

La méthode de la distribution la plus probable
dans le cas des distributions lentement décroissantes :
et la loi de Pareto.

B. ROEHNER

Laboratoire d'Automatique Théorique
Université Paris 7

I. Introduction.

Depuis que la loi de Pareto* a été mise en évidence comme une loi empirique régissant de nombreux phénomènes (distribution de la taille des villes, distribution des revenus, fréquence des mots dans un texte), elle a également suscité de nombreuses tentatives d'explication. Certaines s'appuient sur ce qu'on appelle "la méthode de la distribution la fréquente" introduite en mécanique statistique par Ludwig Boltzmann. Il s'agit, comme on le montre plus loin, d'une méthode permettant de passer d'une hypothèse concernant les constituants microscopiques d'un système à son comportement macroscopique sans faire explicitement référence à la nature de ces constituants; il peut donc aussi bien s'agir d'atomes occupant tel niveau d'énergie que d'individus habitant une ville de telle taille ou ayant tel revenu ou encore de mots ayant telle fréquence d'apparition.

La généralité de cette approche, son aspect en apparence le plus séduisant, est aussi son côté le plus critiquable. N'est-il pas en effet illusoire d'espérer fournir une explication unique pour l'ensemble des phénomènes suivant une loi de Pareto? L'exemple suivant le montre, s'il en était besoin. Si on considère le nombre des étoiles de notre galaxie

* Nous l'écrivons sous sa forme la plus simple:

$$F(x) = \left[1 - \left(\frac{x_0}{x} \right)^\alpha \right] \theta(x - x_0)$$

$F(x)$: fonction de distribution de la variable aléatoire, dont x_0 est la valeur minimum

en fonction de leur masse ou de leur luminosité, on s'aperçoit que (sauf pour les plus petites, celles de la classe M) l'on a en gros une loi de Pareto (figure 1). Comme on sait, si on observe le nombre des villes classées en fonction de leur population on a une loi analogue.

Surtout la raison en est complètement différente; car si les étoiles les plus grosses sont les moins nombreuses, c'est parce que leur durée de vie est la plus courte (plus elles sont grosses, plus elles brûlent leur combustible rapidement) alors que dans le cas des villes on est bien sûr dans une situation diamétralement opposée: ce sont les villes les plus petites qui risquent une disparition prématurée par déclin économique, catastrophe naturelle ou fait de guerre. Ainsi, loin de rechercher une généralité trompeuse nous limiterons nous à l'exemple spécifique de la distribution des revenus et accessoirement à celle des villes.

La loi de Pareto diffère fondamentalement des distributions de type exponentiel qu'on rencontre en général en physique; elle est bien plus lentement décroissante avec comme conséquence que l'éventail des valeurs possibles est extrêmement large. Ainsi dans le cas des villes est-elle vérifiée pour des agglomérations allant de 1000 à plus d'un million d'habitants, alors que pour un gaz à la température T , la dispersion autour de la valeur moyenne $\frac{3}{2} kT$ (où k est la constante de Boltzmann) est de quelques fois cette même valeur moyenne. Il en résulte en particulier que la valeur moyenne (arithmétique) qui est en physique une grandeur importante à la fois expérimentalement et théoriquement n'a plus guère de signification ou même, d'un point de vue théorique, peut ne pas exister. Nous voulons essentiellement examiner quelques conséquences de cette situation.

II. La méthode de la distribution la plus fréquente.

A. En mécanique statistique, on a coutume de distinguer pour un système donné les variables macroscopiques (volume total, masse totale, énergie totale) qui le caractérisent globalement et sont directement mesurable et les variables microscopiques qui caractérisent ses N composants "élémentaires" (par exemple l'énergie ϵ_j de la molécule n° j , pour j allant de 1 à N) et ne sont pas, elles, directement mesurables. Dans une expérience sur le système les conditions du processus expérimental se traduiront par des valeurs fixées pour certaines grandeurs macroscopiques et nous noterons $C_1, C_2, \dots$ ces contraintes; les variables microscopiques pouvant, elles, évoluer à leur guise sous réserve que restent satisfaites les contraintes.

