n° d'ordre : 364

THESE DE DOCTORAT DE 3e CYCLE

Discipline : Mathématiques Appliquées

présentée à

L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE

par

Yves ALEXANDRE

THEORIE DE LA DECISION

ANALYSE DE QUELQUES CRITERES DE CHOIX

-=-

Thèse soutenue le 14 février 1973 devant la Commission d'Examen

M. POUZET : Président du jury

M. VAILLANT : Examinateur

Mlle. MARQUET : Directeur de travail

A Joëlle

Je désire exprimer mes respectueux remerciements à Monsieur le Professeur POUZET qui m'a fait l'honneur de présider le jury, à Monsieur le Professeur VAILLANT qui a bien voulu accepter la charge d'examinateur et à mon Professeur, Mademoiselle MARQUET, qui a assuré la direction des travaux et dont j'ai apprécié l'aide constante.

*   *

*

S O M M A I R E

-=-=-=-=-=-

THEORIE DE LA DECISION

ANALYSE DE QUELQUES CRITERES DE CHOIX

Y. Alexandre

(Février 1973)

Ce travail vise à différencier les divers critères de choix disponibles pour la prise de décision en avenir incertain.

Dans une première grande famille de théories, les critères de choix s'exprime à l'aide de fonctions de risque comme dans la théorie générale de la décision de A. WALD. Ces critères concernent des fonctions de décisions statistiques (décision prise après expérimentation) et les solutions optimales se distinguaient initialement par l'existence ou la non existence d'une distribution de probabilité a priori sur les états de la nature. Plus récemment, d'autres critères ont été introduits, en particuliers des critères mixtes correspondant à des structures de choix intermédiaires (la classification étant faite par rapport à l'existence de la distribution a priori). Nous en examinons quelques cas : le critère "mean-max" introduit apr H. KUDO, le critère "Bayes-minimax" de K.A. DOKSUM et le critère minimax de R.A. BERAU.

En second lieu nous analysons des critères de choix moins développés quant à leurs aspects formels mais d'interprétation plus axiomatisée. Les conditions d'optimalité s'expriment en termes de fonction d'utilité et de probabilités subjectives (théorie statistique de SAVAGE par exemple) et le critère de choix est la maximisation d'une espérance mathématique de celle-ci, espérance mathématique par rapport à la probabilité subjective mentionnée ci-dessus. La théorie générale de P.C. FISHBURN correspondant à ce schéma est en particulier détaillée.

I - PRELIMINAIRE

A la suite de l'idée formulée par J. HAJEK (21), à propos de l'analyse des concepts de base en statistiques, que les statistiques classiques sont riches en méthodes et ne s'attardent pas sur les critères et que la théorie de la décision ajoute un grand nombre de critères mais ne fournit que très peu de méthodes, nous allons nous interroger sur la nature des critères de choix employés en théorie de la décision.

Nous limiterons notre recherche aux seuls aspects formels des quelques critères de choix envisagés sans aborder les questions, pourtant essentielles en pratique, de leur interprétation ou celles des axiomes sur lesquels ils reposent. Les éléments, autres que les critères et ceux sur lesquels ces derniers sont directement définis, ne seront pas discutés.

Nous nous attacherons spécifiquement au caractère incertain avec lequel les problèmes de choix sont envisagés. Ainsi nous n'aborderons pas les critères de choix qui ne visent pas explicitement à lever une indétermination posée par une incertitude. Par exemple les théories diverses d'optimisation, celles de choix collectifs, les méthodes à critères multiples ... seront exclues.

N'ayant absolument pas la volonté d'être exhaustif sur cette question du critère de choix dans l'incertitude, nous nous contenterons de n'analyser que quelques formes de critère que nous jugeons néanmoins significatives.

I A - *Introduction de l'incertitude*.-

Le problème que résoud un processus de prise de décision dans l'incertitude est schématiquement le suivant : quelle décision choisir, sachant que sa conséquence est incertaine ? Dans cette question nous voyons au moins deux types possibles d'incertitudes :

1) Une incertitude entièrement liée à la dimension temporelle du problème. Son existence dépendra de la réponse à la question suivante : si les conséquences futures des décisions sont certaines, le choix actuel de l'une d'entre elles est-il aussi certain ? Ce type d'incertitude peut être implicitement pris en compte dans la recherche d'un critère de choix en fondant ce dernier sur une fonction d'utilité (plus précisément, ce type d'incertitude est essentiellement traduisible dans un modèle de décision par la nature essentiellement subjective qui est attribuée à la fonction d'utilité. Nous y reviendrons ci-dessous en I A 2).

2) Une incertitude liée à une méconnaissance du contexte du problème de choix, et cela indépendamment de sa dimension éventuellement temporelle.

La traduction de cette incertitude n'est pas une question nouvelle puisque, d'un point de vue synthétique, c'est l'objet des statistiques mathématiques. Par opposition à l'incertitude due au temps, le traitement de cette méconnaissance sera indépendante d'une éventuelle fonction d'utilité (voir également I A 2), et d'une manière générale nous distinguerons schématiquement deux types d'éléments dans un modèle mathématique de prise de décisions :

i) ceux qui traduisent cette incertitude atemporelle, comme par exemple les distributions de probabilités sur divers ensembles composant chaque théorie (ensemble des états de la nature, ensemble des décisions possibles,...)

ii) ceux qui prennent en compte cette traduction de l'incertitude (fonctions de risque, fonctions d'utilité, critère de choix ...).

I A 1.- *Incertitudes atemporelles*.-

Rappelons très brièvement d'abord les points classiquement soulevés pour traduire ces incertitudes en théorie de la décision, les théories que l'on rencontre présentant des traductions variées. Un premier point de divergence réside dans la place d'assignation de l'incertitude : si la plupart des statisticiens modernes assignent l'incertitude au niveau des états de la nature (ou évènements méconnus du responsable de la décision ; voir l'exemple précis des éléments de la nature pris par WALD en II A 1) d'autres l'assignent au niveau des décisions elles-mêmes. C'est seulement cette seconde solution qu'il est possible d'adopter si l'on veut englober les divers types de modèles de prise de décision et en particulier ceux qui reposent sur une fonction d'utilité (fonction définie seulement sur les conséquences des décisions à choisir). D'ailleurs dans les modèles statistiques de prise de décision où le choix a lieu après expérimentation, et où l'introduction de l'incertitude se fait au niveau des états de la nature sous forme de probabilité, le report de l'incertitude au niveau des conséquences est possible en respectant les règles de calcul des probabilités.

Si l'on suppose que l'incertitude est traduite par une probabilité, celle-ci pourra reposer sur des interprétations très nombreuses. Citons-en quelques unes : degré de confirmation, de conviction, de croyance rationnelle ; probabilité empirique, géométrique, impersonnelle, induite, intuitive, de jugement, logique, mathématique (au sens de l'axiomatique de Kolmogorov), objective, personnelle, physique, psychologique ; hasard ; fréquence relative ; probabilité subjective, statistique .... Néanmoins ces différents points de vue peuvent se résumer essentiellement à trois écoles :

i) l'interprétation fréquentielle où la probabilité est assimilée à une fréquence relative limite de quelque évènement d'un ensemble donné. Cette interprétation a été suivie par VENN, VON MISES, REICHENBACH, CRAMER, FISHER....

ii) le point de vue logique où la probabilité est une relation unique entre des propositions (KEYNES, JEFFREYS, CARNAP,...) Pour ces deux types de probabilités on parlera de probabilités objectives.

iii) l'approche subjective où la probabilité est l'expression d'une confiance dans la vérité de quelque chose (RAMSEY, DE FINETTI, KOOPMANN, GOOD, SAVAGE....).

Des définitions plus précises de ces interprétations sont données par P.C. FISHBURN dans (13).

Un troisième point de disparité entre les divers modèles de décisions réside dans les transformations qui peuvent être effectuées sur ces probabilités : c'est le cas classique où dans certaines théories, une distribution est introduite comme distribution a priori et où elle est transformée en distribution a postériori après expérimentation.

Remarquons, par ailleurs, que le fait de traduire l'incertitude par une probabilité correspond à un niveau intermédiaire de connaissances : une des limites correspond à une incertitude nulle ou certitude tandis que l'autre limite correspond à l'incertitude maximale acceptable, et qui correspond pour le style de théorie de décision envisagé dans ces lignes, à la seule connaissance de l'existence des évènements incertains possibles sans aucun jugement quant à leur réalisation (Dans les modèles de style "objectif", une incertitude nulle est traitée avec des critères de choix déterministes, une incertitude probabilisable avec des critères d'espérances et une incertitude totale avec des critères minimax). Cette classification des théories de décisions selon le niveau de traduction de l'incertitude pourrait être affinée. A titre d'exemple, notons que des théories se proposant de se situer entre le probabilisable et le totalement incertain en traduisant l'incertitude d'évènements possibles par une relation de préordre (à titre d'illustration, voir FOURGEAUD, LENCLUD, SENTIS, (18). De toute façon, les randomisations utilisées en décision reposent elles-mêmes sur des axiomatiques fort diverses et la littérature est abondante et variée sur ce sujet.

Cet examen sommaire de la nature de la traduction de l'incertitude va nous permettre néanmoins d'isoler une première catégorie de théories de décision. Il s'agit de théories de décision statistique (la décision étant adoptée au vue du résultat d'une expérimentation) où les probabilités introduites sont essentiellement objectives. La présentation de divers critères de choix adoptés dans un tel contexte est l'objet du chapitre II. Notons que cette présentation ne prétend à aucune exhaustivité et ne vise qu'à analyser des critères en faisant essentiellement ressortir les propriétés des décisions optimales obtenues. Pour approcher ces divers critères, nous avons fait varier soit la structure générale de la théorie de décision soit plus particulièrement le niveau de traduction de l'incertitude comme cela a été évoqué ci-dessus. Les critères de choix examinés ne font pas appel aux fonctions d'utilité dans le plein sens du terme mais uniquement à des fonctions de pertes et de risques. De tels critères peuvent ainsi être considérés comme appartenant au domaine des statistiques classiques et l'on retrouve à leur sujet les divergences qui existent par ailleurs entre les diverses écoles de statisticiens.

I A 2.- *Subjectivité et utilité*.-

Introduisons explicitement la dimension temporelle dans le problème de décision (décision en avenir incertain). Illustrons et analysons la notion d'utilité de manière littéraire en suivant une analyse faite dans le contexte spécifique du comportement économique. Cette introduction à l'utilité, tirée de G. BERNARD (4), procède en trois étapes :

1) Soit un objet "i" et la valeur **m** ou mesure de cet objet. (nous ne mentionnerons pas la signification économique exacte des termes employés, ceux-ci étant à prendre dans le sens naturel pour cet exposé). Supposons que la mesure m soit indépendante du temps (atemporelle) et de toute incertitude (certaine). Soit M l'ensemble des mesures des objets "i" possibles ; cet ensemble est alors invariant du temps et de tout opérateur dans l'incertitude. Introduisons alors sur M une fonction u(m) appelée "utilité pour une personne donnée de la valeur m certaine et atemporelle d'un objet i" définie par :

a) elle est une opinion subjective qui dépend essentiellement d'une personne (celle-ci pouvant être prise au large, l'essentiel consistant en l'unicité du décideur).

b) il n'existe pas de dimension utilité, l'utilité étant de la dimension de la valeur qu'elle concerne. Ceci suppose, entre autre, que soit éventuellement résolu par ailleurs le problème de l'expression d'une utilité scalaire pour des objets multidimentionnels.

c) l'axiome de substitution s'applique dans un certain ensemble fermé d'objets (axiome de substitution : étant donnée une valeur m d'un objet i de dimension d, il existe une valeur m' d'un autre objet i' de dimension d' tel que m' et m aient la même disponibilité pour une même personne). Par cette hypothèse, des utilités de différentes dimensions peuvent être composées et l'utilité semble alors être le seul concept quantitatif de la connaissance qui ignore la barrière de l'hétérogénéité tout en demeurant scalaire, mais cela malheureusement en référence à une seule personne par a).

d) l'axiome d'indifférence est vérifié (une personne est capable d'affirmer l'indifférence ou la non-préférence entre deux utilités de la même dimension ou de dimension différente).

e) l'axiome de correspondance aussi (si une personne affirme la non préférence de deux utilités de la même dimension, les valeurs correpondantes sont égales).

2) Introduisons le seul temps en formulant la question suivante : la disponibilité future d'une valeur était incertaine, en est-il de même de son utilité actuelle ? A cette question les deux réponses négatives et positives sont concevables. Mais dans les deux cas une caractéristique généralement relevable est l'aspect subjectif du concept d'utilité.

3) Introduisons l'incertitude pour des valeurs atemporelles. Nous supposerons, bien que cela puisse paraître artificiel, que la définition d'une utilité actuelle des disponibilités incertaines atemporelles de valeurs repose sur deux problèmes :

a) la recherche d'une traduction des incertitudes que contiennent ces valeurs atemporelles. C'est l'objet de toute randomisation, qu'elle soit dans notre cas dépendante ou indépendante du décideur. Nous avons abordé ce point en I.A.1.

b) la recherche d'une mesure, notre utilité, traduisant la perception de cette incertitude lors d'un problème de choix, cette perception s'exprimant par l'intermédiaire de relations de préférence sur les valeurs incertaines. Cette démarche peut elle-même prendre des allures très variées :

- recherche d'une mesure explicative de préférences observées.
- recherche d'une mesure "normative" de préférences (réelles ou artificielles) traduisant des règles de comportements rationnels.

- recherche d'une structure formelle et cohérente pour ces mesures, ces structures fournissant un cadre formel aux mesures explicatives ou normatives qui respecteraient une axiomatique compatible à celle sur laquelle est batie la structure. Cette dernière recherche est l'objet du paragraphe III A ci-dessous sur les éléments de la théorie de l'utilité.

Les clivages réalisés ci-dessus, bien qu'artificiels, ont l'énorme avantage de faire ressentir isolement l'importante question d'interprétation, (objectivité ou subjectivité) contenue dans la quantification de l'incertitude. Ainsi cette qualité d'objectivité ou de subjectivité de la mesure de l'incertitude ne prévaudra en rien dans la rationnalité du processus dans lequel elle s'inserre. De plus, elle permet de ne voir dans l'utilité, quantification des préférences, aucune influence sur le traitement de l'incertitude.

Si l'incertitude provient de la Nature, nous respectons par exemple par cette position le Principe de l'Infifférence de la Nature. Ce principe (parfois appelé Principe de la Raison Insuffisante) est une des bases de la logique de l'incertitude. Comme exemple où ce principe est largement mis en relief, citons le cas des jeux statistiques qui sont des jeux à deux personnes, et à somme nulle, un partenaire étant ledécideur supposé intelligent et l'autre étant la Nature supposée neutre (voir D. BLACKWELL et M.A. GIRSHICK (6)).

Cette conception de séparer la question de la traduction des incertitudes du problème de définitions des fonctions d'utilité a le mérite d'autoriser à considérer comme équivalentes les interprétations suivantes des probabilités (en tant que quantification de l'incertitude) :

- l'interprétation axiomatique des probabilités (BERNOUILLI, LAPLACE, KOLMOGOROV)
- l'approche naturelle ou normative en tant que lois de la structure de l'incertitude (le Principe de la Raison Insuffisante de BERNOUILLI, les critères de LAPLACE, le théorème de BAYES, les chaines de MARKOV, le principe du Minimax de WALD, le critère du Minimax Regret de SAVAGE, définition de la probabilité logique de CARNAP ...)

- l'approche objectiviste qui, en tant qu'objet de l'enseignement de la théorie des probabilités, formalise en un édifice mathématique une certaine connaissance expérimentale des phénomènes du hasard (GAUSS, POISSON, PEARSON, FISCHER, BOREL, GALTON, LEVY, DARNOIS ....)
- l'approche subjectiviste (ou personalistique) qui tente d'associer les mathématiques et le comportement humain en avenir incertain (VON NEUMANN, DREZE, DE FINETTI, SAVAGE ....). Dans ce cas utilité et incertitude temporelle sont formalisées simultanément.

D'une manière générale l'approche subjectiviste donne lieu à de nombreux types de modèles de choix. En matière de critère, c'est seulement pour cette approche que l'on peut trouver les fonctions d'utilité, ces dernières étant alors un élément important des expressions à optimiser pour réaliser le choix. Ainsi le chapitre III sera consacré aux critères de choix en présence de fonctions d'utilité.

I B - *Types de critères envisagés*.-

Dans ce qui suit nous dégagerons ainsi deux grands types de critères de choix :
- ceux qui sont utilisés par la théorie statistique de la décision sous sa forme classique. Le processus de choix comporte une expérimentation à la suite de laquelle le choix est formulé ; les probabilités introduites sont généralement objectives et il n'y a pas de fonction d'utilité au plein sens du terme, les critères de choix s'exprimant à l'aide de fonctions de risque.
- ceux qui sont utilisés par les théories subjectives de la décision inspirées de celle de SAVAGE comportant accessoirement des probabilités objectives ou des expérimentations.

Remarquons que nous laisserons totalement de côté, faute du temps nécessaire, un grand type de critères de choix : les critères de choix dynamiques dont les critères séquentiels sont des exemples.

Entre les deux types de critères précédemments choisis, les liens sont laches et pour construire une théorie générale de la décision dans l'incertitude il serait nécessaire de les resserer.

I - FONCTIONS DE DECISIONS OPTIMALES CLASSIQUES

La théorie générale de la décision élaborée par A. WALD (1) de 1939 à 1950 se base sur l'extension à tous les problèmes de choix du principe qu'une procédure statitiq ue doit être évaluée par l'examen de ses conséquences en diverses circonstances (ce principe étant initialement appliqué par NAYMAN et PEARSON à leur théorie des tests d'hypothèses). Le modèle mathématique de cette théorie de la décision est un cas particulier de celui de la théorie des jeux telle qu'elle a été introduite par E. BOREL en 1921 et généralisée par VON NEUHANN et O. MORGENSTERN (39) de 1928 à 1944. Les contributions au développement de la théorie des décisions statistiques ont été par ailleurs très nombreuses (voir l'importante bibliographie relevée par D. BLACKWELL et M.A. GIRSHICK (3)).

II A - Esquisse de la théorie de la décision de WALD.-

Les éléments définis à propos de cette esquisse seront repris tout au long des autres théories abordées.

II A.1.- Eléments de la théorie générale.-

La théorie de la décision repose sur les éléments formels suivants :
- un espace probabilisé (H,A_H,μ) où les éléments h de H sont couramment appelés états de la nature et la probabilité μ sur la tribu A_H, probabilité a priori des états de la nature.
- une variable aléatoire X à valeur dans R^ν dont la loi de probabilité dépend de $h \in H$, cette loi pouvant être définie par sa fonction de répartition F_h.
- une suite $\{X_i, i \geq 1\}$ de v.a indépendantes de même loi que X. Un n-échantillon Z est formé des n premiers termes de cette suite, $(X_1 \ldots X_n)$, dont la loi de répartition, dépendant également de h, sera noté G_h. Soit Z, l'ensemble des

valeurs de Z avec $Z \subset R^m$ $(m = n \times \nu)$.

- un ensemble $\mathcal{D}$ de décisions d
- un ensemble F de fonctions de décision f de Z dans $\mathcal{D}$.

Une fonction f associe à toute réalisation $(x_1,\ldots,x_n)$ du n-échantillon $(X_1 \ldots X_n)$ une décision d ($(x_1 \ldots x_n)$ étant défini dans R^m avec $m = n \times \nu$). Une telle fonction f, par analogie avec la théorie des jeux, est appelée fonction de décision pure. Remarquons que les fonctions f induisent sur $\mathcal{D}$ une mesure de probabilité dépendant de h.

- une famille Φ de fonctions de décision φ de Z dans Λ, Λ étant une famille de mesures de probabilité λ sur une tribu fixée $\mathcal{A}_{\mathcal{D}}$ de $\mathcal{D}$. Une fonction φ associe à toute réalisation $x = (x_1 \ldots x_n)$ du n-échantillon une mesure de probabilité $\lambda \in \Lambda$.

Les fonctions φ sont appelée fonctions de décisions mixtes (ou randomisées, ou aléatoires). L'ensemble F est un cas particulier de Φ en prenant pour Λ l'ensemble des mesures de proba ponctuelles $\mathcal{E}_d$, $d \in \mathcal{D}$.

- une fonction de perte L de $\mathcal{H} \times \mathcal{D}$ dans R^+. Cette fonction qui peut porter d'autres noms suivant les contextes (fonction de coût, gain ....) est elle-même une v.a, les éléments h et d étant choisis aléatoirement conformément à leur définition.
- une fonction de risque ρ définie par $\rho(\mu,\varphi) = E(L)$, $E(L)$ étant l'espérance mathématique de L. En limitant l'exposé au cas où les mesures μ et G_h introduites ci-dessus admettent des densités, l'expression de la fonction de risque est :

$$\rho(\mu,\varphi) = \int_{\mathcal{H}\times Z\times\mathcal{D}} L(h,d)\ \mu(h)\ dhg_h(z)\ dxd\varphi_z(d)$$

avec $\mu(h)$: densité correspondante à la probabilité μ

$g_h(z)$: densité correspondante à la fonction de répartition G_h

Cette expression peut s'écrire sous diverses formes :

$$* \qquad \rho(\mu,\psi) = \int_H R(h,\psi)\ \mu(h)\ dh$$

avec $R(h,\psi) = \int_{Z\times D} L(h,d)\ g_h(z)\ dx\ d\psi_z(d)$ qui est la fonction de risque associée à ψ conditionnellement à la réalisation de $h \in H$. On peut encore écrire :

$$R(h,\psi) = \int_Z E\psi_z\{L(h,d)\}\ g_h(z)\ dz$$

où $E\psi_z$ est l'espérance mathématique avec la mesure ψ_z.

$$** \qquad \rho(\mu,\psi) = \int_Z r(z,\psi)\ P_\mu(z)\ dz$$

avec $r(z,\psi) = \int_{H\times D} L(h,d)\ \mu_z(h)\ dh\ d\psi_z(d)$ où $\mu_z(h) = \dfrac{\mu(h)\ g_h(z)}{\int_H \mu(h) g_h(z) dh}$ est la densité de probabilité à postériori sur A_H conditionnellement à la réalisation x du n-échantillon et $P_\mu(z) = \int_H \mu(h)\ g_h(z)dh$, densité de probabilité sur Z conditionnellement au choix de la densité à priori μ. La fonction $r(z,\psi)$ est la fonction de risque associée à ψ conditionnellement à la réalisation de z ou fonction de risque à postériori.

II A.2.- *Critères de choix.*-

La résolution d'un problème de décision consiste à choisir une fonction de décision ψ dans Φ (ou f dans F dans le cas des décisions pures, cas qui ne sera plus envisagé par la suite puisqu'il est inclu dans celui des décisions mixtes) relativement à un critère de choix. Il existe essentiellement trois critères de choix classiques pour lesquels les propriétés des fonctions de décision ont été étudiées avec des hypothèses relatives aux divers éléments de la théorie peu restrictives. Il s'agit :

- du critère d'admissibilité ou axiome de préférabilité absolue

- du critère de l'espérance ou critère de Bayes
- du critère du minimax.

Les deux derniers critères sont compatibles avec le premier.

II A.2.1.- *Solutions admissibles*.-

Définition 1.- soit ψ, $\psi' \varepsilon \Phi$.

a) <u>ψ est préféré à ψ'</u> ($\psi \succcurlyeq \psi'$) si :

$$R(h,\psi) \leqslant R(h,\psi') \qquad \forall h \in H$$

b) <u>ψ est équivalent à ψ'</u> ($\psi \asymp \psi'$) si l'on a $\psi \succcurlyeq \psi'$ et $\psi' \succcurlyeq \psi$ simultanément.

c) <u>ψ est strictement préféré à ψ'</u> ($\psi \succ \psi'$) si l'on a $\psi \succcurlyeq \psi'$ sans que $\psi \asymp \psi'$

La relation ($\succcurlyeq$) est un préordre de préférence partiel.

