions Q et \bar{Q}



5. Les cumulatives logarithmiques des distributions P , \bar{P} , Q et \bar{Q}
 (Voir fonctions de fréquence correspondantes aux graphiques 3 et 4)



Annexe III

Table des matières

	Page
Introduction	249
Chapitre I	
<i>Bases mathématiques de la théorie</i>	
§ 1. Variables et fonctions dans l'espace bio-économétrique (t,u,x)	251
§ 2. L'échelle des revenus moyens et son inversion	259
§ 3. Le calcul de la distribution des revenus dans l'espace (t,u,x) à partir de trois fonctions fondamentales.	262
§ 4. Le calcul de la distribution globale des revenus à l'époque t	267
§ 5. Les équations intégrales d'un problème connexe.	277
Chapitre II	
<i>Formules de répartition des revenus</i>	
§ 1. Postulats, méthodes déductives et vue d'ensemble.	280
§ 2. Les distributions hyperbolique (Pareto) P et semi-hyperbolique \bar{P}	285
§ 3. Les distributions semi-normale G et logarithmique normale J	295
§ 4. Les distributions parabolique Q et semi-parabolique \bar{Q}	301
Chapitre III	
<i>L'actuariat économétrique de la sécurité sociale</i>	
§ 1. L'équilibre financier dans l'espace bio-économétrique (t,u,x)	306
§ 2. La rente transitoire moyenne de l'AVS	315
§ 3. La rente ordinaire moyenne de l'AVS	318
§ 4. Problèmes économétriques relatifs aux caisses de pensions	323
§ 5. Un problème économétrique concernant l'assurance-maladie	326
Conclusions	328
Annexe I	
<i>Bibliographie</i>	329
Annexe II	
<i>Représentations graphiques</i>	
1. L'espace bio-économétrique (t,u,x)	331
2. Le plan économétrique (u,x)	331
3. Les fonctions de fréquence des distributions P, \bar{P}, G et J	332
4. Les fonctions de fréquence des distributions Q et \bar{Q}	333
5. Les cumulatives logarithmiques des distributions P, \bar{P}, Q et \bar{Q}	334