Dans la méthode de la distribution la plus fréquente, un ingrédient essentiel est la considération d'un ensemble $\mathcal{E}$ de systèmes virtuels pour les mêmes contraintes macroscopiques $C_1, C_2, \dots$ mais différant au niveau des composants élémentaires aussi bien par l'état initial que par l'évolution ultérieure (rappelons nous qu'en mécanique quantique à un état initial donné correspond tout un éventail d'évolutions possibles). Si on fait par exemple une expérience sur un système isolé, l'ensemble $\mathcal{E}$ sera défini par l'unique contrainte que l'énergie totale du système reste constante.

L'idée de la méthode est de dire: par la pensée je peux envisager la description la plus complète du système, celle où chaque variable microscopique est spécifiée (on l'appellera pour cela description microscopique). Mais des changements dans cette description microscopique (par exemple si les particules n° i et j échangent leur énergie) ne se traduiront pas nécessairement par des effets observables macroscopiquement. Je vais donc envisager une description moins détaillée (obtenue par regroupement des configurations microscopiques) où seuls le nombre, et non plus l'identité, des particules se trouvant dans chaque niveau d'énergie possible sera donné. Ainsi, les niveaux

d'énergie étant numérotés par l'indice k , cette description (qu'on appellera macroscopique ou encore statistique car elle a pour effet de gommer l'identité des constituants élémentaires) sera caractérisée par l'ensemble des nombres n_k .

C'est un résultat connu de l'analyse combinatoire qu'à une description statistique caractérisée par des nombres n_k correspondent :

$$W(n_1, n_2, \dots) = \frac{N!}{n_1! n_2! \dots}$$

configurations microscopiques.

Si donc on fait l'hypothèse que dans un ensemble $\mathcal{E}$ toutes les configurations microscopiques sont également fréquentes, on pourra en déduire le nombre W de systèmes correspondant à une certaine suite $\{n_k\}$ c'est à dire en somme à une certaine fonction densité de probabilité de l'énergie des particules.

Or cette fonction est mesurable (que l'on pense par exemple à la mesure de la distribution des vitesses dans un gaz par la méthode des jets moléculaires) et permettrait donc un test de l'hypothèse envisagée.

Rien ne nous empêche, bien entendu de calculer la suite $\{\tilde{n}_k\}$ pour laquelle compte tenu des contraintes*, W sera maximum. Or il se trouve, qu'à cause des grandes valeurs des nombres mis en jeu, cette distribution $\{\tilde{n}_k\}$ sera, comme on peut le montrer, tellement plus fréquente que toutes les autres que lors d'une observation sur $\mathcal{E}$ on n'observera pratiquement qu'elle.**

C'est en fait cette circonstance qui donne tout son intérêt à la méthode.

* Dans le cas d'un système isolé, d'énergie E , ces contraintes s'écrivent :

$$\sum_k n_k = N \qquad \sum_k n_k \epsilon_k = E$$

** En fait, on observe non seulement $\{\tilde{n}_k\}$ mais également toutes les distributions $\{\tilde{n}_k + \delta_k\}$ où les δ_k sont des nombres inférieurs aux erreurs absolues de mesure.

B- Considérons maintenant une collection de N individus et interrogeons nous sur revenu r_k de chacun (ou si la population p_k de la ville qu'habite chacun). Les r_k (ou les p_k) seront donc les variables microscopiques. Mais aucune grandeur macroscopique directement mesurable ne s'impose à nous et alors que le reste de la démonstration précédente se transpose sans problème, le point doit faire l'objet d'un examen détaillé. Il n'est pas possible dans les domaines envisagés de réaliser des mesures en fixant tel ou tel paramètre. Le statisticien n'a que la possibilité de sélectionner parmi un grand nombre d'observations celles où certaines conditions seront réalisées. Pour fixer les idées considérons le cas de la distribution des revenus aux Etats-Unis. Les annuaires statistiques américains la donne pour toutes les localités de plus de 2500 habitants. A partir de cette somme considérable de données comment va-t-on constituer un ensemble $\mathcal{E}$ de systèmes "macroscopiquement" identiques et à quelle contrainte correspondra-t-il ?