Définition 2.- Une fonction de décision ψ_o est une <u>solution admissible</u> si elle est maximale pour ce préordre. L'ensemble des solutions admissibles est $\mathcal{A}$.

Définition 3.- une partie $C \subset \Phi$ est une <u>classe complète</u> si :

$$\forall \psi \notin C, \quad \exists \psi_o \subset C : \psi_o \succ \psi$$

Définition 4.- une <u>classe complète minimale</u> est une classe complète C dont aucune de ses parties qui diffèrent de C n'est une classe complète.

Les propriétés suivantes découlent aisément de ces définitions :

- toute classe complète contient toutes les solutions admissibles.
- une classe complète peut contenir des fonctions de décision non admissibles
- s'il existe une complète minimale, elle coïncide avec $\mathcal{A}$
- si $\mathcal{A}$ est une classe complète, c'est une classe complète minimale.

Ces propriétés justifient la recherche de la solution optimale dans $\mathcal{A}$, conformément à l'axiome de préférabilité absolue, bien que $\mathcal{A}$ soit souvent difficile à exhiber directement.

II A.2.2.- *Solution de Bayes*.-

Définition 5.- une fonction de décision φ_μ est une solution de Bayes par rapport à μ, mesure de probabilité à priori si :

$$\rho(\mu,\varphi_\mu) \leq \rho(\mu,\varphi) \qquad \forall \rho \in \Phi$$

Définition 6.- une fonction de décision φ est une solution de Bayes s'il existe une mesure de probabilité à priori μ telle que φ soit solution de Bayes p. r. à μ.

En écrivant $\rho(\mu,\varphi) = \int_Z r(z,\Psi) \; P_\mu(z)\, dz$, les solutions de Bayes s'interprètent comme les fonctions de décisions qui minimisent l'espérance mathématique du risque à postériori.

II.A.3.- *Solutions minimax*.-

Les fonctions de décisions suivantes sont utilisées lors des problèmes de choix en incertitude pure, c'est-à-dire dans des situations où l'on se refuse l'introduction d'une probabilité à priori μ.

Définition 7.- une fonction de décision ψ_o est une solution minimax si :

$$\sup_{h\in H} R(h,\varphi_o) \leq \sup_{h\in H} R(h,\varphi) \qquad \forall \varphi \in \Phi$$

II A.3.- *Propriétés des fonctions de décision*.-

Les propriétés qui seront énoncés ci-dessous et qui constituent l'essentiel de la théorie générale de la décision de WALD nécessitent la réalisation d'un certain nombre d'hypothèses de nature topologique sur les divers éléments introduit en II. A.1.

Hypothèse 1 : sur $\mathcal{H}$, il est possible de définir une distance dn telle que

$d_n(h_1,h_2) = \sup_B |P(B|F_{h_1}) - P(B|F_{h_2})|$, où $n > 0$ est quelconque, B est un borélien de R_n et $P(B|G_h)$ la probabilité pour qu'un n-échantillon des f.r. G_h prenne ses valeurs dans B.

H est alors supposé séparable avec la métrique d_n.

Hypothèse 2 : $\mathcal{H}$ est une v.a. discrète ou absolument continue, $\forall h \in \mathcal{H}$

Hypothèse 3 : $L(h,d)$ est une fonction bornée de h et de d

Hypothèse 4 : sur $\mathcal{D}$, il est possible de définir une distance, en posant

$$\delta(d_1,d_2) = \sup_{h \in \mathcal{H}} |L(h,d_1) - L(h,d_2)|$$

$\mathcal{D}$ est alors supposé compact avec la métrique δ.

Hypothèse 5 : si Φ est l'ensemble des fonctions des décisions dont dispose le statisticien et E l'ensemble de toutes les fonctions de décisions possibles $(\Phi \subset E)$, la convergence dans E sera définie par :

$$\psi_m \to \psi_o \quad \text{si} \quad \forall z \in Z \quad \text{et} \quad \forall D \in A_{\mathcal{D}}$$

on a :

$$\psi_m(z)[D] \to \psi_o(z)[D].$$

Alors :

- ψ est supposée convexe
- ψ est une partie fermée de E avec la définition précédente de la convergence.

Hypothèse 6 : l'espace H est compact avec la métrique introduite dans l'hypothèse 1.

Hypothèse 7 : la fonction $R(h,\psi)$ est bornée de $h \in H$, $\forall \psi \in \Phi$. Si les hypothèses 1 à 5 sont vérifiées, on démontre :

Théorème 1.- Φ est un espace compact (selon convergence définie dans l'hypothèse 5).

Théorème 2.- le problème de décision, considéré comme un jeu à deux personnes de somme nulle admet un équilibre, soit

$$\sup_{\mu} \inf_{\psi} \rho(\mu,\psi) = \inf_{\psi} \sup_{\mu} \rho(\mu,\psi)$$

où $\sup\limits_{\mu}$ est pris p.r. à toutes les mesures de probabilité sur H et $\inf\limits_{\psi}$ est pris p.r. aux éléments de Φ.

Théorème 3.- à toute mesure à priori μ sur H, on peut associer une fonction de décision $\psi\mu \in \Phi$, solution de Bayes p.r. à μ. ($\rho(\mu,\psi_{\mu}) \leqslant \rho(\mu,\psi)$ $\forall \psi \in \Phi$).

Théorème 4.- il existe une fonction $\psi_o \in \Phi$ minimax, soit

$$\sup_{h} R(h,\psi_o) \geqslant \sup_{h} R(h,\psi) \qquad \forall \psi \in \Phi$$

Théorème 5.- Si μ_o est une mesure de probabilité à priori la plus défavorable (soit $\inf\limits_{\psi} \rho(\mu_o,\psi) \leqslant \inf\limits_{\psi} \rho(\mu,\psi) \forall \mu$).

alors toute solution minimax est solution de Bayes p.r. à μ_o.

Avec les hypothèses 1 à 6 on a de plus :

Théorème 6.- il existe une mesure de probabilité à priori la plus défavorable.

En ajoutant l'hypothèse 7 :

Théorème 7.- la classe des solutions de Bayes forme une classe complète.

II A.4.- *Prolongements de la théorie générale*.-

A partir de la théorie de WALD schématisée ci-dessus, les directions d'approfondissement ou d'extension sont très variées et nombreuses, comme le témoigne, par exemple, l'importante bibliographie relevée par D. BLACKWELL et M.A. GIRSHICK (6). Nous allons suivre, sans prétendre à aucune exhaustivité, quelques uns de ces axes dans la mesure où ils concernent le critère de choix.

II B - *Théorie générale et théories statistiques classiques*.-

La théorie de l'estimation et la théorie des tests ne sont que des cas particuliers du problème général de la décision. L'analogie de leurs structures est schématisée ci-dessous.

II B.1.- *Fonctions de décision et estimation*.-

Eléments de la théorie de la décision

Ensemble $\mathcal{H}$ des états de la nature h

loi de probabilité à priori μ sur $\mathcal{H}$

Variable aléatoire X sur R de fonction de répartition F_h, $h \in \mathcal{H}$

Echantillon de taille n

Ensemble $\mathcal{D}$ des décisions possibles d.

Ensemble Φ des fonctions de décisions non aléatoires :

$$\varphi : R^m \to \mathcal{D}$$

Ensemble ψ des fonctions de décisions aléatoires

$$\psi : R^m \to \Lambda$$

Λ = ensemble des mesures sur $\mathcal{D}$

Eléments de la théorie de l'estimation

Ensemble Θ des valeurs possible θ du paramètre

Loi de probabilité à priori μ sur Θ

Variable aléatoire X sur R de fonction de répartition $F(x,\theta)$, $\theta \in \Theta$

Echantillon de taille n

Ensemble $\mathcal{D}$ des choix pour les valeurs de θ

. Dans l'estimation ponctuelle $\mathcal{D}$ est une correspondance bijective avec Θ

. Dans l'estimation par intervalle, $\mathcal{D}$ est une famille d'intervalles $I \subset \Theta'$, chaque intervalle étant repéré par son centre θ^* et sa longueur 2ℓ.

. En estimation ponctuelle :

$$\varphi : R^m \to \Theta$$

. En estimation par intervalle :

$$\varphi : R^m \to \mathcal{P}(\Theta)$$

Ensemble ψ des fonctions de décision aléatoires

$$\psi : R^m \to \Lambda$$

Fonction de perte L :

$$L : \mathcal{H} \times \mathcal{D} \to R^{+}$$

. En estimation ponctuelle :

$L(\theta,\theta^{*})$ avec $\theta \in \Theta$

$\theta^{*} \in \mathcal{D} = \Theta$

. En estimation par intervalle ·

$L(\theta,I) = L(\theta,\theta^{*},\ell)$ avec $\theta \in \Theta$, $I \in \mathcal{D}$

($I = I(\theta^{*},\ell)$ avec θ^{*} = centre et ℓ = longueur)

Parmi les questions soulevées par le problème de l'estimation signalons en deux

- l'utilisation de probabilités à priori est refusée, par principe, par certains, bien que la théorie de la décision puisse éclairer sur la non contradiction profonde entre les divers points de vue.
- la définition d'une fonction de perte L est particulièrement délicate en soi (celle-ci représentant des coûts d'erreurs souvent peu cernables dans la pratique) et de plus les hypothèses H3 et H4 ne sont pas respectées avec les fonctions habituelles.

Celles-ci, sont de la forme suivante (voir D. BLACKWELL et M.A. GIRSHICK- (3) ch. 11 - où leurs propriétés sont étudiées) :

* $L(\theta, \theta^{*}) = \lambda(\theta)|\theta - \theta^{*}|$ avec $\lambda(\theta) > 0$

  (perte dépendant de l'erreur absolue)

* $L(\theta,\theta^{*}) = \lambda(\theta)(\theta-\theta)^{2}$ avec $\lambda(\theta) > 0$

  (perte quadratique)

pour l'estimation ponctuelle et, pour l'estimation par intervalles :

* $L(\theta,I) = 0$ si $|\theta-\theta^{*}| \leq \ell$

  $= 1$ si $|\theta-\theta^{*}| > \ell$

* $L(\theta,I) = 0$ si $|\theta-\theta^{*}| \leq \alpha\theta^{*}$

  $= 1$ si $|\theta-\theta^{*}| \leq \alpha\theta^{*}$

II B.2.- Fonctions de décision et tests.-

Eléments de la théorie de la décision	Eléments de la théorie des tests
Ensemble $\mathcal{H}$ des états de la nature h	Ensemble Θ des valeurs possibles pour le paramètre θ. Une hypothèse H est un élément d'une partition Π_θ de Θ
loi de probabilité à priori μ sur $\mathcal{H}$	Loi de probabilité à priori μ sur $\mathcal{H}$
Variable aléatoire X de fonction de répartition F_h, $h \in \mathcal{H}$	Variable aléatoire X de fonction de répartition $F(x,\theta)$, $\theta \in \Theta$
Echantillon de taille n	Echantillon de taille n
Ensemble $\mathcal{D}$ des décisions possibles d	Ensemble $\mathcal{D}$ des décisions, d_i étant la décision d'accepter l'hypothèse H_i ($\mathcal{D}$ est en correspondance bijective avec une partition Π_Θ de Θ)
Ensemble Φ des fonctions de décision non aléatoire φ : $\varphi : R^m \to \mathcal{D}$	Ensemble Φ des fonctions de décision non aléatoire φ : $d = \varphi(x)$
Ensemble ψ des fonctions de décision aléatoire ψ $\psi : R^m \to \Lambda$	Ensemble ψ des fonctions de décision aléatoire ψ : $1 = \psi(x)$
Fonction de pertes $L(h,d)$: $\mathcal{H} \times \mathcal{D} \to R^+$	Fonction de perte $L(\theta,d)$ $\Theta \times \mathcal{D} \to R^+$

Comme pour l'estimation, l'utilisation de probabilités à priori n'est pas admise par tous et il existe ainsi des théories, celles de NEYMAN et PEARSON, qui s'effacent de résoudre le problème sans y faire appel, tout au moins en apparence. Les formes retenues pour les fonctions de pertes sont diverses ; en voici par exemple deux (voir E.L. LEHMANN (24)) :

α) $L(\theta,d_i) = L_i(\theta)$ $(i = 0,1)$ avec $L_o(\theta)$ nul si $\theta \leq \theta_o$ et strictement croissant si $\theta \geq \theta_o$, et $L_1(\theta)$ strictement décroissant si $\theta \leq \theta_o$ et nul si $\theta \geq \theta_o$ dans le test de l'hypothèse $H_o : \theta \leq \theta_o$ entre l'alternative $H_1 : \theta > \theta_o$.

α) $L(\theta,\underline{\theta})$ non négative, non décroissante par rapport à $\underline{\theta}$ pour $\underline{\theta} < \theta$ et nulle pour $\underline{\theta} \geq \theta$ dans le cas particulier de la détermination d'un intervalle de confiance (dont $\underline{\theta}$ est la borne supérieure).

II C - RESTRICTION DE LA THEORIE GENERALE AU CAS FINI.

La restriction du problème général de décision au cas où H ou D est fini donne lieu à d'abondants approfondissements, l'analyse du problème de décision avec de telles hypothèses restrictives étant justifiée tant par la correspondance avec des théories classiques (comme les tests) que par la vérification de ces hypothèses dans de très nombreux modèles de prise de décision élaborés dans la pratique (voir, comme illustration, (29),(34),(30) ou (10)). L'étude particulière des propriétés des fonctions de décision avec de telles hypothèses est l'objet des ouvrages de D. BLACKWELL et M.A. GIRSHICK (6) et de H. CHERNOFF et L.E. MOSES (7).

En matière de critère de choix plus spécialement, l'hypothèse où X est fini permet de faire simplement la liaison entre le problème de décision et ceux plus généraux, d'optimisation. C'est ce que font D. BLACKWELL et M.A. GIRSHICK (6) en employant le langage de la théorie des jeux. Signalons, à propos de leur line deux éléments de la procédure de choix non encore évoqués jusqu'ici :

- dans la théorie générale exprimée au II B, le n-échantillon $(X_1 \dots X_n)$ est de taille fixe. Si n n'est plus fixe, la procédure décrite devient séquentielle (le critère de choix devant alors comporter une règle d'arrêt de l'expérimentation). Ce type de procédure de choix, largement explorée par WALD, et reprise dans (6) sous forme de jeu séquentiel, ne sera pas approfondi dans notre démarche.

- le développement de la théorie des jeux fait appel à une nouvelle fonction non introduite par WALD, l'utilité, dont le rôle en matière de décision est important. Elle fera l'objet du chapitre III.

II C.1.- *Exemple de cas fini*.-

Ci-dessous nous allons illustrer le cas fini en présentant, pour une structure particulière de problème de décision telle qu'elle peut être rencontrée en économie, des propriétés de solutions optimales relatives à divers critères de choix classiques. Ces résultats sont tirés de P. MALGRANGE (29).

Dans le modèle particulier étudié, l'incertitude introduite se limite à la considération de probabilités discrètes attachées aux états de la nature (ceux-ci étant en nombre fini) et le choix se fait sans exprimentation. Les démonstrations sous-jacentes aux propriétés présentées sont réalisées à partir de propriétés utilisés de manière générale dans les problèmes d'optimisation. L'intérêt d'un tel modèle est de refleter une démarche proche de celles suivies par certains auteurs postérieurs à WALD (comme SAVAGE) dans leur théorie de la prise de la décision. On reconnaîtra néanmoins dans ce qui suit un certain nombre de propriétés énoncées ci-dessus dans la théorie générale.

II C.1.1.- *Données de modèles* :

- $\mathcal{H} = \{h_i, i=1,p \text{ avec } p \geq 2\}$, l'ensemble fini des états possibles de la nature
- $\mathcal{D}$, ensemble des décisions réalisables, supposé compact
- G, fonction de gain de $\mathcal{H}\times\mathcal{D}$ dans $\mathcal{R}^+$ (G ayant un rôle identique à la fonction de perte L, au sens du préordre des préférences près).

Posons : $G_i(d) = G(h_i,d)$, fonction continue en d $\forall i$, et

$$G(d) = \begin{pmatrix} G_1(d) \\ \vdots \\ G_p(d) \end{pmatrix} \varepsilon\ \mathcal{R}^p$$

- γ, fonction de choix, de R^p dans R. A chaque critère de choix classique correspond une forme de γ :

  * critère de l'espérance mathématique : $\gamma = \gamma_\mu$ avec

$$\gamma_\mu \circ G(d) = \sum_{i=1}^{n} \mu_i \; G_i(d)$$

avec $\mu = \{\mu_i \mid i=1, p \; ; \; \mu_i \geq 0 \; ; \; \sum_{i=1}^{p} \mu_i = 1\}$

  * critère du maximin : $\gamma = \gamma_{Mm}$ avec :

$$\gamma_{Mm} \circ G(d) = \inf_i G_i(d)$$

  * critère du minimax regret : $\gamma = \gamma_{MR}$ avec :

$$\gamma_{MR} \circ G(d) = \inf_i (G_i(d) - \sup_{d \varepsilon \mathcal{D}} G_i(d))$$

  * critère du maximum de l'espérance minimum : $\gamma = \gamma_{Mem}$ avec :

$$\gamma_{Mem} \circ G(d) = \inf_{\mu \varepsilon M_p} \{ \sum_{i=1}^{p} \mu_i \; G_i(d)\}$$

avec $\mu = \{\mu_i \mid i = 1,p\}$

M_p = sous-ensemble compact du simplexe T_p

- Une décision d^* est optimale dans $\mathcal{D}$, relativement au critère de choix défini par γ, si : $\gamma \circ G(d^*) \geq \gamma \circ G(d)$

II C.1.2.- Solutions admissibles.-

Une décision d est admissible (soit A leur ensemble) si elle est maximale pour le préordre de préférence partiel défini par : d_1 est préféré à d_2 si

$G(d_1) \geqslant G(d_2)$ dans $\mathcal{R}^p$ (Notation : $d_1 \succcurlyeq d_2$)

A étant difficilement étudiable directement, on recherchera dans une partie seulement de A, les solutions optimales en employant un des critères définis ci-dessus comme critère de choix supplémentaire (ces critères ayant été choisis pour respecter l'axiome de préférabilité défini par le préordre précédent).

II C.1.3.- Solution de Bayes.-

Une décision est solution de Bayes par rapport à μ si elle optimale par rapport au critère de l'espérance mathématique, c'est-à-dire si elle rend maximum $\gamma_\mu \circ G(d)$ défini pour le système de probabilité discrete μ. (Une décision sera dite solution de Bayes s'il existe un μ tel qu'elle soit solution de Bayes par rapport à μ).

Posons : * $\mathcal{B}_\mu$ = ensemble des décisions de Bayes p.r. à $\mu = \{d_\mu\}$

* $b(\mu) = \sup\limits_{d \in \mathcal{D}} (\gamma_\mu \circ G(d)) = \gamma_\mu \circ G(d_\mu)$

* $\mathcal{T}_p = \{\mu\}$ = ensemble des probabilités à priori sur $\mathcal{H}$

* $\mathcal{B}$ = ensemble des décisions de Bayes = $\{d_\mu, \forall \mu \in \mathcal{T}_p\}$

les propriétés des solutions optimales d_μ et de la fonction $b(\mu)$ sont :

Propriétés 1.- - $\forall \mu \in \mathcal{T}_p$, β_μ est un ensemble non vide et $\beta_\mu \cap A$ est non vide.

- si $\mu \in \mathring{\mathcal{T}}_p$ avec $\mathring{\mathcal{T}}_p = \{\mu \mid \mu \in \mathcal{T}_p$ avec $\mu_i > 0\ \forall i\}$ alors $\beta_\mu \subset A$.

Propriété 2.- si $\mathcal{D}$ est convexe et si $G_i(d)$ sont concaves sur $\mathcal{D}$, alors $\exists \mu : A \subset \beta_\mu$

Propriété 3.- - la correspondance $\mu \to \beta_\mu$ est semi-continue supérieurement

- la fonction $b(\mu)$ est convexe et continue

- $\mathcal{B}$ est fermé
- si $G_i(d)$ sont strictement concaves :

* β_μ est une fonction continue sur $\mathcal{T}_p$
* $b(\mu)$ est différentiable sur $\mathcal{T}_p$, sa différentielle étant le vecteur $G(d_\mu)$.

On y retrouve, entre autre, l'existence d'une solution de Bayes p.r. μ pour toute probabilité à priori μ, ainsi que la compatibilité du critère de Bayes avec l'axiome de préférabilité.

II C.1.4.- Solution maximin.-

Une décision est une solution maximin si elle est optimale par rapport au critère du maximin, c'est-à-dire si elle rend maximum $\gamma_{Mm} \circ G(d)$. Si les $G_i(d)$ sont concaves on retrouve la propriété que les solutions maximin sont des solutions de Bayes par rapport à une probabilité à priori la plus défavorable (voir théorème 5). Soit d'une manière plus précise :

* $\mathcal{D}^*$ = ensemble des solutions maximin ; décisions d solution du programme d'optimisation I :

$$I = \begin{cases} \text{Max } f(d) \\ f(d) = \gamma_{Mm} \circ G(d) \\ d \in \mathcal{D} \end{cases}$$

* $\mathcal{L}^*$ = ensemble des solutions points selle, $(\mu^*, d^*) \in \mathcal{T}_p \times \mathcal{D}$ tels que

$$II : \begin{cases} \gamma_{\mu^*} \circ G(d) \leq \gamma_{\mu^*} \circ G(d^*) \leq \gamma_\mu \circ G(d^*) \\ \forall \mu \in \mathcal{T}_p, \quad \forall d \in \mathcal{D} \end{cases}$$

* $\mathcal{T}_p^*$ = ensemble des solutions les plus défavorables, décisions d solution de :

$$\text{III} \quad \begin{cases} \text{Min } f(\mu) \\ f(\mu) = \sup\limits_{d \varepsilon \mathcal{D}} (\gamma_\mu \circ G(d)) \\ \mu \; \varepsilon \; \mathcal{T}_p \end{cases}$$

Alors si $\mathcal{L}^*$ n'est pas vide (il suffit que $G_i(d)$ soient concaves par exemple) :

- $\mathcal{L}^* = \mathcal{T}^* \times \mathcal{D}^*$
- $\mathcal{D}^* \subset \mathcal{B}_{\mu^*} \qquad \forall \mu^* \; \varepsilon \; \mathcal{T}_p$
- $\mathcal{L}^*$ est compact

Précisons le comportement du décideur suivant la solution adoptée, le décideur ayant un comportement Γ s'il a un préordre de préférence totale sur les décisions représentable par l'application γ tel que : $d_1 \succsim d_2 \Leftrightarrow \gamma(G(d_1)) \geq \gamma(G(d_2))$. Ainsi, les propriétés ci-dessus montrent que le décideur à comportement maximin choisira la même décision que le décideur à comportement de Bayes qui aura adopté, fortuitement, la probabilité a priori sur les états de la nature la plus défavorable ; mais d'une manière générale ils ne classeront pas toutes les décisions dans le même ordre et ainsi on ne peut interpréter le comportement minimax comme un comportement de Bayes particulier. Cette remarque est à replacer dans le contexte plus général de l'utilité où un comportement maximin suppose une perception ordinale de l'utilité (les préférences sur les décisions étant non plus liées aux gains mais à une fonction strictement croissance quelconque de ceux-ci) tandis que le comportement de Bayes suppose une perception cardinale de l'utilité (utilité définie à une fonction linéaire strictement croissante près) -voir le chapitre III sur l'utilité.

II C.1.5.- Solution minimax regret.

Une décision minimax regret est optimale par rapport au critère correspondant, ce qui implique qu'elle soit maximin pour des fonctions de gain $G_i'(d) = G_i(d) - \sup_{d \in \mathcal{D}} G_i(d)$. Les propriétés de concavités et de continuité de $G_i(d)$ étant nécessaires et suffisantes à celles de $G_i'(d)$, les propriétés des solutions maximin obtenues pour $G_i(d)$ sont analogues pour les solutions minimax regret.

II C.1.6.- *Solution du maximum de l'espérance mathématique minimum.*-

Dans ce cas, le décideur ne dispose que d'une famille M_p de distribution à priori possibles μ, M_p étant donné et supposé compact. Une décision est maximum de l'espérance mathématique minimum si, conformément au critère correspondant, elle maximise la fonction $\theta(d)$:

$$\theta(d) = \inf_{\mu \in M_p} \{ \sum_{i=1}^{n} \mu_i \, G_i(d) \}$$

Le comportement du décideur en présence de M_p supposent que ses préférences soient donc exprimés par $\theta(d)$. Remarquons

- si $M_p \equiv \mu_o$, le décideur à un comportement de Bayes p.r. à μ_o
- si $M_p \equiv T_p$, le décideur à un comportement maximin.

A partir des solutions aux trois problèmes suivants, les propriétés du maximin se généralisent :

Solutions max. de l'esp. math. mini : I
$$\begin{cases} \text{Maximiser } \theta(d) \\ \theta(d) = \gamma_{Mem} \circ G(d) \\ d \in \mathcal{D} \end{cases}$$

Solutions point selle II
$$\begin{cases} \text{trouver } (\mu^*, d^*) \\ \gamma_{\mu^*} \circ G(d) \leq \gamma_{\mu^*} \circ G(d^*) \leq \gamma_{\mu} \circ G(d^*) \\ \forall \mu \in M_p, \quad \forall d \in \mathcal{D} \end{cases}$$

Solution les plus M_p-défavorables : III
$$\begin{cases} \text{Minimiser } \alpha(\mu) \\ \alpha(\mu) = \sup\limits_{d \in \mathcal{D}} (\gamma_\mu \circ G(d)) \\ \mu \in M_p \end{cases}$$

Si $\mathcal{D}^* M_p$, $L^* M_p$, $F^* M_p$ sont les ensembles de solutions I, II et III respectivement, si $B\mu$ est l'ensemble des solutions de Bayes p.r. à μ, si $U(d)$ est l'ensemble de $\mu \in M_p$ minimisant $\gamma_\mu \circ G(d)$ à d donné et si $L^* M_p$ n'est pas vide, alors :

* $L^* M_p = F^* M_p \times \mathcal{D}^* M_p$
* $L^* M_p$ est compact
* $\mathcal{D}^* M_p \subset B\mu^* \qquad \forall \mu^* \in F^* M_p$
* $F^* M_p \subset U(d^*) \qquad \forall d^* \in \mathcal{D}^* M_p$

II C.1.7.- <u>Remarques sur ces solutions</u>.-

* La nature finie des éléments envisagés ci-dessus ainsi que les hypothèses de continuité, compacité et convexité adoptées, permettent pour la recherche pratique des solutions optimales relatives à un de ces critères, d'utiliser des algorithmes classiques d'optimisation. (méthode de gradient par exemple).
* Les propriétés énoncées ci-dessus se généralisent à des théories de type dynamique dans lesquels les états de la nature sont remplacés par des chemins d'états de la nature et les décisions par des stratégies. Un exemple en est donné dans [5]. L'utilisation séquentielle d'un critère de choix se base sur le respect du principe d'optimalité, principe naturel en avenir certain et qui doit être généralisé en avenir incertain. Sur énoncé synthétique est : "une stratégie optimale ne peut être formée que de sous stratégies optimales".

Notons que, si le critère d'espérance peut vérifier un tel principe (programmation dynamique stochastique) et s'il en est de même du critère du maximin, le minimax regret ne correspond pas à un comportement dynamisable, de même qu'en général le maximum de l'esperance minimum.

II D - CRITERE DE CHOIX ET FONDEMENTS DE L'INFERENCE STATISTIQUE.-

Le problème soulevé par d'adoption d'un critère de choix particulier dans une théorie de décision statistique est une question traditionnellement soulevée en inférence statistique et n'est pas spécifiquement lié à la formalisation de l'inférence en terme de décision. L'étude générale d'une théorie de la décision comme celle d'un critère de choix, élément particulier de la précédente, s'intéresse aux décisions optimales sous deux aspects :

- celui de leurs propriétés mathématiques qui doivent être telles que l'analyse et la résolution effective du problème de choix soient possibles.
- celui de leur interprétation vis-à-vis du comportement du décideur.

Nous allons examiner des concepts de base qui, dans l'optique du premier point, guident l'adoption d'un critère. Pour cela nous suivons l'analyse de J. HAJEK (21) sur les concepts de base des statistiques en utilisant les notations définies en II A à propos de la théorie générale de WALD.

Si la probabilité à priori μ sur les états de la nature h est connue ; tous les statisticiens admettent que le choix d'une décision connaissant la réalisation z du n-échantillon Z doit se faire sur la base de la distribution P_μ sur Z, distribution connue conditionnellement à l'adoption de μ. Cette unanimité est liée au fait qu'il n'existe alors qu'une seule distribution sur l'ensemble sur $\mathcal{H} \times Z$ de telle sorte que le problème soit transféré du domaine des statistiques à celui de la théorie des probabilités. Dans ce cas, la fonction de risque de Bayes $r(z,\varphi)$ est mathématiquement naturelle. Ce n'est que lorsque μ n'est pas parfaitement connu que les divergences existent.

Dans une situation non bayesienne bien spécifique, un type de solutions nous est fourni par la théorie du minimax développée par A. WALD (40). Mais les principes à adopter pour justifier de la structure globale d'une procédure d'inférence statistique sont nombreux.

Citons-en deux, la conditionalité et la vraissemblance avant d'examiner en détail deux cas particuliers aux chapitres suivants.

II D.1.- CONDITIONALITE.-

Le principe de conditionalité s'introduit à partir du caractère conditionel de la prise de décision dans la mesure où l'on envisage des fonctions de décisions f construites sur un ensemble de résultats d'expérience Z. Pour raisonner dans un cas un peu plus général, supposons que f est une application de $\mathcal{T}$ dans $\mathcal{D}$, $\mathcal{T}$ étant l'ensemble des statistiques $\tau = t(Z)$. Nous ne considérerons ci-dessous qu'un cas particulier de conditionalité en supposant de plus que la statistiq ue envisagée est ancillaire (rappelons qu'une statistique $v = u(Z)$ est ancillaire par rapport à un paramètre h **si sa distribution est indépendante de** ce paramètre h) ; ceci limite le champ de l'analyse de la conditionalité mais notre analyse des principes de choix ne saurait bien évidemment être exhaustive - par exemple, nous n'envisagerons pas la procédure conditionnelle qui reposerait sur une statistique suffisante pour un paramètre de nuisance comme cela a pu être utili utilisé avec succès dans la construction de test similairesles plus puissants (E.L. LEHMANN (24)). Retenons l'énoncé général suivant de la conditionalité :

Principe de conditionalité :

Etant donné sur l'espace d'échantillonnage Z une statistique $U = u(Z)$ ancillaire par rapport à $h \in \mathcal{H}$, la résolution du problème d'inférence doit être basée sur la distribution conditionnelle $g_h(dz|U = u)$, c'est-à-dire que les distributions $g_h(dz|U = u')$ pour $u' \neq u$ et $\mathcal{P}(du)$ ne doivent pas être prises en compte.

(on suppose, pour l'écriture, que les mesures de probabilités admettent des densités). Remarquons que cet énoncé n'est pas en contradiction avec les procédures de Bayes, pour celles, toutefois, où la distribution a priori μ est indépendante de l'expérimentation. Interprétons en termes de risque cet énoncé général afin de satisfaire une des préoccupations de la théorie générale de la décision :

Autre énoncé du principe de conditionalité :

Une fonction de décision f définie sur Z doit être associée au même risque que celle définie sur la restriction de Z définie par $U = u$, pour $U = u(Z)$ une statistique ancillaire, le transfert de risque n'étant possible qu'une fois l'expérience réalisée et u observé.

Cet énoncé particulier s'écrit : $R(h,U,u) = R(h)$ avec :

- $R(h,U,u) = \int_Z L(h,f(z))\, g_h(z|U = u)dz$
- $R(h) = \int_Z L(h,f(z)\, g_h(z)dz$

Cet énoncé ne stipule pas pour autant quelle statistique ancillaire doit être choisie par batir la conditionalité et cela d'autant qu'il est en général impossible de définir une statistique ancillaire maximale.

Par nature, cette forme est incompatible avec une procédure à risque Bayesien puisqu'elle nie l'intéret de toute randomisation des fonctions de décision. Considérons une restriction de ce principe :

Principe de conditionalité correcte :

Une fonction de décision f définie sur Z et celle définie sur $U^{-1}(u)$ doitvent être telles que le risque vérifie :

$$R(h,f,U,u) = R(h,f)\, b(U,u)$$

Ce principe restrictif n'est pas trop contraignant puiqu'il est applicable aux exemples généraux servant de support au principe de conditionalité en présence de statistiques ancillaire. Ces hypothèses permettent d'obtenir une propriété (théorème 1) et une condition d'existence (théorème 2).

<u>*Théorème*</u> *1*.- si une procédure de choix vérifie le principe de conditionalité correcte relativement à une statistique ancillaire U , alors

$$E(\sup_h \mathcal{R}(h,U,u)) = \sup_h \mathcal{R}(h)$$

<u>*Théorème*</u> *2*.- soit $\mathcal{S} = \{s\}$ un groupe de transformation bijective de Z sur lui-même et soit la famille $\{P_s\}$ générée par la distribution de probabilité P sur Z (P est ci-dessus la distribution de fonction de répartition G_h). Supposons que la fonction de perte L soit $\mathcal{S}$-invariante, c'est-à-dire telle que :

$$L(rs,f(r(z))) = L(s,f(z)) \qquad \forall r,s \in \mathcal{S}$$
$$\forall z \in Z$$

Alors la conditionalité relative à la statistique ancillaire U construite sur les résultats $\mathcal{S}$-invariants est correcte, soit $\mathcal{R}(s) = \mathcal{R}$

$$\mathcal{R}(s,U,u) = \mathcal{R}(U,u)$$

$$= \int_Z L(s,f(r^{-1}(z)))\ p(z,rs|U=u)\ \frac{d\nu(r)}{\bar{p}\ (z)}$$

avec : * ν = mesure de probabilité invariante sur $C_{\mathcal{S}}$, champs des parties de $\mathcal{S}$ ($\nu(Cs) = \nu(sC) = \nu(C)$ $\forall s \in \mathcal{S}$, $\forall C \in C_{\mathcal{S}}$).

$$* \quad \bar{p}(z) = \int_S p(z,s)\ d\nu(s)$$

Le théorème 1 montre qu'avec une conditionalité **non correcte on** obtiendrait :

$$E(\sup_h R(h,U,u)) > \sup_h R(h)$$

ce qui, du point de vue du minimax enleverait tout avantage à la randomisation, ce qui est inacceptable dans des conditions non Bayesiennes.

Le théorème 2 est une application d'un théorème permettant, sous certaines hypothèses, de calculer $\bar{p}[r(z)]$ d'une part et l'espérance conditionnelle par rapport à une tribu d'événements S-invariants (démonstration dans (21)).

Le théorème 3 ci-dessus mentionne au cas de conditionalité illusoire :

Théorème 3.- s'il existe une statistique complète et suffisante $T = t(Z)$, alors $R(h,U,u) = R(h)$ est vérifié sur toute statistique ancillaire U et chaque f qui soit fonction de T.

Mentionnons une nouvelle restriction possible définie par :

Définition 1.- Une statistique ancillaire U vérifie la conditionalité correcte la plus forte, si pour toute fonction convexe non-négative α et toute autre statistique ancillaire $\bar{U}$ vérifiant une conditionalité $\bar{U}$ vérifiant une auditionalité correcte on a :

$$E\ \alpha(b(U,u) \geqslant E\ \alpha(\bar{b}(\bar{U},\bar{u}))$$

avec $b,\bar{b}$ correspondant à la définition p.II 22.

Sous les mêmes conditions que celles utilisées dans le théorème 2, cette définition débouche sur un théorème d'existence (voir (21)) et un principe de conditionalité restreint correspondant est :

Un principe restreint de conditionalité.-

. Si toutes les statistiques ancillaires vérifiant la conditionalité la plus forte donnent le même risque et s'il existe au moins une telle statistique, alors l'emploi d'un risque conditionel est obligatoire.

II D.2.- Vraisemblance.-

Sur l'espace d'échantillonnage (Z, P_h) rappelons que l'expression $P_h(z)$ peut être considérée pour un z donné, à une constante multiplicative près, comme la fonction de h appelée fonction de vraisemblance, et notée $\ell_z(h)$ pour marquer le changement d'optique. Un énoncé général du principe est alors :

Principe de vraisemblance.-

La résolution du problème d'inférence doit être basé seulement sur la fonction de vraisemblance, négligeant tout autre élément des expériences réalisées.

Parmi les diverses interprétation possibles, retenons la suivante exprimée en termes de décision : pour un ensemble H donné, une fonction f de décision doit être regardée comme une fonction construite sur l'ensemble L des fonctions de vraisemblance où toutes les fonctions qui ne diffèrent que par un facteur multiplicatif sont considérées comme équivalentes. Alors pour un espace d'échantillonnage (Z, P_h) donné tel que $\ell_z(h) \in L$ quel que soit z : $f(z) = f(\ell_z(.))$

Le théorème qui suit montre les liaisons qui existent entre le principe de vraisemblance et celui de conditionalité, tels qu'ils ont été respectivement énoncés ci-dessus :

Théorème 4.- Si la résolution du problème de décision est invariant à une transformation biunivoque de l'espace d'échantillonnage dans un autre ensemble homomorphe et si l'on considère l'ensemble E des espaces d'échantillonnages (Z, P_h) tels que leurs fonctions de vraisemblance possibles appartiennent à un espace L donné, alors le principe de conditionalité est équivalent au principe de vraisemblance.

Démonstration : Supposons que le principe de vraisemblance ne soit pas vérifié pour des fonctions de décision f_1 et f_2 ; ainsi il existe deux espaces d'échantillonnage (Z_1, P_{1h}) et (Z_2, P_{2h}) tels que pour un z_1 et un z_2 l'on ait : $f_1(z_1) \neq f_2(z_2)$ bien que $P_{1h}(z_1) = c\, P_{2h}(z_2)$ avec $c = c(z_1 z_2)$ mais indépendant de h. Supposons que Z_1 et Z_2 sont finis et disjoints et considérons un nombre λ tel que : $0 < \lambda < \frac{1}{1+c}$.

Considérons l'espace d'échantillonnage $(Z_1 \cup Z_2, P_h) = (Z, P_h)$

$$\begin{aligned} P_h(z) &= \lambda\, P_{1h}(z) \quad \text{si} \quad z \in Z_1 \\ &= \lambda\, c\, P_{2h}(z) \quad \text{si} \quad z \in Z_2 - \{z_2\}. \\ &= 1 - \lambda(c+1 - P_{1h}(z_1)) \quad \text{si} \quad z = z_2 \end{aligned}$$

Alors l'espace d'échantillonnage $(Z, P_h(z=z \mid z \in Z_1))$ coïncice avec (Z_1, P_{1h}) et $(Z, P_h(Z=z \mid z \in (Z-(z_2)) \cup \{z_1\}))$ coïncide avec (Z_2', P_{2h}) équivalent à (Z_2, P_{2h}) à la transformation bijective près $a(z) = z$ si $z \in Z_2 - \{z_2\}$ et $a(z_2) = z_1$. Ainsi $f_1(z_1) = f(z_1) = f_2'(z_1) = f_2(z_2)$ avec le principe de conditionalité. Ceci est en opposition avec l'hypothèse $f_1(z_1) \neq f_2(z_2)$ et le principe de conditionalité ne peut être vérifié.

Supposons que le principe de vraissemblance soit vérifié pour deux espaces d'échantillonnage (Z_1, P_{1h}) et (Z_2, P_{2h}) tels que si $Z_2 \subset Z_1$ l'on ait $\ell_1 z(h)$ et $\ell_2 z(h)$ différents à une constante multiplicative près pour $z \in Z_2$. Alors ℓ_1 et ℓ_2 sont identiques sur Z_2 et le principe de conditionalité est vérifié.

II D.3.- Remarques.-

Les deux principes abordés ci-dessus ne sont que des exemples de principes adoptés pour guider la construction de la structure de prise de décisions statistiques. Ces principes ne concernent que la définition de l'espace de référence à partir duquel sont prises les décisions et évidemment beaucoup d'autres éléments de la structure du processus de décision soulèvent des problèmes ou sont l'objet de divergence entre les auteurs (voir par exemple dans (9) divers aspects du problème général de décision). En matière de critère de choix les influences réciproques entre les fondements de la structure de choix et la nature du critère sont également très variées. En II D2 et 3, il a été ainsi montré qu'en conditions non bayésiennes, l'espérance du risque conditionel maximum est en général plus grand que le risque maximum et qu'il faut passer à la conditionalité correcte pour avoir l'égalité. Citons ci-dessous un exemple inverse où c'est le critère de choix qui est recherchée pour satisfaire certains principes. Ce cas est tiré de A. PLANT (31).

Le modèle envisagé par PLANT repose sur la structure de base classique : d'une part, un variable aléatoire Z sur les réalisations z de laquelle est batie la fonction de décision f et d'autre part une fonction de perte $L(h,f)$ dépendant du paramètre h. La construction du critère proposé provient de la distinction entre d'une part une perte réelle $L(h,f(x))$ telle qu'elle sera subie après le choix et des pertes idéales $L(h,f(\xi_i))$ où $(\xi_i | i = 1,2)$ est une suite de réalisation passées de z (l'auteur se place dans une situation de décision individuelle caractérisée par la prise d'une décision après l'observation d'un unique résultat z d'expérience, ce résultat étant un élément d'une suite de résultats existant par suite de l'interprétation fréquentielle sur lequel repose l'échantillonnage adopté). Après cette distinction il considère un risque mixte et la définition d'une densité de risque vise à sortir, face au principe de conditionalité, du dilemme suivant : une trop faible conditionalité entraîne trop de mélange entre la perte idéale et la perte réelle et une conditionalité trop forte fait rejeter de l'information. La densité de risque est définie à partir de la contribution d'un voisinage de z de mesure $\Delta\lambda(z)$ à l'espérance de la perte, cette contribution étant estimée à $L(h,f(z))\ G_h(z)\ \Delta\ \lambda(z)$ (On suppose que z appartient à un sous-ensemble A de Z λ-mesurable), l'expression de cette densité de risque est alors :

$$\rho(h,f,z) = G_h(z)\ L(h,f(z))$$

Le risque $\rho(h,f)$ est alors donné par :

$$\rho(h,f) = \int_A \rho(h,f,z)\ d\lambda(z)$$

risque sur lequel un critère minimax peut être appliqué : la fonction de décision sera choisie pour minimiser la densité de risque maximale.

II E - CRITERES MIXTES.-

Les cas d'optimalité envisagés ci-dessus ont été séparés en deux types distincts suivant l'existence ou l'absence d'information à priori disponible (des liaisons entre ces ceux types étant parfois assurées : comme, par exemple, par l'identification de l'un d'eux à un cas particulier de l'autre ou par les relations de convergences ....). De leur côté certains modèles se basent sur des structures intermédiaires à information à priori partielle.

Ces structures de choix intermédiaires comportent évidemment des critères spécifiques. Donnons quelques exemples de telles situations.

II E.1.- Risque "mean-max".-

Ce premier cas est tiré de H. KUDO (22). La fonction de risque initiale $r(h,f)$, $h \in H$ et $f \in F$, est supposée non négative et A_H-mesurable et bornée dans H pour chaque f.

Si μ est une probabilité définie sur A_H et si $\mathcal{B}$ est une sous-tribu de A_H, alors le couple $(\mathcal{B},\mu)$ est appelé information à priori partielle. (traduisant le fait que μ n'est connu que sur $\mathcal{B}$ avant que l'expérimentation n'ai eu lieu). Le risque "mean-max" est alors défini par la fonction $r(\mathcal{B},\mu,f)$ telle que :

$$r(\mathcal{B},\mu,f) = \inf_{C\varepsilon\mathcal{B}} \sum_{i=1}^{h} \sup_{h\varepsilon C} r(h,f) \, . \, \mu(C_i)$$

où C est une sous-tribu définie par une partition de H finie et $\mathcal{B}$-mesurable $\{C_1,..,C_k ; U_{j=1}^{k} C_j = H, C_j \varepsilon \mathcal{B}, C_i \cap C_j = \emptyset \quad i\neq j\}$. Les propriétés d'une telle fonction de risque sont :

* $r(\mathcal{B}',\mu,f) \geqslant r(\mathcal{B},\mu,f) \quad \forall\mu,f$ si $\mathcal{B}'$ est une sous-tribu de $\mathcal{B}$.

* la fonction $r(h,f)$ est $\mathcal{B}$-mesurable à un ensemble de μ-mesure nulle près.

Le concept d'optimalité correspondant est : soit $R(\mathcal{B},\mu) = \inf_{f\varepsilon F} r(\mathcal{B},\mu,f)$, alors une procédure $f^* \varepsilon F$ est dite optimale pour une information à priori $(\mathcal{B},\mu)$ (ou $(\mathcal{B},\mu)$-optimale) si f^* satisfait : $r(\mathcal{B},\mu,f^*) = R(\mathcal{B},\mu)$.

Ce critère d'optimalité pour les informations à priori $(\emptyset,\mu)$ et (A_H,μ), avec $\emptyset$ l'ensemble vide, correspond respectivement au minimax et au critère de Bayes p.r. μ.

Si de plus, μ et ν de A_H coïncident sur $\mathcal{B}$, alors $r(\mathcal{B},\mu,f) = r(\mathcal{B},\nu,f)$ et la $(\mathcal{B},\mu)$-optimalité ne dépend que la distribution marginale de μ sur $\mathcal{B}$ (on retrouve que la procédure minimax ne dépend d'aucune distribution à priori si $\mathcal{B} = A_\emptyset$).

Dans (22), les propriétés de telles fonctions optimales ne sont pas étudiées.

II E.2.- *Critère "Bayes-minimax"*.-

Un second exemple de critère intermédiaire est envisagé ci-dessous, mais cette fois plus à fond puisqu'avec les propriétés du risque, celles des solutions optimales sont présentées. Cet exemple est tiré de K.A. DOKSUM (11) dont l'article fait le point sur un certain nombre de modèle de décision non paramétrique de structure particulière.

II E.2.1.- *Introduction*.-

Le problème de décision envisagé est d'obtenir une fonction de décision optimale lorsque la forme de la distribution de probabilité sur l'ensemble des résultats d'expérience n'est pas complètement connue. Plus précisément la fonction de répartition E des X_i (variables aléatoires à valeur dans R^ν) du n-échantillon $Z = (X_1,\ldots,X_n)$ est supposée appartenir à une classe $\mathcal{E}$ qui n'est pas descriptible de manière naturelle par un paramètre h appartenant à un ensemble $\mathcal{H}$ (le $\mathcal{H}$ envisagé par la suite sera supposé être un espace euclidien de dimension p). Soit $R(E,f)$ la fonction de risque associée à la fonction de décision f lorsque E est la vraie distribution.

Face à ce type de problème, différentes solutions ont été envisagées :
- les solutions minimax minimisant le risque maximum $\sup\limits_{E\in\mathcal{E}} R(E,f)$. De telles solutions ont pu être précisées dans certains cas (par exemple, si $\mathcal{E}$ est la classe de toutes les distributions dans un voisinage d'une distribution normale, alors les solutions minimax dépendent de statistiques dont les espérances mathématiques sont liées)
- les solutions minimax regret qui minimisent l'erreur maximale

$$\sup_{E\in\mathcal{E}}\{R(E,f) - \inf_{f\in\mathcal{F}} R(E,f)\}.$$

- les solutions de Bayes définies à partir d'une distribution de probabilité P sur $\mathcal{E}$ et qui minimisent alors le risque de Bayes $\int_{\mathcal{E}} R(E,f)P(dE)$. De telles solutions ont été explicitées pour certaines forme de P.