On peut penser, à priori, que la distribution des revenus dépend de multiples facteurs qualitatifs : milieu rural ou urbain, prédominance commerciale, industrielle ou administrative, histoire locale etc. S'il en était ainsi, il faudrait examiner une multitude d'aspects avant de pouvoir constituer $\mathcal{E}$. Fort heureusement, les études statistiques sur la loi de Pareto montrent une assez grande indépendance par rapport aux conditions locales.* Pour ce qui est du choix de la contrainte, nous allons procéder à rebours. Supposons nous donc dotés d'un ensemble $\mathcal{E}$ d'observations

* Celles ci influent certes sur la valeur de l'exposant α , mais on peut toujours "limer" les disparités locales en remplaçant dans une petite zone (de la dimension d'un comté par exemple) les distributions par leur valeur moyenne.

correspondant à une certaine contrainte, pour l'instant inconnue, que nous écrirons sous la forme:

$$K = \bar{g}^{-1} \left(\frac{1}{N} \sum_k m_k g(x_k) \right) \quad (1)$$

qui est une généralisation de la moyenne d'ordre h :

$$m_h = \left(\sum_k m_k x_k^h \right)^{1/h}$$

dont la moyenne arithmétique est un cas particulier.

Mettons en œuvre la méthode de la distribution la plus fréquente.

Pour maximiser K , nous prenons d'abord son logarithme et pour la dérivée nous utilisons la formule approchée (dérivée de la formule de Stirling et donc valable en pratique pour $n \geq 4$):

$$\frac{d(\ln n!)}{dn} \approx \ln(n + 1/2)$$

Nous avons par ailleurs à tenir compte en plus de la contrainte (1) que nous préférons réécrire sous la forme:

$$\sum_k m_k g(x_k) = N g(k)$$

de la contrainte:

$$\sum m_k = N$$

exprimant que la population considérée comporte N individus.

La méthode des multiplicateurs de Lagrange conduit alors aux relations (α et λ étant les multiplicateurs):

$$\ln m_k - \alpha - \lambda g(x_k) = 0 \quad (k \text{ de } 1 \text{ à } N)$$

où nous avons négligé le facteur $1/2$ puisque m_k est en général un nombre

bien supérieur à 1 *

Soit :
$$n_k = C' \cdot e^{-\alpha f(r_k)}$$

Comme on peut voir le comportement de n_k sera du type exponentiel et non pas lentement décroissant sauf dans les deux cas suivants :

1° $f(r_k) = p \cdot \ln(r_k + q)$ où on obtient :

$$n_k = \frac{C}{r_k^\alpha}$$

2° $f(r_k) \rightarrow l \neq 0$, mais dans le cas on obtient , pour $r_k \rightarrow \infty$

$N \rightarrow \infty$, une distribution qui n'est plus normalisable et nous la rejetons pour cette raison. Cela écarte en particulier la moyenne harmonique correspondant à $f(x) = 1/x$.

Bien entendu , on pourrait envisager bien d'autres possibilités ; par exemple une fonction $f(r_k)$ qui pour r_k tendant vers l'infini aurait un comportement suffisamment oscillant pour que la distribution devienne normalisable.

Comme il est cependant préférable que la contrainte ait une interprétation simple nous ne conserverons que la première solution avec de plus $p=1$, $q=0$. La contrainte exprime alors la constance de la moyenne géométrique :

$$G = (r_1^{n_1} r_2^{n_2} \dots)^{1/N}$$

* Si on prend pour les r_k tous les revenus mensuels possibles de dollar en dollar de r_0 jusqu'à la valeur maximum, il est clair que pour une population de quelques millions d'individus le nombre des niveaux sera supérieur à celui des individus et que la plupart des niveaux seront soit vides soit occupés par un seul individu. Il se pose exactement le même problème en physique pour un gaz. La solution consiste à regrouper les niveaux par paquets de taille suffisante.

[Janvier 1978]

Lorsqu'en mécanique statistique, on parle de valeur moyenne (de l'énergie, du moment magnétique etc) il s'agit toujours de la moyenne arithmétique dont la contrepartie dans le langage probabiliste est l'espérance mathématique. Cependant les statisticiens savent bien (Cramér page 76) que cette moyenne ne fournit une bonne caractérisation d'une distribution qu'à condition que celle-ci décroisse suffisamment vite pour les grandes valeurs. (voir Appendice).

Lorsqu'ils ont affaire à des distributions lentement décroissantes, ils préfèrent les caractériser par des moyennes dites de position : médiane (quartiles, déciles). Ces caractéristiques malheureusement se prêtent mal aux calculs analytiques.