Ci-dessous la solution envisagée est optimale pour un critère de minimisation intermédiaire entre celui du risque maximum et celui du risque espéré. Le critère a été envisagé pour une situation où la probabilité P sur $\mathcal{E}$ n'est pas entièrement spécifiée et où l'on ne connaît que la distribution de $E(s_1)$. $E(s_\ell)$ pour certains points s_i, $i = 1,\ldots,\ell$ de R^ν avec $s_1 < \ldots < s_\ell$. Plus précisément il consiste à minimiser l'espérance mathématique du risque maximum, espérance relative à la distribution λ sur l'ensemble $\{E(s_i) : i = 1, \ldots, \ell\}$.

II E.2.2.- *Notations et définitions*.-

Soit $Z = (X_1 \ldots X_n)$ le n-échantillon composé des n premiers éléments de la suite $\{X_i\}$ de variables aléatoires indépendantes et identiquement distribuées par la distribution de probabilité E. Soit $\mathcal{E}$ l'ensemble connu des E possibles, en supposant que $\mathcal{E}$ est un sous-ensemble mesurable d'une certaine classe Γ de distribution de probabilité ; nous supposons de plus qu'il existe une probabilité P sur la tribu A_Γ de parties de Γ telle que $P(\mathcal{E}) = 1$.

Soit $L(E,f(z))$ la fonction de perte à valeurs réelles, $f(z) = d$ étant une décision de $\mathcal{D}$.

Soit $R(E,f)$ la fonction de risque, f étant une fonction de décision de $\mathcal{F}$, telle que : $R(E,f) = E_E\, L(E,f(z))$. Ainsi une fonction de décision f^* sera minimax si $R(f^*) = \inf_{f\varepsilon\mathcal{F}} R(f)$ avec $R(f) = \sup_{E\varepsilon\mathcal{E}} R(E,f)$ et f^* sera de Bayes si $r(P,f^*)\ \inf_{f\varepsilon\mathcal{F}} r(P,f)$ avec $r(P,f) = \int_{\mathcal{E}} R(E,f)P(dE)$.

Une fonction de décision mixte Bayes-minimax minimisera une fonction intermédiaire entre $R(f)$ et $r(P,f)$. Cette fonction est le risque espéré lorsqu'une distribution discrète correspondant à P est connue, l'espérance étant calculé en respectant cette distribution. Précisons la définition de cette fonction.

Soit $C(E)$, $E \varepsilon \mathcal{E}$, le plus petit sous-ensemble compact de A_H tel que $E(C) = 1$. Soit $S(\mathcal{E}) = \cup_{E\varepsilon\mathcal{E}} C(E)$. Soit s_i, $i=1,\ldots,\ell$, ℓ points de $S(\mathcal{E})$ tels que $s_1 < \ldots < s_\ell$. $\{E(s_i)\}$ est un ensemble de ℓ nombres q_i dans $[0,1]$, dont la distribution, connaissant P, est notée $\lambda(.;P,\ell)$. Si l'on pose $q = (q_1,\ldots,q_\ell)$ avec $0 \leq q_1 \leq \ldots\ldots q_\ell \leq 1$, alors :

$$\lambda(q;P,\ell) = P(E;E(s_1) \leq q_1, \ldots, E(s_\ell) \leq q_\ell).$$

Notons $E(q,\ell)$ la classe de distribution de E dont la probabilité correspondante de s_i est q_i, $i=1,\ldots,\ell$; soit $E(q,\ell) = \{E \in E \; ; \; E(s_i) = q_i, \; i=1,\ldots,\ell\}$.
Alors l'espérance du risque maximum d'une fonction de décision f est définie par :

$$r_\ell(P,f) = \int_{R^k} \Big[\sup_{E\in E(q,\ell)} R(E,f)\Big] \, \lambda(dq \; ; \; P,\ell)$$

et une fonction de décision f_o sera mixte Bayes-minimax si elle minimise $r_\ell(P,f)$, soit

$$r_\ell(P,f_o) = \inf_{f\in F} r_\ell(P,f).$$

Supposons que $Z \in R^\nu$. Définissons la probabilité $\tilde{P}$ sur l'espace $(R^\nu \times \Gamma, \; B^\nu \times A_\Gamma)$ où B^ν est la tribu borélienne sur R^ν, par :

$$\tilde{P}(B \times S) = \int_S G_E(B) \; P(dE)$$

où G_E est la probabilité sur R^ν correspondante à la distribution E sur A_H.
On suppose que G_E est A_Γ-mesurable. De plus on suppose qu'il existe une distribution conditionnelle de Z étant donné E telle que :

$$\tilde{Q}(X_1 \leq x_1,\ldots,X_n \leq x_n | E) = \Pi_1^n \, E(x_i)$$

et qu'il existe une distribution conditionnelle sur E étant donné $E(s_i) = q_i$, $i=1,\ldots,\ell$, notée P^q.

II E.2.3.- Théorème d'existence.-

Lemme : Les risques définis ci-dessus vérifient les relations :

* $R(f) \geq r_{\ell}(P,f) \geq r(P,f)$ avec $\ell \geq 1$
* si $\{\Pi_m : s_{m,1} < .. < s_{m,\ell_m}\}$ est, pour $m=1,2,\ldots,$ une suite de partitions telle que chaque partition soit plus fine que la précédente, alors :

$$r_{\ell_m}(P,f) \geq r_{\ell_{m'}}(P,f) \quad \text{pour} \quad m < m' \qquad (1)$$

Ce terme découle directement des définitions adoptées.

Soit une fonction de décision f_{ℓ} qui soit, dans certains cas, mixte Bayes-minimax. Soit $s_1 = \inf_{s \in S(E)} (s)$ et $s_{\ell} = \sup_{s \in S(E)} (s)$. Alors $E(s_{\ell}) = 1$. Nous supposons $E(s_1) = 0$, $-\infty < s_1 < \ldots. < s_{\ell} < +\infty$, et $\ell \geq 3$. Soit $q = \{q_1,..,q_{\ell}\}$ avec $0 = q_1 \leq \ldots. \leq q_{\ell} = 1$. Soit $E_{q,\ell}$ la distribution de probabilité dont les valeurs sont q_i pour s_i, $i=1,\ldots,\ell$ et dont les autres valeurs sont obtenues par interprétation linéaire sur chacun des intervalles $[s_i,s_{i+1}]$, $i=1, \ell-1$. Soit E_{ℓ} la fonction aléatoire dont les valeurs $E_{q,\ell}$ sont obtenues pour des q distribués par $\lambda = \lambda(.;P,\ell)$. Nous supposons que E_{ℓ} est une fonction mesurable d'un certain espace sur (Γ,A_{Γ}). Soit P_{ℓ} la distribution de E_{ℓ} et soit f_{ℓ} la solution de Bayes pour $E_{q,\ell}$ quand q admet une distribution à priori λ (ainsi f_{ℓ} minimise $r(P\ell,f) = \int_{R^{\nu}} R(E_{q,\ell},f)\, \lambda(dq)$. Alors :

Théorème : Si $E_{q,\ell} \in E$ pour presque tout q appartenant à $C(\lambda)$, si f_{ℓ} (solution de Bayes pour $E_{q,\ell}$ par rapport à λ) existe et si :

$L(E,f_{\ell}) = L(F,f_{\ell}) \qquad \forall E,\ F \in E(q,\ell)$ pour presque tout q de $C(\lambda)$, alors f_{ℓ} est une solution mixte Bayes-minimax.

La démonstration dans sa grande ligne est la suivante :

Soit N_i, le nombre de valeurs de Z dans $]s_i, s_{i+1}]$, $i=1, \ell-1$. Alors $N = (N_1, \ldots, N_{\ell-1})$ est suffisant pour $E_{q,\ell}$ et f_ℓ, fonction de Z, ne dépend que de N. La distribution de f_ℓ, en tant que fonction de Z, est la même pour tout E dans $\mathcal{E}(q,\ell)$. Ceci et la relation (1) impliquent :

$$\mathcal{R}(E,f_\ell) = \mathcal{R}(F,f_\ell) \qquad \forall E,\ F \in \mathcal{E}(q,\ell).$$

Alors $$r_\ell(P,f_\ell) = r(P_\ell,f_\ell)$$

et f_ℓ étant optimale pour $\mathcal{P}_\ell$, alors :

$$r(\mathcal{P}_\ell,f_\ell) \leq r(\mathcal{P}_\ell,f) \qquad \forall f.$$

Finalement, puisque $E_{q,k} \in \mathcal{E}$ pour presque tout q :

$$r(\mathcal{P}_\ell,f) \leq r_\ell(\mathcal{P},f)$$

II E.2.4.- Convergence de la solution mixte vers la solution de Bayes.-

Le théorème d'existence ci-dessus est complété par des propriétés de convergence.

L'étude de la convergence repose sur l'introduction de l'ensemble $D[0,1]$ des fonctions sur $[0,1]$ présentant des discontinuités du premier ordre au moins et définies par y tel que :

(i) pour $0 \leq s < 1$ $\quad \lim\limits_{r \downarrow s} y(r)$ existe et $y(s_+) = y(s) = \lim\limits_{r \downarrow s} y(r)$.

(ii) pour $0 < s \leq 1$ $\quad \lim\limits_{r \uparrow s} y(r)$ existe et $y(s_-) = \lim\limits_{r \uparrow s} y(r)$

$D[0,1]$ est doté d'une topologie dont l'idée est que dans cet ensemble deux fonctions y et y' seront proches l'une de l'autre si l'on peut passer du graphe de $y(s)$ à celui de $y'(s)$ par de petites déformations uniformes des abscisses s et également des ordonnées. Une telle topologie a été définie par Skorohod (voir (5), par exemple, chapitre 3) :

Soit Λ l'ensemble des fonctions de $[0,1]$ sur lui-même continues et strictement croissante. Pour y et y' appartenant à D, $d_o(y,y')$ défini comme l'infimum des α positifs tel qu'il existe dans Λ un λ satisfaisant à : $||\lambda|| \leqslant \alpha$ avec $||\lambda|| = \sup_{r \neq s} \left| \log \frac{\lambda s - \lambda r}{s-r} \right|$ est une métrique (modificied Skorohod metric dans son appellation anglo-saxone) pour laquelle D est complet et séparable.

Avec les notations précédentes, nous pouvons prendre $\Gamma = D[0,1]$ et si $\{E_\ell\}$ est une suite de fonctions aléatoires de distribution P_ℓ, E_ℓ convergera en loi vers E si $\int_\Gamma \gamma(E)\, P_\ell(dE) \to \int_\Gamma \gamma(E)\, P(dE)$ pour chaque fonction γ de Γ à valeurs réelles, continue et bornée (ou convergence faible de P_ℓ vers P). Il s'agit alors de montrer que le risque de Bayes pour la solution de Bayes par rapport à la distribution à priori P_o converge vers le risque de Bayes de la solution de Bayes pour la distribution à priori P.

La première étape consiste à utiliser une fonction de distribution G_ℓ qui soit constante sur $[s_i, s_{i+1}[$ et pour laquelle la fonction distribution associée Q_ℓ des $G_\ell(s_1),\dots,G_\ell(s_\ell)$ soit égale à celle des $E(s_1),\dots,E(s_\ell)$. On montre alors que : si $E \subset D[0,1]$, alors G_ℓ converge en distribution vers E.

Ensuite soit $G_{q,\ell}$ la fonction de distribution qui soit constante sur $[s_i, s_{i+1}[$ et dont la valeur en s_i est q_i avec $0 = q_1 \leqslant \dots \leqslant q_k = 1$. Soit Q_ℓ la foncti on de distribution des $G_{q,\ell}$ lorsque q admet une distribution λ ; G_ℓ est alors une fonction de distribution aléatoire obtenue en donnant à q de $G_{q,\ell}$ la distribution λ.

Soit f_ℓ la fonction de décision qui minimise le risque de Bayes $r(P_\ell, f)$. cette nouvelle étape consiste en le théorème suivant :

<u>*Théorème*</u> : Si $G_{q,\ell} \in \mathcal{E}$ pour presque tout q appartenant à $C(\lambda)$, si une solution de Bayes f_ℓ existe, si $L(E,f_\ell(z)) = L(F,f_\ell(z))$ est vérifié pour tout $E, F \in \mathcal{E}(q,\ell)$ et presque tout q de $C(\lambda)$ et si f a dans E un risque continu et borné $R(E,f)$, alors pour $\mathcal{E} \subset D[0,1]$, f_ℓ converge vers f dans le sens que le risque de Bayes $r(P,f_\ell)$ de f_ℓ converge vers le risque de Bayes $r(P,f)$ de f.

<u>Démonstration</u> : Soit $\hat{f}_\ell$ la solution de Bayes pour la distribution à priori Q_ℓ $[r(Q_\ell,\hat{f}_\ell) = \inf_{f \in F} r(Q_\ell,f)]$. Comme la fonction de répartition $G_{q,\ell}$ est constante entre les s_i, $\hat{f}_\ell$ ne dépendra sur Z que de $S = (S_1,\ldots, S_{\ell-1})$ où S_i est le nombre de valeurs de Z égales à s_{i+1} pour $i = 1,\ldots,\ell-1$ (Par hypothèse $E(s_1) = 0$) S a la même distribution que celle de N introduit précédemment pour $G_{q,\ell}$, celle-ci étant elle-même celle de N pour $F_{q,\ell}$. Alors f_ℓ est aussi solution de Bayes pour Q_ℓ, ce qui se traduit par :

$$(1) \quad r_\ell(P,f_\ell) = r(P_\ell,f_\ell) = r(Q_\ell,f_\ell) = r(Q_\ell,f_\ell) = r(Q_\ell,\hat{f}_\ell) \leq r(Q_\ell,f)$$

Avec le précédent théorème de convergence :

$$\lim_{\ell\to\infty} r(Q_\ell,f) = r(P,f)$$

puisque G_ℓ (de distribution Q_ℓ) converge en loi vers E (de distribution P).

Ces deux précédentes inégalités donnent :

$$\lim_{\ell\to\infty} r(Q_\ell,f) \leq r(P,f).$$

Par ailleurs nous avons avec le lemme de c)

$$r_\ell(P,f_\ell) \geq r(P,f_\ell)$$

et si f est solution de Bayes pour P :

$$r(P,f_{\ell}) \geq r(P,f).$$

L'ensemble d'inégalité conduit à

$$\lim_{\ell\to\infty} r_{\ell}(P,f_{\ell}) = r(P,f)$$

$$\lim_{\ell\to\infty} r(P,f_{\ell}) = r(P,f)$$

Lorsque E est l'ensemble des fonctions de distribution continue sur $[0,1]$ ($E \subset C[0,1]$, ensemble des fonctions continues sur $[0,1]$, ensemble sur lequel les propriétés principales de convergence sont présentées, comme pour $D[0,1]$ dans [17]), la convergence de f_{ℓ} vers f peut être démontrée avec des conditions moins restrictives : si $G_{q,\ell} \varepsilon E$ pour presque tout q de $C(\lambda)$, si f_{ℓ} est une solution mixte Bayes-minimax, si une solution de Bayes f existe telle que f ait un risque $R(E,f)$ continue et borné, alors pour $E \subset C[0,1]$, $\lim_{\ell\to\infty} r(P,f_{\ell}) = r(P,f)$.

La démonstration de même inspiration que la précédente repose sur un lemme de convergence analogue à celui utilisé ci-dessus, à savoir : si $E \subset C[0,1]$, alors E_{ℓ} converge en distribution vers E.

II E.2.5.- *Exemple : un problème de test*.-

Examinons un exemple d'obtention d'une solution mixte qui, à partir du théorème d'existence précédent, peut être obtenue (sous certaines conditions sur la fonction de risque) en calculant la distribution à postériori d'un paramètre $p = (p_1 \ldots p_{\ell-1})$ ayant comme distribution à priori la distribution G de $E(s_q)-E(s_1)$, $Es_3-E(s_2)$, ... $E(s_{\ell})-E(s_{\ell-1})$.

Supposons le test suivant : avec $E = E_o \cup E_1$ ($E_o \cap E_1 = \emptyset$;

$\mathcal{E}_o$, $\mathcal{E}_1 \in \Gamma$) nous voulons tester $H_o = \{E \in \mathcal{E}_o\}$ contre $H_1 = \{E \in \mathcal{E}_1\}$. Appelons $f = f(z)$, la fonction de décision, fonction de test. Prenons par exemple comme fonction de perte, la fonction suivante. Classique en théorie des tests : $L(E,f) = L_i$ avec $L_i > 0$ si H_i est rejeté à tort, $i = 0,1$ et $L(E,f) = 0$ autrement. Le risque de Bayes est alors

$$r(P,f) = L_o \int_{\mathcal{E}_o} E_E(f)\ P(dE) + L_1 \int_{\mathcal{E}_1} (1-E_E(f))\ P(dE)$$

et l'espérance du risque maximum :

$$r_\ell(P,f) = L_o \int_{Q_o} \Big(\sup_{E\in\mathcal{E}(q,\ell)} E_E(f)\Big)\ \lambda(dq;P,\ell)$$

$$+ L_1 \int_{Q_1} \Big(\sup_{E\in\mathcal{E}(q,\ell)} (1-E_E(f))\Big)\ \lambda(dq;P,\ell)$$

où $Q_i = \{(E(s_1),\dots,E(s_k))\ ;\ E \in \mathcal{E}_i\}$ $i=1,2$ (ainsi Q_o et Q_1 sont les ensembles de $\mathcal{R}^k$ correspondants à $\mathcal{E}_1$ et $\mathcal{E}_2$). On suppose que $Q_o \cap Q_1$ a une probabilité nulle. On vérifie facilement que la relation (1) du théorème d'existence est vérifiée et si de plus $E_{q,\ell} \in \mathcal{E}$, le résultat correspondant peut être appliqué si $E_{q,\ell} \notin E$, $L(E_{q,\ell},f)$ n'est pas défini). Il y a un moyen naturel de définir $L(E_{q,\ell},f)$ pour que $E_{q,\ell} \in \mathcal{E}(q,\ell)$ pour $q \notin \mathcal{E}_o \cap \mathcal{E}_1$, $L(E,f)$ a la même valeur pour chaque $E \in \mathcal{E}(q,\ell)$ et soit $L(E_{q,\ell},f)$ cette valeur connue. Nous le supposerons dans la suite.

Soit $e(z|q)$ la densité de $E_{q,q}$. Avec le théorème d'existence nous pouvons conclure que la fonction de test mixte Bayes-minimax f_k rejète H_o si :

$$L_1 \int_{Q_1} \prod_{i=1}^{\ell-1} (q_{i+1}-q_i)^{N_i}\ \lambda(dq) \geq L_o \int_{Q_o} \prod_{i=1}^{\ell-1} (q_{i+1}-q_i)^{N_i}\ \lambda(dq) \qquad i=1,2$$

Si l'on précise $p_i = q_{i+1} - q_i \qquad i=1 \ldots k-1$

$$A_i = \{(E(s_2)-E(s_1)), \ldots\ (E(s_\ell)-E(s_{\ell-1}))\ |\ E \in \mathcal{E}_i\}$$

la relation précédente s'écrit :

$$L_1 \int_{A_i} \prod_{i=1}^{\ell-1} P_i^{N_i}\, \Pi(dp) \geqslant L_o \int_{A_o} \prod_{i=1}^{k-1} P_i^{N_i}\, \Pi(dp)$$

où Π est la distribution de $(E(s_2)-E(s_1),\ E(s_\ell)-E(s_{\ell-1}))$ quand E a une distribution P. Cette dernière expression est la solution de Bayes du test multinomial dans lequel $(N_1 \ldots N_{k-1}|P)$ est une variable multinomial de paramètre p et n, le test étant celui de $H_o = \{p \in A_o\}$ contre $H_1 = \{P \in A_1\}$.

La solution est la fonction de test f_ℓ qui rejète H_o lorsque le rapport des probabilités à postériori de H_1 et de H_o est supérieur à L_o/L_1, soit si :

$$\frac{\int_{A_1} g(p|N)dp}{\int_{A_o} g(p|N)dp} \geqslant \frac{L_o}{L_o}$$

où p à la densité a postériori $g(p|N)$ avec $N = (N_1, \ldots, N_{\ell-1})$.

II E.3.- RISQUES "PLUS HAUTS" ET "PLUS BAS".-

Ce dernier exemple, tiré de R.J. BERAN (3), est plus limité puisqu'il concerne essentiellement une procédure d'estimation. Mais par ailleurs, il a l'avantage de bien faire ressortir l'influence de la nature de l'interprétation sur laquelle repose la traduction de l'incertitude, c'est-à-dire les probabilités, sur le critère de choix à adopter (cet aspect est primordial dans l'étude de l'utilité)

II E.3.1.- *Structure de la théorie*.-

Si (Z, A_Z) est un espace d'échantillonnage, nous supposons que l'on a :

$$z = E(x,h) \qquad h \in H$$
$$x \in X$$

Nous supposons que E et p, probabilité sur A_H sont connus. Le problème est de construire une procédure de décision concernant h à partir de la donnée d'une réalisation z.

Après la réalisation de l'expérience, le triplet $(z,h,x) \in Z \times H \times X$ est connu par : z a été observé, x est un élément arbitraire de l'ensemble $H_z(x) = \{h \in H \mid z = E(x,h)\}$. Soit $\{(z,h_i,x_i) \mid i=1,2...\}$ une suite d'hypothétiques expérimentations indépendantes ; l'évaluation de la procédure statistique sera basée sur le comportement moyen des choix faits à partir de ces expérimentations ; si la notation générale $\text{prop}_n(\Pi_i)$ est adoptée $-\text{prop}_n(\Pi_i)$ = proportion de vraie propositions sur $(\Pi_1 \ldots \Pi_n)$ - la perte moyenne située après les n premières expériences hypothétiques est :

$$\frac{1}{n}\sum_{i=1}^{n} L(h_i,d) = \int_0^{\infty} \text{prop}_n\left[L(h_i,d) > y\right] dy$$

si L est une fonction de perte $H \times D \to R^+$.

Le risque plus bas $r_z(L,d)$ et le risque plus haut $s_z(L,d)$ sont les espérances de pertes minimum et maximum obtenues suivant les valeurs de f. Ces risques et leur interprétation fréquentielle conduisent à adopter le critère de choix suivant :

"une décision $d \in D$ est minimax pour une fonction de perte ℓ et une observation z si $s_z(L,d) \leqslant s_z(L,d)$ $\forall d' \in D$ et une fonction de décision $f \in F$ est minimax pour une fonction de perte ℓ si $s_z(L,f) \leqslant s_z(L,f')$ $\forall f' \in F$ et $\forall z \in Z$."

Le théorème central démontré dans (16) est alors le suivant :

<u>*Théorème*</u> : si H et D sont des espaces métriques compacts et si la fonction de perte $L : H \times D \to R$ est continue, alors

- $s_z(L,d)$ et $r_z(L,d)$ sont uniformément continues sur D
- les extrémum de $r_z(L,d)$ sur D sont atteints

L'adoption d'un tel critère minimax dans une procédure de choix statistique vise à remplacer celui de la théorie classique, ce dernier ayant montrer de graves lacunes (on pourrait citer celle, en théorie des test, qu'un test dont l'échantillon est de petite taille mais dont la puissance est élevée puisse donner un choix qui soit en contradiction avec les données).

FONCTIONS D'UTILITE ET DECISIONS

La notion d'utilité dépasse largement le contexte de la seule théorie de la décision. Elle est employée dès que l'on parle de représentation numérique de "préférences" sur des éléments de description donnée. A ce titre les voies de réflexion sur l'utilité sont très variées et pour notre démarche d'analyse des critères de choix employés pour la prise de décision, nous aborderons l'utilité essentiellement sous les deux aspects suivants :

- la théorie générale de l'utilité avec son axiomatique, ses propriétés et ses théorèmes d'existence.
- différentes formes de fonction d'utilité employées dans les modèles de décision.

Dans ce paragraphe nous rencontrerons des utilités de forme très spécifique par rapport à celles rencontrées au paragraphe correspondant au premier aspect. Cette spécificité est principalement due à la structure particulière des ensembles qui interviennent dans les modèles de décision et à la grande part accordée aux questions de traduction de l'incertitude. Les formes d'utilités présentées seront examinées pour l'axiomatique sur laquelle elles sont baties et pour les propriétés qui en découlent ; les discussions sur l'adéquation des axiomes (adéquation dépendante du but recherché - descriptif ou normatif par exemple - et du contexte - décision prise par un individu ou par un groupe d'individus, décision économique, décision médicale, ...) ne seront pas abordées.

III A - *Eléments de la théorie de l'utilité*.-

En général une fonction d'utilité est la représentation numérique de relations de préférences (la nature précise de telles relations sera rappelée en III A.1) et la structure d'une telle fonction dépend en général de trois facteurs principaux :

. la nature de la relation de préférence que cette fonction d'utilité est censée traduire (par exemple : la relation de préférence est-elle totale ? ou dénombrable ?

- propriétés précisées en III A.1)

. la nature de l'ensemble des éléments sur lesquels portent les préférences (par exemple : cet ensemble a-t-il une structure topologique, vectorielle ou de treillis ?)

. le type de représentation numérique qui définit cette utilité (par exemple : le modèle dans lequel rentre la fonction d'utilité est-il une loi économique ou un processus de choix ?)

Les éléments présentés ci-dessous seront principalement extraits de l'Introduction à la théorie de l'utilité de MORINEAU (28). Dans la mesure où cette présentation de la structure générale de l'utilité, structure à laquelle appartiennent les fonctions d'utilités prises en compte en décision, sera assez schématique, nous renvoyons à (28) pour connaître les détails omis.

III A.1.- Relations de préférences.-

Définition 1.- Une relation binaire notée $\preccurlyeq$, dans un ensemble X ayant plus de deux éléments sera dite relation de préférence dans X (pour $x, y \in X$, $x \preccurlyeq y$ sera lue : x n'est pas préféré à g), si elle est :

- réflexive ($\forall x \in X$, $x \preccurlyeq x$ est vrai)
- non symétrique ($\exists x, y \in X$ tels que $x \preccurlyeq y$ soit vraie et $y \preccurlyeq x$ soit faux).

Le couple ordonné formé d'une relation de préférence $\preccurlyeq$ et de l'ensemble X dans lequel elle est définie, noté $(\preccurlyeq, X)$, sera appelé système de préférence.

Dans la définition ci-dessus, la condition essentielle est la réflexivité, les conditions de deux éléments au moins pour X et de non symétrie évitant seulement les cas triviaux.

Définition 2.- Une relation binaire, notée $\asymp$, dans un ensemble X ayant au moins deux éléments sera dite relation d'indifférence dans X (pour $x, y \in X$, $x \asymp y$ sera lue : x et y sont indifférents) si elle est :

- réflexive ($\forall x \in X$, $x \asymp x$ est vraie)
- symétrique ($\forall x, y \in X$, si $x \asymp y$ est vraie, alors $y \asymp x$ l'est aussi)

Définition 3.- une relation binaire, notée $\prec$, dans un ensemble X ayant au moins deux éléments sera dite relation de préférence stricte dans X. (pour $x, y \in X$, $x \prec y$ sera lue : y est préféré à x) si elle est :

- irréflexive ($\forall x \in X$, $x \prec x$ est faux)
- asymétrique ($\forall x, y \in X$, si $x \prec y$ est vraie, alors $y \prec x$ ne l'est pas).

En général les relations de préférence, d'indifférence ou de préférence stricte ne sont pas totales, deux éléments de X non liés par de telles relations seront dits incomparables.

Notations.- Soit $\mathcal{R}$ l'ensemble des relations binaires dans un ensemble X. Précisons rapidement quelques éléments de $\mathcal{R}$. Deux éléments x et y de X liés par R de $\mathcal{R}$ seront notés xRy tandis que deux éléments de X non liés par R seront notés $x \not{R} y$. La relation complémentaire de R de $\mathcal{R}$ est une relation de $\mathcal{R}$, notée $\tilde{R}$, telle que $\forall x, y \in X$, $x \tilde{R} y \Rightarrow x R y$. La relation duale de R de $\mathcal{R}$ est une relation de $\mathcal{R}$, notée R^{-1}, telle que : $\forall x, y \in X$ $x R^{-1} y \Rightarrow y R x$. La relation d'égalité dans $\mathcal{R}$ sera notée Δ ; c'est une relation de $\mathcal{R}$ telle que : $\forall x, y \in X$, $x \Delta y \Rightarrow x = y$. La relation vide de $\mathcal{R}$ sera notée O. $\forall x, y \in X$, $x\, O\, y \Rightarrow x = y = \emptyset$. De plus nous nous définissons les relations suivantes dans $\mathcal{R}$:

si $R,S,T \in \mathcal{R}$, $T = R \cap S \Rightarrow$ si xTy alors xRy et xSy $\forall x,y \in X$

si $R,S,T \in \mathcal{R}$, $T = R \cup S \Rightarrow$ si xTy alors xRy ou x,S,y $\forall x,y \in X$

si $R,S \in \mathcal{R}$, $R \subseteq S \Rightarrow$ si xRy alors xSy $\forall x,y \in X$

si $R,S,T \in \mathcal{R}$, $T = R.S \Rightarrow \forall x,y \in X$ si xTy alors $\exists z \in X$ tel que xRz et zSy.

On démontre alors facilement que si $(\precsim, X)$ est un système de préférence, les relations $\precsim \cap \precsim^{-1}$ et $\precsim \cap \overset{\sim}{\succsim}$ sont respectivement d'indifférence et de préférence stricte.

Définition 4.- Soit $(\precsim, X)$ un système de préférence. Ce système sera dit prétransitif s'il existe une relation binaire dans X, notée $\precsim^*$, vérifiant

- $\precsim^*$ est transitive
- $\precsim \subseteq \precsim^*$
- $(\precsim \cap \overset{\sim}{\succsim}^{-1}) \subseteq (\precsim^* \cap \overset{\sim}{\succsim}^{*-1})$

$\precsim^*$ sera appelé extension transitive de $\precsim$ dans X.

Cette proposition n'est pas vide de sens grace à la proposition suivante :

Proposition 1.- Soit $(\precsim, X)$ un système de préférence. Soit :

$$\prec \; = \; \precsim \cap \overset{\sim}{\precsim}^{-1} \quad ; \quad \asymp \; = \; \precsim \cap \precsim^{-1}$$

Si $\forall m,n$ entiers positifs on a $\precsim^m \cap (\precsim^{-1})^n = 0$, alors le système $(\precsim, X)$ est prétransitif.

D'une manière plus précise, la relation $\precsim^*$ définie dans X par

$\precsim^* = \prec^* \cup \asymp^*$ avec :

$\prec^* = \bigcup\limits_{s>0} (\precsim^s - \asymp^s)$

et

$\asymp^* = \bigcup\limits_{s>0} \asymp^s$

est une extension transitive de $\precsim$

Démonstration par construction $\prec^* \supseteq \prec$ et $\asymp^* \supseteq \asymp$, donc $\precsim^* \supseteq \precsim$, on vérifie aisément que $\prec^*$ et $\asymp^*$ sont des relations transitives et un raisonnement par récurrence montre que $\overset{*}{<} \cap \overset{*}{\asymp} = 0$ et que $\overset{*}{\leqslant}$ est transitive.

Cette propriété de prétransitivité d'un système de préférence est moins contraignante que la propriété de transitivité.

Définition 5.- Soit $(\leqslant, X)$ un système de préférence prétransitif ; il existe alors une plus petite extension transitive $\overset{*}{\leqslant}$ de $\leqslant$ ($\overset{*}{\leqslant}$ est une extension transitive de $\leqslant$ et pour toute autre extension transitive $\overset{*}{\leqslant}$ de $\leqslant$, $\overset{*}{\leqslant} \subseteq \overset{*}{\leqslant}$) ; notons $\overset{*}{\asymp} = \overset{*}{\leqslant} \cap \overset{*-1}{\leqslant}$ la relation d'équivalence qu'elle définit. Le système quotient $\left(\frac{\overset{*}{\leqslant}}{\overset{*}{\asymp}}, \frac{X}{\overset{*}{\asymp}}\right)$ est alors appelé système associé au système prétransitif et sera noté $(\dot{\leqslant}, \dot{X})$, ce système est ordonné ($\dot{\leqslant}$ étant transitif et antisymétrique dans X).

III A.2.- Notion de fonction d'utilité.-

Le but d'une fonction d'utilité est de fournir une représentation numérique d'un système de préfèrence, l'adjectif numérique exigeant d'être précisé. Le sens de ce dernier s'appuie sur la définition du nombre, extension du nombre réel habituel.

Définition 6.- On appelle nombre un couple, noté $(L_{\leqslant}, N)$, où N est un ensemble dont les éléments sont des ensembles de nombres réels (éventuellement réduits à un élément) et où $L_{\leqslant}$ est une relation d'ordre totale sur N.

(Une relation binaire est une relation d'ordre si elle est reflexive, transitive et antisymétrique. Elle est totale si tous les éléments de X sur lequel elle est définie sont comparables). La relation $L_{<} = L_{\leqslant} \cap \tilde{L}_{\leqslant}^{-1}$ est la relation d'ordre stricte correspondante à $L_{\leqslant}$.

La notion d'utilité est alors donnée par :

Définition 7.- Soit $(\preccurlyeq, X)$ un système de préférence avec $\asymp \, = \, \preccurlyeq \cap \preccurlyeq^{-1}$ et $\prec \, = \, \preccurlyeq \cap \tilde{\preccurlyeq}^{-1}$ respectivement comme relation d'indifférence et de préférence stricte associées. Une fonction u sera une fonction d'utilité sur le système de préférence $(\preccurlyeq, X)$ si c'est une application strictement monotone de $(\preccurlyeq, X)$ dans un certain nombre $(\underset{\leqslant}{L}, N)$, cela signifiant que u est une application de X dans N telle que : $\forall x,y \in X$

$$x \prec y \Rightarrow u(x) \underset{<}{L} u(y)$$
$$\text{et} \quad x \asymp y \Rightarrow u(x) = u(y)$$

Cette notion est précisée par une définition plus contraignante :

Définition 8.- Soit $(\preccurlyeq, X)$ un système de préférence. Une fonction u sera une fonction d'utilité totale sur le système de préférence $(\preccurlyeq, X)$ si c'est un homomorphisme de $(\preccurlyeq, X)$ dans un certain nombre $(\underset{\leqslant}{L}, N)$, cela signifiant que u est une application de X dans N telle que :

$$x,y \in X \qquad x \preccurlyeq y \Leftrightarrow u(x) \underset{\leqslant}{L} u(y)$$

Remarquons que, pour assurer la comparaison des deux définitions précédentes, que la condition $x \preccurlyeq y \Leftrightarrow u(x) \underset{\leqslant}{L} u(y)$ est équivalente à l'ensemble de x $x \prec y \Leftrightarrow u(x) \underset{<}{L} (y)$ et $x \asymp y = u(x) = u(y)$ avec les définitions de $\prec$ et $\asymp$ données dans la définition 7. De plus, le passage entre ces définitions est assuré par la proposition suivante :

Proposition 2.- Soit une fonction d'utilité u sur le système de préférence $(\preccurlyeq, X)$ à valeur dans $(\underset{\leqslant}{L}, N)$. Une condition nécessaire et suffisante pour que u soit une fonction d'utilité totale est que la relation de préférence $\preccurlyeq$ soit totale sur X ($\preccurlyeq$ étant alors un préordre total sur X).

L'étude de l'existence de fonctions d'utilité peut être abordée par des côtés différents suivant que l'on mette en relief les caractéristiques de la relation de préférence, du nombre ou des éléments préférés intervenant pour définir l'utilité. Nous allons donner les principaux résultats correspondants à ces divers optiques.

a) <u>Théorèmes d'existence de l'utilité dont les hypothèses sont fonction de la relation de préférence utilisée</u>.-

Théorème 1.- Il existe une fonction d'utilité sur un système de préférence $(\lesssim, X)$ si et seulement si :

(i) $(\lesssim, X)$ est prétransitif

(ii) en appelant $(\dot{\lesssim}, \dot{X})$ le système associé à $(\lesssim, X)$ (f. Def. III.A.5), il existe une application monotone de $(\dot{\lesssim}, \dot{X})$ dans un certain nombre $(L_{\leq}, N)$, application qui soit bijective de X sur N, c'est-à-dire que si u est une fonction d'utilité sur $(\lesssim, X)$, on peut écrire $u = v \circ \varphi \circ i$ avec :

* i est l'application identique de X sur lui même utilisée comme application strictement monotone de $(\lesssim, X)$ sur $(\lesssim^{*}, X)$ où $\lesssim^{*}$ est l'extension transitive la plus petite de $\lesssim$

** φ est l'application canonique de X sur $\dot{X}$ ($\dot{X} = \frac{X}{\asymp^{*}}$ où $\asymp^{*}$ est la relation d'équivalence définie par le préordre $<^{*}$).

*** v est une application monotone de $(\dot{<}, \dot{X})$ dans un certain nombre $(L_{\leq}, N)$ avec v bijective de $\dot{X}$ sur N. Autrement dit si $\dot{x}$ est sa classe d'équivalence dans X pour $\asymp^{*}$, on a : $u(x) = v(\dot{x})$.

Condition nécessaire : soit u une fonction d'utilité sur $(\lesssim, X)$, application strictement monotone de $(\lesssim, X)$ dans un certain nombre $(L_{\leq}, N)$. En définissant $<^{*}$ par $u(x) \dot{L}_{\leq} u(y) \iff x <^{*} y \quad \forall x,y \in X$ et $\asymp^{*}$ par $u(x) = u(y) \iff x \asymp^{*} y$, on peut définir la relation $\lesssim^{*} = <^{*} \quad \asymp^{*}$ qui est alors une extension transitive (et totale) de $\preccurlyeq$ telle que :

$$x \preccurlyeq^* y \iff u(x) \underset{\leqslant}{L} u(y)$$

L'application canonique de X sur la classe d'équivalence $\dot{X} = \frac{X}{\asymp^*}$ étant un homomorphisme de $(\preccurlyeq, X)$ sur $(\frac{\preccurlyeq^*}{\asymp^*}, \frac{X}{\asymp^*}) = (\dot{\preccurlyeq}, \dot{X})$ on a : $x \preccurlyeq^* y \iff \dot{x} \dot{\preccurlyeq} \dot{y}$. Enfin définissons une application v de $\dot{X}$ dans N par $v(\dot{x}) = u(x)$. Alors v est bijective de $\dot{X}$ sur N et monotone de $(\dot{\prec}, \dot{X})$ dans $(\underset{\leqslant}{L}, N)$.

Condition nécessaire : Soit $(\preccurlyeq, X)$ un système de préférence prétransitif et $\preccurlyeq^*$ l'extension transitive minimale de $\preccurlyeq$ $(\preccurlyeq^* = \prec^* \cup \asymp^*)$. Le système ordonné associé est $(\dot{\preccurlyeq}, \dot{X})$ soit v-une application monotone de $(\dot{\prec}, \dot{X})$ dans un certain nombre $(\underset{\leqslant}{L}, N)$. L'application identique i de X sur X est une application strictement monotone de $(\preccurlyeq, X)$ sur $(\preccurlyeq^*, X)$. L'application canonique ϕ de X sur $\dot{X}$ est un homomorphisme de $(\preccurlyeq^*, X)$ sur $(\dot{\preccurlyeq}, \dot{X})$. De par la définition de v, on montre que $v(\phi(i(x))) \underset{<}{L} v(\phi(i(y)))$ si $x = y$. La fonction $v \circ \phi \circ i$ est alors l'utilité recherchée.

Pour une relation de préférence transitive, l'ensemble se simplifie (corollaire 1) comme dans le cas où cette relation est préordre total (corollaire 2).

Corollaire 1.- soit $(\preccurlyeq, X)$ un système transitif de préférence et soit $\asymp$ la relation d'indifférence associée. Il existe une fonction d'utilité sur $(\preccurlyeq, X)$ si et seulement s'il existe un quasi-isomorphisme v du système ordonné des classes d'indifférences (l'ensemble quotien $\frac{X}{\asymp}$ étant appelé l'ensemble des classes d'indifférence) dans un certain nombre $(\underset{\leqslant}{L}, N)$. Plus précisément si u est une fonction d'utilité elle s'écrit pour tout x de X : $u(x) = v(\frac{x}{\asymp})$ où $\frac{x}{\asymp}$ est la classe d'équivalence de x.

Corollaire 2.- il existe une fonction d'utilité totale sur un système de préférence $(\leqslant, X)$ si et seulement si

(i) $\leqslant$ est une relation de préordre totale sur X

(ii) le système totalement ordonné des classes d'équivalence $(\frac{\leq}{\asymp}, \frac{X}{\asymp})$ est isomorphe à un certain nombre $(\underset{\leqslant}{L}, N)$.

b) Premières propriétés des fonctions d'utilité.-

Enonçons rapidement quelques propositions liées au problème de trouver, grâce à la fonction d'utilité, les éléments maximaux de X pour $\leqslant$, c'est-à-dire les éléments $y \in X$ tels qu'il n'existe pas d'élément $x \in X$ qui soit strictement préféré à y (quand on adopte cette démarche de choisir, en accord avec les préférences, les éléments maximaux de X, on dit souvent que le choix se fait en suivant l'axiome de rationalité)

Proposition 3.- S'il existe une fonction d'utilité u sur le système de préférence $(\leqslant, X)$ et si $y \in X$ est tel que $u(y) \in \mathrm{Max}\ u(X)$ - avec $\mathrm{Max}\ u(X)$ défini comme l'ensemble des éléments maximaux de $u(X) \subseteq N$ pour la relation induite par $\underset{\leqslant}{L}$ dans $u(X)$ - alors $y \in \underset{\leqslant}{\mathrm{Max}}\ X$.

La réciproque est fausse et d'une manière générale la connaissance d'une fonction d'utilité sur $(\leqslant, X)$ ne permet pas de reonstituer la relation $\leqslant$. dans X.

Proposition 4.- s'il existe une fonction d'utilité totale u sur le système de préférence $(\leqslant, X)$, alors $\underset{\leqslant}{\mathrm{Max}}\ X = \{y : u(y) \in \mathrm{Max}\ u(X)\}$. De plus $\mathrm{Max}\ u(X)$ définit alors l'ensemble des éléments les plus grands pour $\leqslant$ dans X, c'est-à-dire l'ensemble des $y \in X$ tels que $x \leqslant y \quad \forall x \in X$.

Proposition 5.- soit $(\leqslant, X)$ un système de préférence et supposons qu'il existe une fonction d'utilité u appliquant $(\leqslant, X)$ dans le nombre $(\underset{\leqslant}{L}, N)$, alors pour toute application strictement croissante ω

de $(\underset{\leqslant}{L},u(X))$ dans $(\underset{\leqslant}{L},N)$, la fonction composée $\omega \circ u$ est encore une fonction d'utilité sur $(\leqslant,X)$. De plus il n'existe pas en général de fonction d'utilité u telle que $\omega \circ u$ décrive l'ensemble de toutes les fonctions d'utilité sur $(\leqslant,X)$ lorsque ω décrit l'ensemble des applications strictement croissantes de $(\underset{\leqslant}{L},u(X))$ dans $(\underset{\leqslant}{L},N)$.

c) Rôle de la structure des nombres sur l'existence de l'utilité.-

La théorie de l'utilité exige, à ce stade, de préciser la nature de l'image numérique que la fonction utilité donne du système de préférence. Nous esquissons très sommairement cette vue, dans la mesure où nous ne serons concernés ultérieurement que par un de ses cas particulier. Pour connaître les définitions précises de termes employés nous renvoyons à (28), notre but n'étant ici que de saisir l'esprit général de ce développement de la théorie de l'utilité.

Définition 9.- Soit $(\leqslant,X)$ un système de préférence. On dira qu'il existe une fonction d'utilité sur $(\leqslant,X)$ s'il existe une application strictement monotone de $(\leqslant,X)$ dans une puissance ordinale $(\underset{\leqslant}{L},{}^{\alpha}L)$ dont la base L est une partie de R et l'exposant α un certain nombre ordinal.

On dira que la fonction d'utilité est ${}^{\omega\alpha}2$- l'exicographique. Si la puissance ordinale est de la forme $(\underset{\leqslant}{L}, {}^{\omega\alpha}2)$; ${}^{\alpha}R$-lexicographique si elle est de la forme $(\underset{\leqslant}{L},{}^{\alpha}R)$; scalaire dans le cas où la puissance ordinale se ramène à $(\underset{\leqslant}{L},R)$ (R = ensemble de réels)

Mentionnons un théorème d'existence (sans ici non plus expliciter tous les termes employés) :

Théorème 2.- Si l'axiome de choix tient, et si l'hypothèse généralisée du contenu est admise, alors pour qu'il existe une fonction d'utilité ${}^{\omega\alpha}2$-lexicographique sur un système de préférence $(\leqslant,X)$, il faut et il suffit que

- $(\leqslant, X)$ soit prétransitif (ou a fortiori préordonné)
- il existe une partie Y de X, de cardinal $\leq X\alpha$ dont les éléments sont 2 à 2 comparables pour $\leqslant$ mais non indifférents, et telle que pour tout $x, z \in X$ avec $x < z$, il existe $y \in Y$ avec $x \leqslant y$ et $y \leqslant z$.

Si $(\leqslant, X)$ est totalement préordonné, la fonction d'utilité sera totale.

Le cas particulier de la fonction d'utilité scalaire est :

<u>*Théorème 3.-*</u> Une condition nécessaire, et si l'axiome de choix tient, suffisante pour qu'il existe une fonction d'utilité scalaire sur le système de préférence $(\leqslant, X)$ s'énonce :

- $(\leqslant, X)$ est prétransitif (ou à fortiori préordonné)
- il existe une partie au plus dénombrable y de X, dont les éléments sont deux à deux comparables et non indifférents et telle que $\forall x, z \in Z$ avec $x \leqslant z$, il existe $y \in Y$ avec $x \leqslant y$ et $y \leqslant z$.

si $(\leqslant, X)$ est totalement préordonné, la fonction d'utilité est totale.

Précisons néanmoins que par axiome de choix il faut entendre : tout ensemble peut être bien ordonné, c'est-à-dire que pour tout ensemble X, il existe toujours une relation $\leqslant$ définie sur X telle que le système $(\leqslant, X)$ soit bien ordonné.

Ce théorème 3 a été présenté ici comme un cas particulier de 2. Mais historiquement l'existence d'une fonction d'utilité scalaire a été souvent assurée directement par des conditions imposées au système de préférence (conditions de "continuité", de "substitution", d'Archiméde). Cela a été le cas de L.J. SAVAGE (37) et VON NEUMANN et O. MORGENSTERN (39).

Le cas intermédiaire des fonctions d'utilité ${}^{\omega}R$-lexicographique (suites dénombrables de nombres réels) et ${}^{n}R$-lexicographiques (suites finies de nombres réels) ne seront pas abordés dans ces lignes.

d) <u>Rôle de la structure des éléments préférés sur l'existence de l'utilité</u>.-

Les problèmes d'existence d'utilité peuvent également être abordés en terme de topologie sur l'ensemble X des alternatives. Nous présenterons ci dessous une application de cela au cas particulier où l'ensemble des alternatives possède une structure d'espace vectorielle :

<u>*Définition*</u> *10*.- le système $(\lesssim, X)$ est appelé système vectoriel de préférences si l'ensemble des alternatives X est une partie d'espace vectoriel réel et si la relation de préférence $\lesssim$ satisfait les conditions suivantes :

(i) pour tout $x,y,z \in X$, si $x \lesssim y$ est vrai alors $(x+z) \succ (y+z)$ est faux

(ii) pour tout $x,y \in X$ et $a \in R^{+}$ (réels strictement positifs) si $x \lesssim y$ est vrai alors $ax \succ ay$ est faux.