Il existe cependant d'autres types de moyennes de grandeur que la moyenne arithmétique, telle la moyenne géométrique ou harmonique :

$$G(x) = \frac{E(\log x)}{e} \quad H(x) = \frac{1}{E(1/x)}$$

qui sont mieux adaptées aux distributions lentement décroissantes et ceci pour deux types de raisons :

- 1° Elles peuvent exister pour des distributions dont l'espérance diverge.
- 2° Pour des distributions lentement décroissantes qui ont cependant une espérance, la moyenne géométrique sera en général assez proche de la médiane ou contraire de l'espérance.

On peut illustrer cela par l'exemple de la distribution de Pareto dont la fonction densité de probabilité et la fonction de distribution s'écrivent :

$$g(x) = \frac{\alpha x_0^\alpha}{x^{\alpha+1}} \theta(x-x_0) \quad F(x) = \left[1 - \left(\frac{x_0}{x} \right)^\alpha \right] \theta(x-x_0)$$

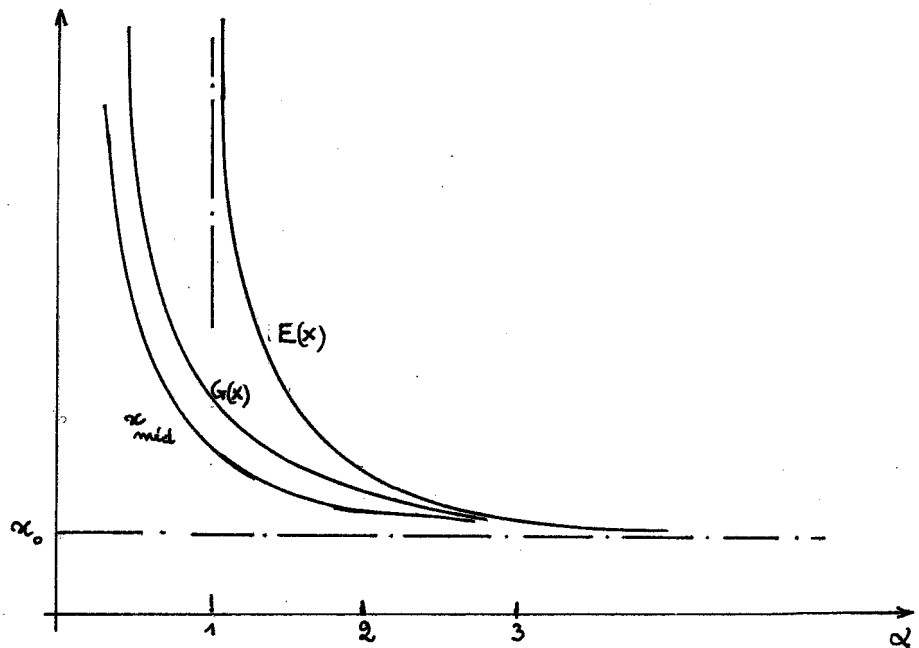
où $\theta(x)$ désigne la fonction échelon et α le coefficient de Porto : $\alpha > 0$
 On obtient pour la médiane et la moyenne géométrique :

$$\alpha_{\text{méd}} = 2^{1/\alpha} \alpha_0 \quad G(X) = 2^{1/\alpha} \alpha_0$$

L'espérance n'existe que si $\alpha > 1$ et dans ce cas :

$$E(X) = \frac{\alpha}{\alpha-1} \alpha_0$$

Ces résultats sont résumés par le graphique suivant :



* vers

Nous allons mettre à profit ces remarques pour étendre la méthode de la distribution la plus probable bien connue en mécanique statistique au cas de distributions lentement décroissantes.

Cette méthode s'applique à un système constitué d'un grand nombre d'éléments que nous appellerons microscopiques et auquel on impose pour la grandeur g étudiée une certaine contrainte macroscopique (g moyen fixé, par exemple) ; elle permet, à partir d'une hypothèse d'équiprobabilité concernant les états microscopiques, d'obtenir la probabilité pour la répartition des éléments microscopiques entre les états possibles. En cela, la méthode déborde du cadre où elle est

* Un autre exemple est fourni par la loi ayant pour densité de probabilité:

$$f(x) = \frac{\alpha (\log x_0)^\alpha}{x (\log x)^{\alpha+1}} \theta(x-x_0) \quad \begin{matrix} \alpha > 0 \\ x_0 > 1 \end{matrix}$$