Dans ce cas un théorème d'existence est le suivant :

<u>*Théorème*</u> *4*.- Soit $(\lesssim, X)$ un système vectoriel de préférences. Il existe une fonction d'utilité ${}^{n}R$-lexicographique sur $(\lesssim, X)$ qui soit linéaire de X dans une partie de R^{n}, si et si seulement :

- la relation $\lesssim$ est prétransitive
- l'espace vectoriel des classes d'équivalence de X pour une extension transitive de $\lesssim$ est de dimension finie n.

<u>*Corollaire*</u> *3*.- Avec les hypothèses du théorème précédent, il existe une fonction d'utilité totale ${}^{n}R$-lexicographique sur $(\lesssim, X)$ qui soit linéaire de X dans une partie de R^{n} si et seulement si :

- la relation $\preccurlyeq$ est un préordre total.

- l'espace vectoriel des classes d'indifférence est de dimension finie n.

Alors la fonction d'utilité est unique à une transformation linéaire près qui fait passer d'une base lexicographique $(e_1,\ldots,e_n)$ de l'espace des classes d'indifférence, à une base lexicographique $(e'_1,\ldots,e'_n)$; une telle transformation s'écrit d'ailleurs :

$$e'_i = r^i_i\, e_i + r^i_{i+1}\, e_{i+1} + \ldots + r^i_n\, e_n \quad (\forall i=1,2,\ldots,n)$$

$$\text{avec} \quad r^i_i \in R \quad \text{et} \quad r^j_i > 0 \quad (\forall i,j=1,\ldots,n).$$

Corollaire 4.- Il existe une fonction d'utilité scalaire sur le système vectoriel $(\preccurlyeq,X)$ qui soit linéaire de X dans R si et seulement si $\preccurlyeq$ est prétransitive et l'espace vectoriel des classes d'équivalence de X pour une extension transitive de $\preccurlyeq$ est de dimension 1.

Une telle fonction d'utilité scalaire est totale si et seulement si la relation de préférence est un préordre total.

Souvent les conditions d'existence de fonction d'utilité sur des ensembles préordonnées sont exprimés en termes algébriques. Le cheminement esquissé ci-dessus part de l'exploitation de propriétés topologiques des espaces ordonnées pour retrouver des conditions courantes qui en constituent alors des cas particuliers. Comme exemple d'approche directe basée sur la condition courante dite d'Archimède (ou de continuité ou de substitution suivant les auteurs) citons :

Définition 11.- Un espace vectoriel de préférence $(\preccurlyeq,X)$ est dit archimédien si la relation de préférence $\preccurlyeq$ vérifie la condition : pour tout nombre naturel positif k, si $x \prec ky$ est faux, alors $y \prec 0$ est aussi faux.

Théorème 5.- Soit $(\preccurlyeq,X)$ un espace vectoriel de préférence ; s'il est archimédien et de dimension finie, il existe une fonction d'utilité scalaire (respectivement scalaire et totale) qui soit linéaire de X dans R si et seulement si la relation de préférence est prétransitive (respectivement prétransitive et totale).

III A.3.- *Fonction d'utilité de VON NEUMANN et MORGENSTERN*.-

III A.3.1.- *Introduction*.-

Les fonctions d'utilité couramment rencontrées en théorie de la décision ne sont que des cas particuliers de celles abordées au paragraphe III A, cas particuliers essentiellement marqués par

- une structure d'espace vectorielle pour l'ensemble des éléments préférés, ceux-ci étant des aléas.
- une fonction de type scalaire.

Le caractère scalaire de la fonction d'utilité permet l'emploi, comme critère de choix, de la maximisation de cette utilité. Avec les notations précédentes, un problème de décision est donné lorsque l'on a défini un système de préférences $(\leqslant, \mathcal{D})$ et une fonction d'uitlité u sur $(\leqslant, \mathcal{D})$. La résolution de ce problème est alors la recherche de la décision $d \in \mathcal{D}$ qui maximise $u(d)$ sur $\mathcal{D}$, $\mathcal{D}$ pouvant être composé d'éléments incertains.

L'introduction d'une utilité dans un modèle de choix simplifie la schématisation, mais ne fait que reporter la richesse de la formalisation sur un seul élément, l'utilité. Nous allons dans ce chapitre examiner différentes formes d'utilité proposée lorsque $\mathcal{D}$ est incertain.

Si nous remontons rapidement dans l'histoire des théories d'utilité construite dans l'incertitude, nous trouvons essentiellement des théories d'espérances où l'utilité s'exprime sous forme d'espérance mathématique : selon la nature de cette utilité ainsi que celle des probabilités sur lesquelles l'opérateur d'espérance mathématique est construit nous trouvons :

- la théorie de la valeur espérée, qui a été la première de ce type, suppose une interprétation en termes économiques.

Cette théorie postule que le choix d'une décision incertaine se fait au vue de la valeur de l'espérance mathématique de la valeur monétaire certaine de cette décision (cette valeur monétaire étant calculée en supposant connu l'état de la nature, soit $m(h,d)$). La fonction d'utilité u est alors définie par :

$$u(d) = \underset{P}{E}\, m(h,d)$$

P étant une probabilité objective de réalisation pour l'état de la nature h.

- la théorie de l'utilité espérée (ou théorie de BERNOUILLI) où la fonction d'utilité u de la décision incertaine d est égale à l'espérance mathématique d'une utilité de la valeur monétaire certaine :

$$u(d) = \underset{P}{E}\; v(m)$$

A ce stade l'incertitude est déjà prise en compte de manière multiple, les probabilités mise en oeuvre étant objectives.

- la théorie l'utilité subjective espérée qui est liée à la définition d'une probabilité subjective sur $\mathcal{D}$ où l'utilité d'une décision incertaine n'est plus exprimée analytiquement en fonction d'éléments certains. Cette troisième théorie a pour base les travaux de VON NEUMANN et MORGENSTERN (39) va être l'objet de la suite de ce chapitre III.

III A.3.2.- Analyse de cette fonction d'utilité.-

Nous présentons son axiomatique à travers une généralisation qui en a été faite par M. HAUSNER (20) et qui est extraite de (28). Le point de départ est la définition d'une loterie sur un ensemble d'alternatives incertaines : si $d_o, d_1 \in \mathcal{D}$ et si $\lambda \in [0,1]$, on interprète (d_o, d_1, λ) comme le pari qui donne

d_o avec la probabilité λ et d_1 avec la probabilité $1-\lambda$ et tel que l'on ait :

- $(d_o,d_1,\lambda) = (d_1,d_o,1-\lambda)$
- si $\lambda_1,\lambda_2 \in]0,1[$: $(d_o,d_1,d_2,\lambda_1) = ((d_o,d_1,\mu),d_2,\nu)$ avec $\frac{\lambda_1}{\lambda_1+\lambda_2-\lambda_1\lambda_2}$ et $\nu = \lambda_1+\lambda_2-\lambda_1\lambda_2$.
- $(d_o,d_o,\lambda) = d_o \quad \forall\lambda \in [0,1]$
- si pour $\lambda \in]0,1]$, $(d_o,d_2,\lambda) = (d_1,d_2,\lambda)$ alors $d_o = d_1$.

Le théorème d'existence est alors (avec les notations définies au III.A) :

__Théorème__ 1.- a) si $\mathcal{D}$ est un ensemble d'alternatives incertaines sur lequel sont définis des paris, $\mathcal{D}$ est isomorphe à une partie convexe d'un espace vectoriel réel.

b) soit $\preccurlyeq$ une relation de préférence définie dans $\mathcal{D}$. Si les conditions suivantes sont satisfaites

(i) $\preccurlyeq$ est prétransitive

(ii) si $d_o \preccurlyeq d_1$, alors $\forall d_2 \in \mathcal{D}$ et pour $\lambda \in]0,1[$ $(d_1,d_2,\lambda) \in \mathcal{D}$ et $(d_o,d_2,\lambda) \in \mathcal{D}$ sont tels que $(d_1,d_2,\lambda) \not\prec (d_o,d_2,\lambda)$.

(iii) si $(d_o,d_2,\lambda) \preccurlyeq (d_1,d_2,\lambda)$ pour $\lambda \in]0,1[$, alors $d_1 \not\prec d_o$.

(iiii) si l'espace vectoriel des classes d'équivalence de $\mathcal{D}$ pour une extension transitive de $\preccurlyeq$ est de dimension finie n.

alors il existe une fonction d'utilité ^{n}R-lexicographique u pour $(\preccurlyeq,\mathcal{D})$ telle que :

$$u(d_o,d_1,\lambda)) = \lambda\, u\,(d_o) + (1-\lambda)\, u(d_1)$$

c) si de plus la relation $\preccurlyeq$ satisfait la condition suivante :

si $d_2 \prec (d_o,d_1,\lambda)\ \forall\lambda \in]0,1]$ alors $d_1 \not\prec d_2$ alors l'application u est une fonction d'utilité **scalaire**.

Comme nous le remarquons plus haut, ce théorème est considéré comme un cas particulier des théorèmes généraux d'existence. Mais la démarche dans un autre sens est envisageable. Soit D une partie finie de $\mathcal{D}$, ensemble des alternatives ; soit p_D une distribution de probabilité arbitraire sur le sous-ensemble D ; soit la relation de préférence $\preccurlyeq$ définie sur l'ensemble des couples (D,p_D), il est alors possible de montrer qu'il existe, sous **certaines** conditions une fonction v scalaire ou vectorielle telle que :

$$(D,p_D) \preccurlyeq (D',p_{D'}) \Rightarrow \underset{p_D}{E} v(d) \leq \underset{p_{D'}}{E} v(d)$$

où E_p est l'opérateur d'espérance mathématique pour la distribution p. Alors les théorèmes d'existence d'une fonction d'utilité sur un système quelconque de préférence $(\preccurlyeq,\mathcal{D})$ seraient considérés comme un cas dégénéré, les parties finies D de $\mathcal{D}$ étant réduite à un seul élément d et les distributions de probabilité p étant telles que $p(d) = 1$.

Remarquons par ailleurs que dans le modèle de VON NEUMANN et MORGENSTERN, les probabilités que définissent les loteries ne sont pas exprimées et peuvent être ignorées du décideur. N'ayant pas de support d'interprétation, le choix du décideur a été parfois jugé trop artificiel et ne tenant pas compte d'éventuelles probabilités à priori et objectives de surcroit. Les contraintes reconnues à leur axiomatique vont souvent dans ce sens.

III B - THEORIES GENERALES DES PROBABILITES UTILITES SUBJECTIVES.-

III B.1.- Théorie de SAVAGE.-

La théorie de L.J. SAVAGE (37), largement discutée et citée (discussion portant tant sur la validité de la cohérence interne que sur le bien fondé des axiomes postulés) comprend une traduction de l'incertitude sous forme d'une distribution de probabilité particulière subjective, plus sophistiquée que celle adoptée par VON NEUMANN et MORGENSTERN. Nous donnons ci-dessous un aperçu détaillé du modèle de SAVAGE en nous inspirant fortement d'une note de lecture de R. GUESNERIE et J.Y. JAFFRAY (19) (les auteurs ont d'ailleurs repris certaines démonstrations jugées fragiles et erronnées). Les démonstrations de SAVAGE sont basées sur sept axiomes, et ceux-ci seront présentés au fur et à mesure de la définition des éléments qui les composent. Un schéma de l'organisation des démonstrations est donné en annexe I à titre de guide.

III B.1.1.- *Premiers éléments du modèle.*-

- les états de la nature h dont l'ensemble est l'univers H. Les évènements sont des ensembles d'états de la nature. L'univers H est un évènement certain ; l'évènement $\emptyset$ est l'évènement impossible.
- une décision est toute situation que le décideurpeut et veut caractériser. L'ensemble des décisions est $\mathcal{D}$.
- un acte est une application quelconque de H dans $\mathcal{D}$ telle que f(h) de $\mathcal{D}$ soit l'image de h par f. L'ensemble des f est noté F.
- résoudre le problème de décision revient à choisir un acte f.

Remarquons que suivant la structure globale du problème de décision (décision après expérimentation ou non, rôle de l'ensemble des états de la nature - voir par exemple (22) pour avoir un aperçu des divers formes de modélisation du problème de décision avec utilité) les fonctions de décision sont construites sur divers espaces (Z en III.B.1, H ici) mais ce point n'influe pas la discussion sur l'utilité, elle-même seulement construite sur F.

III B.1.1.- *Relation de préférence sur les fonctions de décision.*-

La résolution du choix de f contre g dans F respecte une relation de préférence. Le premier axiome impose que cette préférence soit toujours prononcée (même pour l'indifférence) et que les choix soient transitifs :

Axiome 1 : il existe un préordre complet, noté $\lesssim$, sur l'ensemble F des actes.

A propos de la relation $\preccurlyeq$, nous conserverons les notations introduites dans III A. En particulier f et g coïncident dans $A \subset \mathcal{H}$ si $f(h) = g(h)$, $\forall h \in A$.

Le second axiome impose que si deux actes coïncident dans un ensemble $\tilde{A}$, le complémentaire de A dans $\mathcal{H}$, l'ordre relatif dans $\mathcal{H}$ ne soit pas changé par une modification commune dans A :

<u>Axiome 2</u> : si f, g, f', g' sont tels que

1) dans $\tilde{A}$, f coïncide avec g ; f' avec g' ,

2) dans A, f coïncide avec f' ; g avec g'

3) $f \preccurlyeq g$ alors $f' \preccurlyeq g'$

Ces deux axiomes permettent la définition d'un préordre de préférence sur les actes conditionnellement à un évènement A :

<u>*Définition*</u> 1.- f est dit non préféré à g conditionnellement à A si $f' \preccurlyeq g'$ avec :

$f'(h)=f(h) \quad \forall h \in A$

$f'(h)$ quelconque si $h \in \tilde{A}$

$g'(h) = g(h) \quad \forall h \in A$

$g'(h) = f'(h)$ si $h \in \tilde{A}$

Ce préordre noté $(f \preccurlyeq g)/A$ est démontré complet. Une autre relation est alors ajoutée :

<u>*Définition*</u> 2.- un évènement A est dit nul si $\forall f,g \in \mathcal{F}$ on a

$$(f \preccurlyeq g)/A$$

On a alors immédiatement les théorèmes suivants :

<u>*Théorème*</u> 2.- 1) ϕ est nul

2) A nul $\Leftrightarrow \forall f,g : (f \asymp g)/B$

3) A nul, $A \supset B \Rightarrow B$ nul

4) $\tilde{A}$ nul, $(f \leqslant g)/A \Rightarrow f \leqslant g$

5) $(f \leqslant g)/\mathcal{H} \Leftrightarrow f \leqslant g$

6) $\mathcal{H}$ nul $\Rightarrow f \asymp g \quad \forall f$ et g

Théorème 3.- si $\{A_i, i \in I\}$ est une partition finie de A, alors

$\forall i \quad (f \leqslant g)/A_i \Rightarrow (f \leqslant g)/A$

si, de plus, $\exists\, i_o \in I$ tel que $(f < g)/A_{i_o}$, alors :

$(f < g)/A$.

Corollaire 1.- L'union d'un nombre fini d'évènements nuls est nulle

Définition 3.- un acte est dite constant si son image dans $\mathcal{D}$ est indépendante des états de la nature : f constante $\Leftrightarrow f(h) = f \quad \forall h \in \mathcal{H}$.

Un acte constant sera notée d.

On peut alors définir une relation de préférence sur les décisions $(d \leqslant d')$ si et seulement si $f = d$, $f' = d'$ et $f \leqslant f'$. Cette relation ne sera véritablement une relation de préférence sur les décisions que si elle ne dépend pas de l'ensemble des états de la nature réalisables. Ceci est alors assuré par le troisième axiome :

Axiome 3 : soit $f = d$, $f' = d'$ et A non nul ; alors $(f \leqslant f')/A \Leftrightarrow d \leqslant d'$.

Le théorème 3 implique alors :

Théorème 4.- soit $\{A_i, i \in I\}$ une partition finie de A. Si $\forall i$, $\forall h \in A_i$, $f(h) = d_i$, $g(h) = \delta_i$ où $d_i \leqslant \delta_i$, alors $(f \leqslant g)/B$

si de plus $\exists i_o \in I$, $d_{i_o} < \delta_{i_o}$, alors $(f < g)/B$.

Il faut remarquer que les axiomes 1 et 2 et le théorème 4 entrainent l'axiome 3.

c) Probabilité qualitative (subjective) sur les évènements.-

Un acte f_A sera définie pour un évènement A par

$$f_A(h) = d \quad \text{si} \quad h \in A$$
$$f_A(h) = d' \quad \text{si} \quad h \in \tilde{A} \quad \text{avec} \quad d' \prec d$$

On dira que l'acte f_A offre une <u>prime</u> si A est réalisé. Les préférences entre les actes offrant des primes "semblables" sur les divers évènements permettent de définir un préordre sur ces évènements qui traduira la plus ou moins grande probabilité que leur accorde le décideur. Une modification du montant des primes offertes ne devra pas modifier le préordre. Il faut aussi qu'il existe au moins une possibilité de prime. Ceci est contenu dans les axiomes 4 et 5.

<u>Axiome 4</u> : si $d, d', \delta, \delta', f_A, f_B, g_A, g_B$ sont tels que :

1) $d' \prec d$, $\delta \prec d$
2) $f_A(h) = d$, $g_A(h) = \delta$ pour $h \in A$;
   $f_A(h) = d'$, $g_A(h) = \delta'$ pour $h \in \tilde{A}$.
3) $f_B(h) = d$, $g_B(h) = \delta$ pour $h \in B$;
   $f_B(h) = d'$, $g_B(h) = \delta'$ pour $h \in \tilde{B}$
4) $f_A \preccurlyeq f_B$

alors $g_A \preccurlyeq g_B$

<u>Axiome 5</u> : il existe au moins un couple de décisions d, d' telle que $d \prec d'$.

Cet axiome 5 entraîne que H n'est pas nul. On dit que A n'est pas plus probable que B, noté $A \preccurlyeq\cdot\, B$ si : lorsque $d' \prec d$ et f_A, f_B sont tels que :

$$f_A(h) = d \text{ pour } h \in A \; ; \; f_A(h) = d' \text{ pour } h \in \tilde{A}$$
$$f_B(h) = d \text{ pour } h \in B, \; f_B(h) = d' \text{ pour } h \in \tilde{B}$$

on a $f_A \preccurlyeq f_B$.

La relation $\preccurlyeq\cdot$ entre évènements vérifie les conditions suivantes (B,C,D étant des évènements) :

1°) $\preccurlyeq\cdot$ est un préordre complet

2°) $B \leq\cdot C = B \cup D \leq\cdot C \cup D$ lorsque $B \cap D = C \cap D = \emptyset$

3°) $\emptyset \leq\cdot B$; $\emptyset \leq\cdot S$

On dit, par définition, que la relation est une probabilité qualitative (c'est la probabilité subjective qualitative du décidem). Cette probabilité qualitative possède, de par les conditions 1°), 2°), 3°) les possibilités suivantes :

1) $B \subset C = \emptyset \leq\cdot B \leq\cdot C \leq\cdot S$

2a) si $B \cap D = C \cap D = \emptyset$ alors : $B <\cdot C \Leftrightarrow B \cup D <\cdot C \cup D$

2b) si $\emptyset <\cdot C$ et $B \cap C = \emptyset$, alors $B <\cdot B \cup C$

3) si $B \leq\cdot C$, $\tilde{C} \leq\cdot \tilde{B}$

4a) $B \leq\cdot C$ et $C \cap D = \emptyset \Rightarrow B \cap D \leq\cdot C \cup D$

4b) $B \cup D \leq\cdot C \cap D$ et $B \cap D = \emptyset \Rightarrow B \leq\cdot C$

4c) $B <\cdot C$ et $C \cap D = \emptyset \Rightarrow B \cup D <\cdot C \cup D$

4d) $B \cup D <\cdot C \cup D$ et $B \cap D = \emptyset \Rightarrow B <\cdot C$

5a) $B_1 \leq\cdot C_1$, $B_2 \leq\cdot C_2$, $C_1 \cap C_2 = \emptyset \Rightarrow B_1 \cup B_2 \leq\cdot C_1 \cup C_2$

5b) $B_1 \cup B_2 \leq\cdot C_1 \cup C_2$ et $B_1 \cap B_2 = \emptyset \Rightarrow B_1 \leq\cdot C_1$ ou $B_2 \leq\cdot C_2$

5c) $B_1 <\cdot C_1$, $B_2 \leq\cdot C_2$, $C_1 \cap C_2 = \emptyset \Rightarrow B_1 \cup B_2 \leq\cdot C_1 \cup C_2$

5d) $B_1 \cup B_2 <\cdot C_1 \cup C_2$ et $B_1 \cap B_2 = \emptyset \Rightarrow B_1 <\cdot C_1$ ou $B_2 <\cdot C_2$

d) <u>Probabilité quantitative sur les évènements</u>.-

Contrairement à l'usage courant, SAVAGE définit une probabilité quantitative ou mesure de probabilité comme une fonction d'ensemble $P(B)$ attachant à tout $B \subset H$ un nombre réel tel que :

1) $P(B) \geq 0$ et $P(H) = 1$

2) si $B \cap C = \emptyset$, $P(B \cup C) = P(B) + P(C)$

Toute probabilité quantitative permet de définir une probabilité qualitative sur les évènements.

La mesure de probabilité P est dite strictement compatible avec la probabilité qualitative $\leq\cdot$ si $P(B) < P(C) \Leftrightarrow B \leq\cdot C$.

Elle est dite presque compatible avec celle-ci si $B \precsim. C \Rightarrow P(B) \leq P(C)$. Certaines conditions sur $\precsim.$ assurent l'existence d'une mesure de probabilité presque ou strictement compatible avec elle. Le sixième axiome entrainera la vérification de ces conditions et ainsi donc l'existence d'une probabilité quantitative dite subjective strictement compatible avec la probabilité qualitative subjective construite plus haut.

Une n-partition presque uniforme, notée "p.u" de B est une n-partition de B telle que l'union de r de ses éléments n'est jamais plus probable que l'union de $(r+1)$ d'entre eux.

Théorème 6.- Si l'ensemble des n tels qu'il existe une n-partition p.u de S n'est pas bornée, il existe une unique mesure de probabilité P presque comaptible avec $\precsim.$; de plus, pour tout $\rho \in [0,1]$, tout $B \subset S$, il existe $C \subset B$ tel que $P(C) = \rho\, P(B)$.

Une probabilité qualitative est dite fine, si, quelque soit $B \supset \emptyset$, il existe une partition de S dont tout élément est moins probable que B. Deux évènements B et C sont dits presque équivalents (noté $B =. C$) si $\forall G, H$ non nuls et tels que $B \cap C = C \cap H = \emptyset$ on a :

$$B \cup G \precsim. C \quad \text{et} \quad C \cup H \precsim. B$$

Si deux évènements presque équivalents sont toujours équivalents, la probabilité qualitative est dite serrée. Alors :

Théorème 7.- si $\precsim.$ est fine, l'ensemble des n tels qu'il existe une n-partition p.u de H n'est pas borné ; les conclusions du théorème 6 sont alors valables ; de plus $B \succ. 0 \Rightarrow P(B) > 0$

Corollaire 1.- si $\precsim.$ est fine et serrée, la seule probabilité presque compatible avec $\precsim.$ est compatible avec elle, et il existe des partitions de H en des nombres arbitrairement grands d'évènements équivalents.

Théorème 8.- $\leqslant\cdot$ est fine et serrée si et seulement si, $\forall B <\cdot C$, il existe une partition de H telle que l'union avec B d'un quelconque de ses éléments est $<\cdot$ C.

Un axiome minimum pour qu'il existe une mesure de probabilité compatible avec la probabilité qualitative $<\cdot$ serait : "si $B <\cdot C$, il existe une partition de H telle que l'union avec B d'un quelconque de ses éléments soit moins probable que C". Mais SAVAGE pose un axiome plus fort qui lui servira pour la construction de l'utilité.

Axiome 6 : Si $g < f$ et d une décision quelconque, il existe une partition de H telle que si g (ou f) est modifié sur un élément de la partition de manière à y prendre en tout h la valeur d, f demeure préféré à g modifié (ou f modifié demeure préféré à g).

Soit $\leqslant\cdot$ une probabilité qualitative et $\emptyset <\cdot D$. On dira que "B n'est pas qualitativement plus probable que C, sachant D réalisé" - ce qui est noté $B \leqslant\cdot C/D$- si $B \cap D \leqslant\cdot C \cap D$.

Théorème 9.- Si $\leqslant\cdot$ est une probabilité qualitative, $(\leqslant\cdot)/D$ l'est également. Si de plus $\leqslant\cdot$ est finie ou serrée, $(\leqslant\cdot)/D$ l'est aussi.

e) Utilité.-

La construction d'une utilité sur des actes quelconques se fait en passant par celle d'une utilité sur une catégorie particulière de fonctions (les loteries) pour lesquelles les axiomes 1 à 6 sont suffisants, tandis qu'un nouvel axiome, le septième, sera nécessaire dans le cas général.

Théorème 10.- Si 1) $d_1 \ldots d_n$ sont n éléments de $\mathcal{D}$ $(n \geqslant 1)$

2) $\rho_1 \ldots\ldots \rho_n$ sont des nombres tels que $\sum_i \rho_i = 1 \quad i=1,\ldots n$

3) g et f sont des fonctions de décision telles que :

$$P\{g(h)=d_i\} = P\{f(h)=d_i\} = \rho_i \qquad i=1,\dots,n$$

alors $g \asymp f$.

<u>*Corollaire*</u> *2*.- Si $E \subset H$ tel que $P(E) = \rho$.

Si $\rho'_1 \dots \rho'_n$ sont des nombres tels que $\sum_1^n \rho'_i = \rho$

Si g et f sont des fonctions de décision telles que :

$$P\{g(h)=d_i \cap E\} = P\{(f(h)=d_i) \cap E\} = \rho'_i$$

alors $(g \asymp f)/E$.

On peut ainsi caractériser les actes dans un nombre fini de décisions par la liste des décisions et les probabilités discrètes associées. On notera $\sum \rho_i\, d_i$ la classe de toutes les fonctions de décision pour lesquelles il existe des partitions B_i de H telles que $P(B_i) = \rho_i$ et $f(h) = d_i$ pour $h \in B_i$. On posera $\sum \rho_i\, d_i = \underline{f}$, $\underline{f}$ étant appelé une loterie (gamble en anglais). Des combinaisons de loteries sont définies de la façon suivante. Soit $\{\underline{f}_j\}$ une suite finie de loteries $(\underline{f}_j = \sum_i \rho_{ij}\, d_{ij})$ et σ_j des nombres réels positifs tels que $\sum \sigma_j = 1$. Alors :

$$\sum_j \sigma_j\, \underline{f}_j = \sum_{i,j} (\sigma_j\, \rho_{ij}) d_{ij} \qquad \text{avec} \quad \sum_{i,j} \sigma_j\, \rho_{ij} = 1$$

L'extension du préordre précédent des actes aux loteries donne :

$$\underline{f} \preccurlyeq \underline{g} \iff \forall f \in \underline{f},\ g \in \underline{g}\ ,\quad f \preccurlyeq g$$

<u>*Théorème*</u> *11*.- soit $\underline{f}$, $\underline{g}$, $\underline{h}$ et $0 \leq \rho \leq 1$. On a alors

$$\rho\underline{f} + (1-\rho)\underline{h} \preccurlyeq \rho\, \underline{g}(1-\rho)\underline{h} \iff \underline{f} \preccurlyeq \underline{g}$$

<u>*Théorème*</u> *12*.- soit $f \prec g$ et $0 \leq \sigma < \rho \leq 1$. On a alors :

$$\rho\underline{f} + (1-\rho)g \prec \sigma\, \underline{f} + (1-\sigma)\underline{g}$$

<u>*Théorème*</u> *13*.- soit $\underline{f}' \preccurlyeq \underline{f}''$ et $\underline{f}' \preccurlyeq \underline{g} \preccurlyeq \underline{f}''$. Alors il existe un ρ et un seul

$$\rho \underline{f}' + (1-\rho)\underline{f}' \asymp g$$

Une fonction d'utilité est une fonction à valeur réelle des actes telle que si $\underline{f} = \sum \rho_i \, d_i$, $\underline{g} = \sum \sigma_j \, \delta_j$, on ait $\underline{f} \leqslant \underline{g}$ si et seulement si, v étant elle-même une fonction à valeur réelle des décisions :

$$\sum_i \rho_i \, v(d_i) \leqslant \sum_j \sigma_j \, v(\delta_j)$$

avec $u(\underline{f}) = \sum_i \rho_i \; v(d)$ comme définition de cette fonction d'utilité.
Les problèmes d'unicité et d'existence sont résolus de la manière suivante :

Théorème 14.- si u est une fonction d'utilité et si ρ et σ sont des nombres réels avec $\rho > 0$, alors $u' = \rho u + \sigma$ est aussi une fonction d'utilité.

Théorème 15.- si u et u' sont des fonctions d'utilité, il existe des nombres ρ et σ tels que $u' = \rho u + \sigma$ avec $\rho > 0$.

Théorème 16.- il existe une fonction d'utilité.

L'extension de l'utilité à des actes quelconques utilise le septième axiome :

Axiome 7 : Si $(f \leqslant (\geqslant) \; g(h))/B$ $\quad \forall h \in B$
alors $(f \leqslant (\geqslant) \; g)/B$

Des lemmes intermédiaires sont d'abord montrés :

Lemme 1.- Si pour toute décision d, $f \leqslant d$ et $g \leqslant d$ alors $f \asymp g$.

Lemme 2.- s'il existe une décision d_o telle que $f \leqslant d_o$ et si $u(f(h)) < u_o$ $\forall h$, alors il existe une loterie $\underline{g}$ telle que $\underline{f} \leqslant \underline{g}$ et $u(\underline{g}) \leqslant u_o$

Lemme 3.- soit une partition $B_i (i=1,\ldots n)$ et u_i des nombres associés.
Soit f un acte tel que

$$u(f(h)) \leq u_i \quad \text{pour} \quad h \in B_i$$

alors $u(\underline{f}) \leq \sum_i u_i \, P(B_i)$.

A ce stade le théorème important est :

__Théorème__ 17.- si f et g sont bornés, alors :

$$f \lesssim g \iff u(f) \leq u(g)$$

Si la fonction u est bornée, **tous** les actes le sont aussi. Ainsi le théorème précédent s'applique à des actes quelconques. Maintenant, le fait que u soit borné découle directement des axiomes donnés.

__Théorème__ 18.- sous les axiomes 1 à 7, u est borné.

III B.2.- __Remarques__.-

A la suite de la théorie de SAVAGE précédemment abordée précisons deux points avant d'aborder une généralisation.

III B.2.1.- __Structure du problème de décision__.-

Au cours du chapitre II, nous avons abordé des problèmes de décision pour lesquels la résolution consistait à trouver une fonction de décision f (pure ou mixte) optimale. La condition d'optimalité s'exprimait à l'aide d'une fonction de perte L sur $\mathcal{H} \times \mathcal{D}$, et cela par l'intermédiaire des diverses fonctions de risque qui prenaient en compte les traductions de l'incertitude adoptées.

Les schémas mathématiques de résolution que nous montrons dans ce chapitre III correspondent à une logique différente sans pour cela être systématiquement contradictoire (nous intéressant à la structure mathématique de ces modèles nous nous abstiendrons d'intervenir dans le conflit de leurs justifications philosophiques). A partir des mêmes ensembles $\mathcal{H}$, états de la nature et $\mathcal{D}$ ensemble des décisions constituées de toute situation que le décideur veut caractériser, la résolution du choix consiste à trouver une décision d "satisfaisante" compte tenu de l'existence d'une relation de préférence sur $\mathcal{D}$. Cette formalisation, qui par ailleurs ne distingue pas l'existence ou l'absence d'expérimentation, s'appuie sur une notion d'optimalité quelque peu différente de la précédente puisqu'une partie importante du critère de choix est inclue dans cette relation de préférence (cette inclusion étant le reflet d'une analyse informelle réalisée par le décideur préalablement à la construction de la structure formelle du choix telle que nous l'analysons ici).

C'est cette différenciation de structure formelle du critère de choix qui fait l'intérêt de cette comparaison des deux résolutions du problème de choix dans l'incertitude.

III B.2.2.- Probabilités subjectives.-

Dans le chapitre préliminaire (en I.A.1) nous rappelions les diverses écoles d'interprétation des probabilités. Précisons actuellement le cas des probabilités subjectives en rappelant que deux approches ont principalement été utilisés pour les définir.

- d'une part l'approche intuitive utilisée par KOOPMAN, KRAFT, PRATT, SEIDENBERG, SCOTT, GOOO, VILLEGAS, DE FINETTI (voir leur bibliographie dans (16)) et dans laquelle l'axiomatique applique une relation de probabilité comparative "est plus probable que" à un ensemble d'évènements.
- d'autre part l'approche basée sur une relation de préférence-indifférence. C'est cette approche qui est suivie par FISHBURN, RAMSEY, SAVAGE, SUPPES, DAVIDSON, ANSCOMBE, AUMANN, PRATT, RAIFFA, SCHLAIFER .... (voir également dans (16)). L'axiomatique permet la définition conjointe d'une mesure de probabilité et d'une fonction d'utilité subjective espérée, cette définition étant compatible avec la relation de préférence.

Avant d'aborder la seconde approche qui est donc celle utilisée dans la suite de ce chapitre, mentionnons les quelsques grandes lignes de la formalisation de la première approche afin de bien saisir les différences des structures qui existent. Pour cela référons-nous à C. VILLEGAS (38).

Sur une algèbre d'évènements $\mathcal{A}$, une relation binaire $\leqslant$ sera dite une probabilité qualitative si elle satisfait les axiomes suivants :

- <u>axiome de préordre</u> : la relation $\leqslant$ sur $\mathcal{A}$ est un préordre total des évènements avec $\emptyset$ comme premier élément et A comme dernier.
- <u>axiome de monotonie</u> : si $A_1 \cap B_2 = \emptyset$, $B_1, B_2 \in \mathcal{A}$, alors pour $A_1 \leqslant B_1$ et $A_2 \leqslant B_2$, $A_1, A_2 \in \mathcal{A}$, on a $A_1 \cup A_2 \leqslant B_1 \cup B_2$ (et si $A_1 \leqslant B_1$ et $A_2 \leqslant B_2$, alors $A_1 \cup A_2 < B_1 \cup B_2$ ou si $A_1 \leqslant B_1$ et $A_2 < B_1$, alors $A_1 \cup A_2 < B_1 \cup B_2$)

(voir par exemple SAVAGE - III B.1.c)) et le couple $(\leqslant, \mathcal{A})$ est appelé algèbre de probabilité qualitative. Si $\mathcal{A}$ est une σ-algèbre et si $\leqslant$ est monotone continue, (voir définition en (38)) $(\leqslant, \mathcal{A})$ sera appelé σ-algèbre de probabilité qualitative. Si deux évènements A, B de $\mathcal{A}$ sont tels que $A \leqslant B$ on dira que "B est au moins aussi probable que A" quelque soit, pour le point de vue axiomatique avec lequel on présente de telles probabilités le contexte d'interprétation objectif ou subjectif.

Par ailleurs une mesure de probabilité est, rappelons-le, définie comme une fonction P qui assigne à tout évènement A un nombre vérifiant :

(i) $P(\emptyset) = 0$; $P(A) = 1$ et $\forall A \in \mathcal{A} \quad 0 \leqslant P(A) \leqslant 1$

(ii) si $A \cap B = \emptyset$ alors $P(A \cup B) = P(A) + P(B) \quad \forall A, B \in \mathcal{A}$

Si P vérifie (i) et (ii), $(\mathcal{A}, P)$ est une algèbre de probabilité. Si de plus $\mathcal{A}$ est une σ-algèbre et si :

(iii) si $\{A_i : i=1,2....\}$ est une suite d'évènements tels que pour $i \neq j \quad A_i \cap A_j = \emptyset$, alors

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i)$$

alors P est une σ-algèbre de probabilité.

On définit alors sur un ensemble d'évènements $\mathcal{A}$ la compatibilité entre probabilité additive $\leqslant$ et une mesure de probabilité P, si et seulement si

$A \leqslant B \iff P(A) \leqslant P(B) \quad \forall A,B \in \mathcal{A}$.

Donnons à simple titre d'illustration quelques propriétés et conditions d'existence sur ces probabilités qualitatives et mesures de probabilité compatibles (voir (38)) :

(1) " Une condition nécessaire et suffisante pour qu'une mesure de probabilité P qui est compatible avec une probabilité qualitative $\leqslant$, soit σ-additive, est que la probabilité qualitative soit monotone continue."

(2) " Si une σ-algèbre de probabilité qualitative est "sans atome" (voir (38)) alors il existe une et une seule mesure de probabilité compatible, et celle-ci est σ-additive."

(3) " Si une algèbre de probabilité qualitative est fine et "serrée" (voir (38)), alors elle peut être étendue à une σ-algèbre de probabilité qualitative."

(4) " Si une σ-algèbre de probabilité qualitative est sans atome, alors elle est fine et serrée."-

Ces propriétés rappellent celles de σ-additivité fondamentales en théorie de la mesure. Cette théorie de σ-additivité développée par VILLEGAS ne constitue qu'un exemple parmi d'autres théories voisines (voir par exemple celle de R.D. LUCE (25) plus générale et dont T. FINE (17) montre que la liaison avec celle de VILLEGAS).

III B.3.- Théorie de FISHBURN.-

P.C. FISHBURN présente dans (24) une théorie générale de la décision dans l'incertitude comprenant comme critère de choix la maximisation d'une utilité subjective espérée. Cette théorie est une généralisation d'un ensemble de théories récentes issues de la formalisation du problème de choix de SAVAGE. Elle vise à définir, sur une base axiomatique, une mesure de probabilité et une fonction d'utilité subjectives compatibles avec une relation de préférence.

III B.3.1.- Structure des éléments.-

a) L'ensemble H des états de la nature est supposé non vide. Aucune autre hypothèse particulière n'est formulée sur H alors que pour les théories de H. CHERNOFF (8), R.D. LUCE et H. RAIFFA (26), F.J. ANSCOMBE et R.J. AUMANN (2), J.W. PRATT, H. RAIFFA et R. SCHLAIFER (33), P.C. FISHBURN (16) H est supposé fini et que pour celle de SAVAGE H doit être infini.

b) L'ensemble $\mathcal{D}$ des décisions (traduction adoptée pour l'anglais "consequence") est supposé non vide et contenant au moins deux éléments. Les théories de D. DAVIDSON et P. SUPPES (12), F.P. RAMSEY (35) et P. SUPPES (36) supposent que $\mathcal{D}$, est infini.

c) P est une famille de mesure de probabilité simple sur $\mathcal{D}$, notées p_1 et appelées lotteries (appellation prise pour l'anglais "gamble". Les probabilités introduites par les subjectivistes sont essentiellement caractérisées par :

* $p \in P \Rightarrow \exists D \subset \mathcal{D}$, D fini : $p(D) = 1$

* P est un ensemble mixte, à savoir que l'opération définie sur $P^2 \times [0,1]$ dans P par $\alpha p+(1-\alpha)q$ a les propriétés :

(i) $1p + 0q = P$

(ii) $\alpha p+(1-\alpha)q = (1-\alpha)q + \alpha p$

(iii) $\alpha(\beta p+(1-\beta)q) + (1-\alpha)q = (\alpha\beta)p + (1-\alpha\beta)q$

pour $p,q \in P$ et $\alpha,\beta \in [0,1]$

Ces probabilités sont dites externes dans la mesure où leur interprétation n'est pas précisée par la construction du modèle. Cela a déjà été signalé à propos de la construction de l'utilité de NEUMANN et MORGENSTERN et c'est à rapprocher de l'introduction de variables randomisées dans les théories statistiques classiques.

d) Φ est un ensemble de fonctions φ de $\mathcal{H}$ dans $\mathcal{P}$, fonctions appelées paris (en guise de traduction de "horselotterie"). Dans Φ, la combinaison linéaire de φ_1 et φ_2 est définie par :

$$\alpha \in [0,1] \; . \; (\alpha\varphi_1+(1-\alpha)\varphi_2)(h) = \alpha\varphi_1(h) + (1-\alpha)\varphi_2(h).$$

On démontre facilement que Φ est également un ensemble mixte.

e) Précisons encore le sens de divers termes en notations.

- $\psi \in \Phi$ est constant sur un sous-ensemble H de $\mathcal{H}$ si et seulement si :

$$\varphi(h) = \varphi(h') \qquad \forall h,h' \in H$$

et l'on écrit $\varphi = p$ sur H quand $\varphi(h) = p \quad \forall h \in H$

- φ_1 et $\varphi_2 \in \Phi$ correspondent sur $H \subset \mathcal{H}$ si et seulement si :

$$\varphi_1(h) = \varphi_2(h) \qquad \forall h \in H$$

et l'on écrit $\varphi_1 = \varphi_2$ sur H.

- Si $\preccurlyeq$ est une relation binaire sur Φ, on définit les relations binaires $\preccurlyeq$ sur $\mathcal{P}$ à l'aide de la première par :

$$p \preccurlyeq q \iff \varphi_1 \preccurlyeq \varphi_2 \text{ lorsque } \varphi_1 = p \text{ et } \varphi_2 = q \text{ sur } \mathcal{H}$$

- $H \subset \mathcal{H}$ est dit évènement nul si et seulement si :

$$\varphi_1 \asymp \varphi_2 \text{ lorsque } \varphi_1 \text{ et } \varphi_2 \text{ correspondent sur } \tilde{H}$$

($\tilde{H}$ étant le complément de H dans $\mathcal{H}$ et la relation $\asymp$ est déduite de $\preccurlyeq$ conformément aux notations adopétées en III A.1).

- $\varphi \in \Phi$ est dit borné si et seulement s'il existe des nombres réels a et b :

$$P\{h : a \leqslant \underset{\varphi(h)}{E}(u) \leqslant b\} = 1$$

III B.3.2.- Axiomatique.-

Les caractéristiques suivantes de ces éléments sont précisées sous forme d'axiomes. Dans ce type de théorie de probabilité et d'utilité subjectives associées à une relation de préférence, on peut distinguer trois niveaux d'axiomes :

(1) Les axiomes qui permettent d'obtenir une utilité exprimée sous forme d'une espérance mathématique d'une fonction à valeurs réelles sur les états de la nature, la distribution de probabilité associée étant construite pour des évènements sans rapport direct avec les états de la nature.

(2) Des axiomes supplémentaires, le cas échéant, rendent chaque utilité précédente égale à une somme d'utilités dépendantes chacune d'un des états de la nature.

(2) Une dernière série d'axiomes permettent de définir chaque utilité dépendante d'un état de la nature sur un même ensemble. Ce n'est qu'à ce niveau qu'apparaît la nature subjective des probabilités prises en compte ainsi que la forme d'espérance mathématique prise par la fonction d'utilité, elle-même subjective.

Ces axiomes sont les suivants, pour $\psi_1, \psi_2, \psi_3 \in \Phi$

Axiome 1 : $\preccurlyeq$ est un préordre total sur Φ.

Cet axiome se retrouve chez SAVAGE. Il existe également dans le modèle de VON NEUMANN et MORGENSTERN (cet axiome permet en particulier d'assurer la cohérence des préférences avec une structure vectorielle de l'ensemble des éléments préférés), mais la généralisation de ce modèle présenté ci-dessus ne suppose qu'une relation de préférence prétransitive.

Axiome 2 : Si $\psi_1 \prec \psi_2$ et $\alpha \in [0,1]$ alors $\alpha\psi_1+(1-\alpha)\psi_3 \prec \alpha\psi_2+(1-\alpha)\psi_3$

Il est appelé axiome d'indépendance ou axiome de la chose certaine. Il est également courant (VON N et M, SAVAGE, RAIFFA....)

<u>Axiome 3</u> : Si $\varphi_1 < \varphi_2$ et $\varphi_2 < \varphi_3$ alors $\exists\ \alpha,\beta \in [0,1]$:

$\alpha\varphi_1+(1-\alpha)\varphi_3 < \varphi_2$ et $\varphi_2 < \beta\ \varphi_1 + (1-\beta)\varphi_3$

Cet axiome est dit archimédien.

Ces trois premiers axiomes assurent les deux premiers niveaux précédemment signalés.

<u>Axiome 4</u> : $\exists p,p' \in P$ tel que : $p < p'$

La possibilité d'un cas trivial est seulement écarté.

<u>Axiome 5</u> : Si $H \subset \mathcal{H}$ est non nul

Si $\varphi = p$ et $\varphi' = p'$ sur H

si $\varphi = \varphi'$ sur $\tilde{H}$

alors : $\varphi \preccurlyeq \varphi' \iff p \preccurlyeq p'$

<u>Axiome 6</u> : Si $\varphi_1(h) < \varphi_3 \quad \forall h \in \mathcal{H}$ alors $\varphi_1 \preccurlyeq \varphi_3$

Si $\varphi_3 < \varphi_2(h) \quad \forall h \in \mathcal{H}$ alors $\varphi_3 \preccurlyeq \varphi_2$

III B.3.3.- <u>Propriétés d'existence</u>.-

Le théorème central de cette théorie est le suivant :

<u>Théorème</u>.- Les axiomes A1 à A6 impliquent qu'il existe une distribution de probabilité P σ-finie sur la tribu $A_{\mathcal{H}}$ et une fonction réelle u sur $\mathcal{D}$ telle que l'on ait pour tout $\varphi_1,\varphi_2 \in \Phi$:

$$\varphi_1 \preccurlyeq \varphi_2 \iff \underset{P}{E}\Big(\underset{\varphi_1(h)}{E}\ u(d)\Big) \leq \underset{P}{E}\Big(\underset{\varphi_2(h)}{E}\ u(d)\Big)$$

De plus, pour de tels φ_1 et φ_2, P et u on a les propriétés :

(i) chaque pari φ est borné pour P et u

(ii) $\forall H \subseteq \mathcal{H}$ $P(H) = 0 \iff H$ est nul

(iii) u est borné si et seulement s'il existe une partition finie de telle que chaque partie ait une probabilité positive avec P.

(iiii) une mesure de probabilité Q sur $\mathcal{A}_H$ et une fonction réelle u' sur $\mathcal{D}$ tels que $\forall \psi_1, \psi_2 \in \Phi$:

$$\psi_1 \leq \psi_2 \iff E\Big(\underset{Q\ \psi_1(h)}{E} u'(d)\Big) \leq E\Big(\underset{Q\ \psi_2(h)}{E} u'(d)\Big)$$

implique

(j) $P = Q$

(jj) $\exists a,b$ réels avec $a > 0$ tels que :

$$u'(d) = a\, u(d) + b \quad \forall d \in \mathcal{D}$$

Examinons les étapes de la démonstration.

<u>*Lemme*</u> *1.*- Si M est un ensemble mixte et si $\preccurlyeq$ est une relation binaire satisfaisant les axiomes directement analogues aux axiomes 1, 2, 3, alors il existe une fonction w à valeur réelle sur M telle que, $\forall m_1, m_2 \in M$ et $\alpha \in [0,1]$:

(1) $m_1 \preccurlyeq m_2 \iff w(m_1) \leq w(m_2)$

(2) $w(\alpha m_1 + (1-\alpha) m_2) = \alpha w(m_1) + (1-\alpha) w(m_2)$

si w et w' sur M satisfont chacun (1) et (2), alors il existe des nombre a et b, $a > 0$ tels que

(3) $w'(m) = a\, w(m) + b \quad \forall m \in M$.

Ce lemme a été largement utilisé dans d'autres théories et ne sera pas démontrer ici (voir par exemple JENSEN (22)).