Son espérance diverge pour toute valeur de α ; mais sa médiane vaut:

$$x_{\text{méd}} = x_0^{\frac{1}{2} \alpha} \quad \forall \alpha > 0$$

et sa moyenne géométrique existe pour tout α supérieur à 1:

$$G(x) = x_0^{\frac{\alpha}{\alpha-1}}$$

Dès que α devient supérieur à 4, $x_{\text{méd}}$ et $G(x)$ diffèrent de moins de 15%.

usuellement utilisée et peut s'appliquer à toute population comportant un grand nombre de degrés de liberté. Pour fixer les idées nous allons raisonner sur la répartition du revenu* dans une population de N individus; nous les supposons distinguables les uns des autres et nous les identifions par leur numéro k , variant de 1 à N .

Le système sera complètement décrit par la donnée pour chaque k du revenu correspondant ce que nous appellerons une configuration détaillée. Contrairement à ce qui se lie en physique, les configurations détaillées sont ici observables (par les déclarations individuelles de revenu) mais ces données n'acquiescent de signification statistique qu'après agrégation.

Notre postulat sera que pour un système isolé, toutes les configurations détaillées sont également probables. Dans la mesure où, de même qu'en physique, tout système isolé est inobservable (par un observateur extérieur), il s'agit bien là d'un postulat uniquement vérifiable par ses conséquences. Remarquons qu'en mécanique statistique, ce postulat devrait en principe pouvoir être déduit des équations du mouvement (classiques ou quantiques); un problème qui reste, comme on sait, en suspens.

Nous appellerons configuration grossière la donnée pour chaque niveau (discret) de revenu r_k , du nombre d'individus n_k ayant ce revenu.

Nous considérons une situation où par suite des interactions du système avec le milieu extérieur (climat, ressources naturelles, état de développement technique etc), la distribution de revenu prend une valeur moyenne géométrique fixée:

$$\langle r \rangle_g = [r_1^{n_1} r_2^{n_2} r_3^{n_3} \dots]^{1/N} = G$$

* mais on pourrait faire un raisonnement semblable pour d'autres situations présentant une distribution lentement décroissante, par exemple la répartition des villes selon leur taille en nombre d'habitants.

Comme à chaque configuration grossière $\{n_1, n_2, n_3, \dots\}$ correspondent :

$$W(n_1, n_2, n_3, \dots) = \frac{N!}{n_1! n_2! n_3! \dots}$$

configurations détaillées, nous sommes conduits pour obtenir la distribution la plus probable à maximiser W compte tenu des deux contraintes :

$$\sum_k n_k = N \quad (1)$$

$$\langle r \rangle_g = G \quad (2)$$

La méthode des multiplicateurs de Lagrange conduit de suite au résultat :

$$n_k = \frac{\frac{-\alpha}{e}}{r_k^{\alpha+1}}$$

où α et α sont déterminées par les relations :

$$\frac{-\alpha}{e} = \frac{N}{\sum_k r_k^{-(\alpha+1)}} \quad \sum_k r_k^{-(\alpha+1)} \log r_k = \frac{\log G}{\sum_k r_k^{-(\alpha+1)}}$$

il s'en résulte en particulier que α doit être supérieur à 0 et varie en sens inverse de G .

Cet argument de caractère général conduisant à la distribution de Pareto se fait donc apparaître comme le pendant, pour la classe des distributions lentement décroissantes, de la distribution de Maxwell-Boltzmann. Et, à l'instar de cette dernière dans son propre domaine d'application, elle couvre en effet un grand nombre de phénomènes distincts.

Appendice. Que faut-il entendre par bonne caractérisation ? Les statisticiens ont pour cela un certain nombre de critères (condition de Yule), mais pour le physicien la justification essentielle de l'utilisation de la moyenne arithmétique vient en dernier ressort du théorème central-limite. Celui-ci lui assure que si l'on mesure une certaine valeur moyenne (par ex. celle de la vitesse des particules d'un gaz) plusieurs fois de suite, il obtiendra des valeurs fortement groupées autour d'une valeur centrale. Certes d'un point de vue mathématique, la loi forte des grands nombres, qui requiert seulement, dans le cas de variables indépendamment distribuées l'existence de ^{la variance} ~~l'espérance~~, suffit à assurer ce regroupement autour d'une valeur centrale. Mais dans la pratique, la moyenne porte sur un nombre fini d'observations et on est donc surtout intéressé par la rapidité de la convergence ; or celle-ci est d'autant meilleure que la loi considérée possède un plus grand nombre de moments.

Pour des lois qui décroissent trop lentement à l'infini pour avoir une variance, le théorème central limite ne s'applique plus et on n'est plus assuré de cette stabilité d'une série de mesures à une autre. Dans la pratique, on observe alors des fluctuations importantes lors de mesures répétées de la valeur moyenne. Ainsi par exemple, la mesure du revenu moyen d'une population qui correspond à une distribution lentement décroissante sera très sensible à la prise en compte plus ou moins complète des revenus élevés.

Est-ce que le passage de la moyenne arithmétique à la moyenne géométrique atténue ces fluctuations ? C'est difficile à dire du fait justement que le théorème central-limite ne s'appliquant pas, on ne connaît pas la distribution de :

$$A_n = \frac{1}{n} [x_1 + x_2 + \dots + x_n]$$

Quant à celle de :

$$G_n = [x_1 \cdot x_2 \cdot \dots \cdot x_n]^{1/n}$$

on peut affirmer que sauf si la distribution commune des x_i est très lentement décroissante (comme par ex: $\frac{1}{x (\log x)^{\alpha+1}}$ avec $0 < \alpha < 1$),

G_n suivra une loi log-normale. Cela n'assure certes pas la stabilité que posséderait la loi normale, mais la loi log-normale possède néanmoins des moments à tous les ordres et décroît donc plus vite que toute puissance de $1/x$.

Il est vrai que la moyenne géométrique présente un certain nombre d'inconvénients qui expliquent qu'elle ait été jusqu'ici assez peu utilisée :

- sa définition suppose une variable aléatoire de signe constant et non nulle.

- la moyenne géométrique d'une somme de variables aléatoires indépendantes et non pas égales, mais supérieure ou égale à la somme des moyennes géométriques (Polya page 69)

Notons que cette méthode suppose la quasi-indépendance des "particules" et donc une interaction faible, une hypothèse qui n'est vérifiée que pour des systèmes physiques très spécifiques (gaz rares). Dans les systèmes réels on ne sait pas mesurer l'interaction mais il est sûr qu'il y en a une.

Cr  mer The elements of probability theory and some
of its applications (1954)

P  lya & Szeg   Problems and theorems in analysis (1972)

Sur la m  thode de la distribution la plus probable :

Huang Statistical Mechanics (1963) pages 75 et suivantes
206 et suivantes.

UNIVERSITE PARIS VII
LABORATOIRE
D'AUTOMATIQUE THEORIQUE

2, Place Jussieu - 75221 PARIS CEDEX 05
Tour 14-24 - 5e étage
TEL. 338-25-26

*transmis le 30.04.75 à
tous les chercheurs -*

Je propose une série (prolongée) d'exposés de caractère introductif sur le thème :

MODELES MATHEMATIQUES EN ECONOMIE ET ECONOMIE POLITIQUE

Y seraient abordées, à titre indicatif, les questions suivantes :

- Application des équations aux différences finies : modèles de croissance,
- Systèmes d'équations différentielles : théorie générale de l'équilibre,
- Modèles mathématiques d'économie marxiste,
- Applications des méthodes d'optimisation et de la théorie des jeux.

Sur tous ces sujets, on marquera une nette prédilection pour les exemples simples afin :

d'une part, de donner à ces exposés de fin de semaine (un consensus a semblé se dégager pour le vendredi après-midi) un caractère récréatif,

d'autre part, de ne pas obscurcir le débat de fond par un appareil mathématique trop lourd.

Il va de soi que les contributions de tous les intéressés, sous forme d'exposés complémentaires par exemple, seraient les bienvenues.

La première séance aura lieu le :

VENDREDI 16 MAI 1975, à 14 h. 30 (Salle 03)

et en voici le programme :

Ouverture : Un modèle dynamique élémentaire d'inflation par le revenu

1er mouvement : L'équation aux différences finies du 2ème ordre et son application au modèle de croissance de Samuelson

2ème mouvement : Remarques générales sur le cadre de ces modèles : l'optique de Keynes

etc, etc ...

A PARTIR DE 16 h. 30 : DISCUSSION

BERTRAND ROEHNER