<u>*Lemme*</u> *2.*- Si les axiomes 1,2, ..., 5 sont vérifiés et si $\{H_1, \dots, H_n\}$ est une partition finie de H, alors il existe n nombres non négatifs $F_H^*(H_i)$ de somme 1 et une fonction réelle u_H sur $\mathcal{D}$ tels que, $\forall\, p_i, q_i \in P$:

si $\psi_H = p_i$ et $\psi'_H = p'_i$ sur H_i ; $i=1,\dots,n$ alors :

(4) $\psi_H \preccurlyeq \psi'_H \iff \sum\limits_{i=1}^{n} p_H^*(H_i) \cdot E\, u_H(d) \leq \sum\limits_{i=1}^{n} P_H^*(H_i) \cdot \underset{P'_i}{E}\, u_H(d)$

De plus P_H^* et u_H ont les propriétés suivantes :

(i) $P_H^*(H_i) = 0 \Leftrightarrow H_i$ est nul

(ii) Les nombres non-négatifs $P_H^*(H_i)$ de somme 1 sont uniques et u_H n'est unique qu'à une transformation linéaire près.

Démonstration :

a) Supposons que les axiomes 1, 2 et 3 soient vérifiés ; alors le lemme 1 est vrai pour $M = \Phi$. Ainsi il existe une fonction v_H sur Φ vérifiant (1) et (2).

Soit $I = \{i : i=1,\dots,n\}$. Soit $p_o \in P$ fixé. Définissons la fonction $v_{H_i}(p_i)$, $\forall p_i \in P$ et $\forall i \in I$ par :

$$v_{H_i}(p_i) = v_H(\varphi_i) - v_H(\varphi_o)\,\frac{n-1}{n}$$

avec φ_i tel que : $\varphi_i = p_i$ sur H_i ; $\varphi_i = p_o = \tilde{H}_i$

φ_o tel que : $\varphi_o = p_o$ sur H

On en déduit :

$$\text{(A)} \qquad \sum_{i=1}^{n} v_{H_i}(p_i) = v_H(\varphi_H)$$

avec φ_H tel que : $\forall i \in I$ $\varphi_H = p_i$ sur H_i par application de (1) aux deux combinaisons convexes égales $\frac{1}{n}\varphi_H + \frac{n}{n-1}\varphi_o$ d'une part et $\sum_{i=1}^{n} \frac{1}{n}\varphi_i$ d'autre part, et par application de (2) à chaque utilité.

On obtient ensuite :

$$\text{(B)} \qquad v_{H_i}(\alpha p+(1-\alpha)p') = \alpha\, v_{H_i}(p) + (1-\alpha)\, v_{H_i}p' \qquad \forall p,p' \in P \quad \alpha \in [0,1]$$

en posant $\varphi_{p\rho_o}$ et $\varphi_{p'\rho_o}$ tels que :

$$\varphi_p = p \text{ sur } H_i \qquad \varphi'_p = p' \text{ sur } H_i$$
$$\varphi_p = p_o \text{ sur } \tilde{H}_i \qquad \varphi'_p = p_o \text{ sur } \tilde{H}_i$$

et en calculant $v_{H_i}(\alpha p+(1-\alpha)p')$ avec (A) d'une part et avec (2) et (A) d'autre part.

On démontre enfin que, si v_H et v_{H_i} et v'_H et v'_{H_i} satisfont chacun les relations (1), (2) et (A), il existe alors des nombres b_i tels que :

$$\text{(C)} \qquad v'_{H_i}(p) = a\, v_{H_i}(p) + b_i \qquad \forall p \in P \text{ et } i=1, \ldots, n$$

avec $b = b_1 + \ldots + b_n$ et a et b définis par la relation (3).

b) Soit v_H sur Φ vérifiant (1) et (2). Supposons que les axiomes 4 et 5 soient vérifiés et choisissons $\varphi_o \in \Phi$. Soient v_{H_i} sur P vérifiant (A). Soient v'_H et v'_{H_i} les fonctions transformées par (3) et (C) pour que $v'_H(\varphi_o) = 0$ et $v'_{H_i}(p_o) = 0$ respectivement (pour simplifier nous reprendrons les notations v_H et v_{H_i} au lieu de v'_H et v'_{H_i}). Si H_i est nul, posons $P^*_H(H_i) = 0$ de telle sorte que $P^*_H(H_i) \cdot v_{H_i}(p) = 0$, $\forall p \in P$; ceci correspond à $v_{H_i} \equiv 0$ découlant de $v_{H_i}(p_o) = 0$, de la définition d'un évènement nul donné en III B.3.1. et de (A).

Soit $I = \{i : H_i$ n'est pas nul$\}$. L'axiome 4 assure que J n'est pas vide.

Appliquons pour $i \in I$, la définition de $\preccurlyeq$ sur P (voir III B.3.1), l'axiome 5 et les relations (1) et (A) ; on obtient :

$$v_{H_i}(p) \leq v_{H_i}(p') \iff p \preccurlyeq p' \qquad \forall p,p' \in P$$

Chaque v_{H_i} vérifie également une relation similaire à (2) de par (B). Appliquons alors le lemme 1 à P pour les v_{H_i} avec i=k d'une part et un i quelconque d'autre part. Par (3), il existe des nombres $a_i > 0$ et c_i :

$$v_{H_i}(p) = a_i\, \lambda(p) + c_i \qquad \forall p \in P \text{ avec } \lambda = v_{Hk}$$

$(a_k=1 \; ; \; c_k=0)$.

Par construction $v_{H_i}(p_o) = v_{H_k}(p_o) = 0$, ce qui implique que $c_i = 0 \ \forall i \in I$.

Ainsi :

$$\forall i \in I \qquad v_{H_i}(p) = a_i \ \lambda(p)$$

D'après (A)

$$v_H(\psi_H) = \sum_{i\in I} a_i \ \lambda(p_i) \qquad \forall \psi_H \in \Phi$$

Posons : $P_H^*(H_i) = \dfrac{a_i}{\sum_{i\in I} a_i} \qquad \forall i \in I$

et

$$\mu_H(p) = (\sum_{i\in I} a_i) \ \lambda(p) \qquad \forall p \in P.$$

Nous obtenons :

$$\text{(D)} \qquad v_H(\psi_H) = \sum_{i=1}^{n} P_H^*(H_i) \ \mu_H(p_i) \qquad \forall \psi_H \in \Phi$$

avec

$$\text{(E)} \qquad P_H^*(H_i) \geq 0 \quad \text{et} \quad \sum_{i=1}^{n} P_H^*(H_i) = 1$$

et en appliquant (1) sur Φ avec les notations du lemme 2.

$$\text{(F)} \quad \psi_H \preccurlyeq \psi'_H \iff \sum_{i=1}^{n} P_H^*(H_i) \ \mu_H(p_i) \leq \sum_{i=1}^{n} P_H^*(H_i) \ \mu_H(p'_i)$$

Montrons que les $P_H^*(H_i)$ ainsi obtenus sont uniques. Soient μ_H et P_H^* satisfaisant (1), (2), (D) et (E) et supposons que les axiomes 4 et 5 sont vérifiés. I n'étant pas vide, il existe p et $p' \in P$ tels que $\mu_H(p) < \mu_H(p')$ et $p \prec p'$. Soient μ'_H et $P_H^{*'}$ satisfaisant (1), (2), (D) et (E) aussi. De plus le lemme 1 : $\mu'_H = a \ \mu_H + b$ avec $a > 0$; d'après (C) il existe des b_i de somme b tels que :

$$P_H^*(H_i) \ \mu'_H(p) = a \ P_H^*(H_i) \ \mu_H(p) \qquad \forall p \in P \ , \ i = 1, \dots n$$

la résolution de ces équations entraîne que $P_H^* \equiv P_H^*$ et de plus on montre facilement que si $P_H^*(H_i) = 0$, H_i est nul.

Enfin le passage de $\mu_H(p_i)$ à $\underset{p_i}{E}\ u_H(d)$ s'obtient en suivant le cheminement réalisé en a) et b) avec $\mathcal{P}$ à la place de Φ et $\mathcal{D}$ pour $\mathcal{H}$ et $\mathcal{P}$ et en prennent une partition finie de $\mathcal{D}$, ce qui est possible puisque les $p \in \mathcal{P}$ sont σ-finis. Cette analogie conduit à une relation de la sorte :

$$\mu_{HD}(p_D) = \sum_{j=1}^{m} p_D^*(D_j)\ u_{HD}(d_j)$$

et l'identification de p_D^* à p_i lorsque $p_D = p_i$ est réalisée en appliquant les propriétés d'unicité du lemme 1 à l'ensemble mixte $\mathcal{P}_D{*}$ défini par $\mathcal{P}_D{*} = \{p : p \in \mathcal{P}$ et p est constant sur D^* et $\tilde{D}^*\}$ si $D \subseteq \mathcal{D}$ (voir raisonnement similaire suivi dans la démonstration du lemme 3).

<u>*Lemme*</u> 3.- Les axiomes 1 à 5 impliquent qu'il existe une mesure de probabilité P^* sur $\mathcal{H}$ et une fonction réelle sur $\mathcal{D}$ telle que :

$$(5)\quad \varphi \preccurlyeq \varphi' \Longleftrightarrow \underset{P^*}{E}\big(\underset{\varphi(h)}{E}\ u(d)\big) \leqslant \underset{P^*}{E}\big(\underset{\varphi'(h)}{E}\ u(d)\big).$$

pour $\varphi, \varphi' \in \Phi_o$ où $\Phi_o = \{\varphi :\ \varphi \in \Phi$ et φ est constant sur chaque élément d'une partition finie de $\mathcal{H}\}$.

De plus, avec les axiomes 1 à 5 et P^* et u vérifiant (5)

$\forall \varphi, \varphi' \in \Phi_o$:

(i) $\forall A \subseteq \mathcal{H},\ P^*(A) = 0 = A$ est nul

(ii) $P^{*\prime}$ et u' satisfont (5) au lieu de P^* et u, et cela $\forall \varphi, \varphi' \in \Phi_o$ si et seulement si $P^{*\prime} = P^*$ et u' est une transformation linéaire positive de u.

<u>Démonstration</u> :

Appliquons le lemme 2 avec deux partitions différentes de $\mathcal{H}$ en considérant deux loteries $\varphi = p$ et $\varphi = p'$ constantes de $\mathcal{H}$. Il en résulte :

$$\underset{p}{E}\ u_B(d) \leqslant \underset{p'}{E}\ u_B(d) \Longleftrightarrow \underset{p}{E}\ u_o(d) \leqslant \underset{p'}{E}\ u_o(d)$$

Dans un premier temps on montre en raisonnant sur l'ensemble mixte

P que u_A et u_B ne sont définis qu'à une transformation linéaire positive près. On écrit alors (4) sous la forme :

$$\emptyset \lesssim \emptyset' \iff \sum_{i=1}^{n} P_B^*(B_i) \underset{p_i}{E}\ u(d) \leq \sum_{i=1}^{n} P_B^*(B_i) \underset{p'_i}{E}\ u(d) .$$

Soit ensuite un évènement $A \subseteq H$ tel que :

$$\text{(G)} \quad \Phi_A = \{\emptyset : \emptyset \in \Phi \text{ et } \emptyset \text{ est constant sur } A \text{ et } \tilde{A}\}$$

Φ_A est un ensemble mixte. Choisissons les partitions $\{B_1 \ldots B_n\}$, $\{C_1 \ldots C_n\}$ contenant A. Alors, notant que $\emptyset \in \Phi_A$, $\emptyset = p$ sur A et $\emptyset = \tilde{p}$ sur $\tilde{A}$, (G) implique que :

$$\forall \emptyset, \emptyset' \in \Phi : P_B^*(A) \underset{p_A}{E}\ u(d) + (1-P_B^*(A)) \underset{\tilde{p}_A}{E}\ u(d)$$

$$\leq P_B^*(A) \underset{p'_A}{E}\ u(d) + (1-P_B^*(A)) \underset{\tilde{p}'_A}{E}\ u(d)$$

si et seulement si :

$$P_C^*(A) \underset{p_A}{E}\ u(d) + (1-P_C^*(A)) \underset{\tilde{p}_A}{E}\ u(d)$$

$$\leq P_C^*(A) \underset{p'_A}{E}\ u(d) + (1-P_C^*(A)) \underset{\tilde{p}'_A}{E}\ u(d)$$

Avec le lemme 1 il s'ensuit que $P_B^*(A) = P_C^*(A)$ et l'on peut écrire pour une partition non spécifique que H :

$$\text{(H)} \quad \emptyset \lesssim \emptyset' \iff \sum_{i=1}^{n} P^*(B_i) \underset{p_i}{E}\ u(d) \leq \sum_{i=1}^{n} P^*(B_i) \underset{p'_i}{E}\ u(d)$$

Avec le lemme 2, P^* est unique et $P^*(A) = 0 \iff A$ est nul et u est unique à une transformation linéaire positive près.

En utilisant les partitions $\{A, B, \widetilde{(A \cup B)}\}$ et $\{A \cup B, \widetilde{A \cup B}\}$ on montre qu'avec le raisonnement qui a donné (H) que P^* est σ-finie. (avec $A \cap B = \emptyset$)

Enfin on applique (H) à la partition :

$$\{B_i \cap C_j \ : \ i=1 \ldots n, j=1 \ldots\ldots m \ ; \ B_i \cap C_j = \emptyset\}$$

et on obtient :

$$\psi \leqslant \psi' \iff \sum_i \sum_j \ P^*(B_i \cap C_j) \underset{p_i}{E} \ u(d) \leqslant \sum_i \sum_j \ P^*(B_i \cap C_j) \underset{p_i'}{E} \ u(d)$$

et par σ-additivité de P^* on obtient (5).

<u>*Lemme*</u> 4.- Les axiomes 1 à 6 impliquent que I soit vrai sur l'ensemble des loteries bornées.

<u>Démonstration</u> : Les étapes sont succintement les suivantes. H étant un ensemble mixte il existe une fonction réelle v sur H vérifiant le lemme 1. Si l'on pose $\omega(\psi) = \underset{P^*}{E}\left(\underset{\psi(h)}{E} \ u(d)\right)$ pour $\psi \in \Phi_o$ et l'on montre que ω sur Φ_o est, à une transformation linéaire positive près, la restriction de v à Φ_o. Il s'agit donc de montrer que :

$$v(\psi) = \underset{P^*}{E}\left(\underset{\psi(h)}{E} \ u(d)\right)$$

est vraie pour tout ψ borné.

Une première étape consiste à montrer que si $P^*(A) = 1$ et si c et d définis ci-dessous sont finis alors :

$$\text{(I)} \qquad c = \inf_{\psi(h) | h\varepsilon A} \ E \ u(d) \leqslant v(\psi) \leqslant \sup_{\psi(h) | h\varepsilon A} \ E \ u(d) = d$$

(c'est à ce niveau que l'axiome 6 est utilisé)

Après avoir écarté le cas simple où $c = d$, on considère $c < d$ et pour simplifier $c = 0$ et $d = 1$. On considère alors la partition $\{A_1 \ldots A_n\}$ sur H définie par :

$$A_i = \{h : \frac{i-1}{n} < \underset{\psi'(h)}{E} \ u(d) \leqslant \frac{i}{n}\} \qquad i=1, \ldots n$$

pour $\varphi'(h)$ vérifiant I et tel que $v(\varphi') = v(\varphi)$.

On associe aux A_i les $p_i \in \mathcal{P}$ tels que :

$$\frac{i-1}{n} \leqslant \underset{P_i}{E}\ u(d) \leqslant \frac{i}{n} \quad i=1, \ldots, n$$

En introduisant un p_o convenable $(p_o = \frac{1}{n} \sum_{i=1}^{n} p_i)$ on calcule $v(\varphi')$ par encadrement.

Lemme 5.- Si les axiomes 1 à 6 sont vérifiés et s'il existe une partition dénombrable de $\mathcal{H}$ telle que chaque élément de cette partition ait une probabilité positive avec $\mathcal{P}^*$, alors u est borné sur $\mathcal{D}$.

Démonstration : A partir de la partition $\bar{A}$ dénombrable de $\mathcal{H}$ avec $P^*(A) > 0$ $\forall A \in \bar{A}$, soit $A_1 \subset \bar{A}$ l'élément qui maximise $P^*(A)$. On définit ensuite $A_2 \subset \bar{A}$ qui maximise $P^*(A)$ sur $\bar{A} - \{A_1\}$ et ainsi de suite. On suppose alors la partition A_i et on définit p_i par : $\underset{p_i}{E}\ u(d) = \frac{1}{P^*(A_i)}$. On définit ensuite φ et φ'_n pour

$$\varphi = p_i \quad \text{sur} \quad A_i, \quad i=1,\ldots,n$$

φ'_n : constant sur chaque A_i pour $i \leqslant n$ avec

$$\underset{\varphi'_n(h)}{E}\ u(d) = \frac{1}{P^*(A_n)} - \frac{1}{P^*(A_i)} \quad , \quad h \in A_i$$

constant sur $\cup_{i=n+1}^{\infty} A_i$ avec $\underset{\varphi'_n(h)}{E}\ u(d) = 0$, $h \in \cup_{i=n+1}^{\infty} A_i$.

Enfin on suppose que u n'est pas borné et l'on montre alors que le $v(\psi)$ obtenu sur Φ_o au lemme 4 $(v(\varphi) = \underset{P^*}{E}\ (\underset{\varphi(h)}{E}\ u(d))$ doit être infini

$$(v(\varphi) \geqslant \frac{1}{P^*(A_n)} - \frac{\sum_{i=1}^{n} P^*(A_i)}{P^*(A_n)} + n \geqslant n).$$

Lemme 6.- Si les axiomes 1 à 6 sont vrais et s'il existe un entier n tel qu'il n'y ait pas de partition de $\mathcal{H}$ en n parties telles que chaque partie ait une probabilité positive sous P^*, alors toutes les loteries sont bornées.

<u>Démonstration</u> : Remarquons d'abord que si pour chaque n il existe une partition telle que chacune des n parties ait une probabilité positive sous P^* il existe alors une partition dénombrable de H pour laquelle chaque partie a une probabilité positive et le problème est résolu.

Dans le cas contraire envisagé dans ce lemme 6, il existe donc un m, unique, tel qu'une partition de taille m ait au plus m parties de probabilité positive. Supposons que $u(\delta) = 0$ sur $\mathcal{D}$. Soit un $\psi' \in \Phi$ non borné supérieurement. Soit ψ obtenu à partir de ψ' en remplaçant chaque d où $\psi(h)(d) > 0$ et $u(d) < 0$ par δ. ψ est alors un borné supérieurement et pour chaque $n > 0$:

$$P^*\{\underset{\psi(h)}{E} u(d) \geq n\} > 0 .$$

Ainsi il existe un entier N et un $\alpha > 0$ tel que

$$P^*\{\underset{\psi(h)}{E} u(d) \geq n\} = \alpha \qquad \forall n \geq N .$$

Soit $\underset{p_i}{E} U(d) = i$ pour $i=1,2,\ldots$

$$\psi'_n = \begin{cases} \psi & \text{sur } \{h : \underset{\psi(h)}{E} u(d) \geq n\} \\ \psi_n & \text{sur } \{h : \underset{\psi(h)}{E} u(d) \leq n\} \end{cases}$$

$$\psi''_n = \begin{cases} \psi_n & \text{sur } \{h : \underset{\psi(h)}{E} u(d) \geq n\} \\ \psi & \text{sur } \{h : \underset{\psi(h)}{E} u(d) < n\} \end{cases}$$

On montre alors que pour un $v(\psi)$ comme celui trouvé sur H_o au lemme 4 on a :

$$v(\psi) + n = v(\psi'_n) + v(\psi''_n)$$

et on aboutit à $v(\emptyset) \geq n\alpha-1 \quad \forall n \geq N$ ce qui est contradictoire. Egalement ψ' ne peut être borné.

Situons rapidement cette théorie par rapport à d'autre vis-à-vis de la fonction d'utilité. D. DAVIDSON et P. SUPPES (12) supposent, $\mathcal{D}$ étant fini, que l'utilité u vérifie

$$u(d_1) - u(d_2) = u(d_4) - u(d_3) \quad \text{si} \quad d_1 \prec d_2 \quad \text{et} \quad d_3 \prec d_4$$

et si dans $\mathcal{D}$ il n'y a pas de décision entre d_1 et d_2, d_3 et d_4.

F.P. RAMSEY (35) et P. SUPPES (36) supposent que, $\mathcal{D}$ étant également fini, si $u(d_1) < u(d_2)$ alors il existe $d_3 \in \mathcal{D}$ tel que :

$$u(d_3) = 0{,}5 \; u(d_1) + 0{,}5 \; u(d_2).$$

On a vu que SAVAGE, qui demande que H soit infini, suppose que pour tout n entier, il existe une partition de H en n parties telles que P^* vale $1/n$ sur chaque partie. K.J. ARROW (1) suppose aussi cela sur P^*.

L'utilité obtenue, $v(\emptyset) = E_{P^*}^{*}(E_{\emptyset(h)} u(d))$, fait partie, en tant qu'utilité espérée aux utilités dites additives ; sur le plan pratique la littérature est abondante en ce qui concerne leurs techniques de détermination : voir par exemple P. C. FISHBURN (15) ou H. RAIFFA et R. SCHLAIFFER (34).

III B.4.- *Théorie conditionnelle de la décision.*-

Le modèle ci-dessus de FISHBURN était une généralisation de l'idée de SAVAGE. Présentons ci-dessous une variante de cette idée qui s'appuie sur l'introduction d'éléments conditionnels dans la formalisation. Il s'agit d'une théorie qualitative de décisions conditionnelles. De telles théories conditionnelles ont été suggérées par certains auteurs, dont P.C. FISHBURN (13) et présentée par ailleurs sous forme axiomatique dans un cas simple par J. PFRANZAGL (32). C'est une axiomatisation généralisant cette dernière que nous suivons ; elle a été faite par R.D. LUCE et D.H. KRANTZ (27).

III B.4.1.- *Introduction du modèle*.-

Le problème de décision se présente sous une forme conditionnelle où $\mathcal{D}$ est l'espace de décisions possibles et où une décision conditionnellement un évènement A sera une fonction de A dans $\mathcal{D}$, fonction définie que dans A et notée g_A. L'ensemble G des actes, au lieu d'être l'ensemble des applications de H (ensemble des états de la nature) dans $\mathcal{D}$, sera l'ensemble des applications de la tribu A_H dans $\mathcal{D}$, celles-ci étant notées g_A si $A \in A_H$. G possède une relation de préférence $\preccurlyeq$. L'axiomatique proposée sur G vise à permettre la construction d'une fonction d'utilité v sur $\mathcal{D}$ et une distribution de probabilité conditionnelle $P(.|A)$ de telle sorte que le choix respecte la macimisation d'une utilité dans l'incertitude égale à une espérance conditionnelle

$$v(g_A) = \underset{P(.|A)}{E} u(d_A)$$

Deux points essentiels différencient cette démarche de celle de SAVAGE :

- toutes les fonctions de décisions ne sont pas supposées avoir les mêmes domaines de définitions.
- il existe plusieurs espérances mathématiques définissant l'utilité, celles-ci dépendant de plusieurs distributions de probabilité.

III B.4.2.- *Axiomatique*.-

Soit N l'ensemble des évènements nuls N de A_H pour une distribution donnée (pour cette distribution de probabilité sur A_H, la probabilité de N est nulle) les axiomes suivants s'entendent

$\forall A,B \in A_H - N$

$\forall R,S \in A_H$

$\forall g_A, g_A^{(i)}, g_{A \cup B}, g_{A \cup R}, g_B^{(i)}, h_A^{(i)}, k_B^{(i)} \in G$

<u>Axiome</u> A : (i) si $A \cap B = \emptyset$ alors $g_A \cup g_B \in G$

(ii) si $B \subset A$, alors g_A réduite à B appartient à G

<u>Axiome</u> B : $\succcurlyeq$ est un préordre complet sur G

<u>Axiome</u> C : si $A \cap B = \emptyset$ et $g_A \prec g_B$, alors $g_A \cup g_B \prec g_A$

Axiome D : si $A \cap B = \emptyset$ alors $g_A^{(1)} \succsim g_B^{(2)}$ si et seulement si
$g_A^{(1)} \cup g_B \succsim g_A^{(2)} \cup g_B$.

Axiome E : si $A \cap B = \emptyset$, $g_A^{(i)} \asymp g_B^{(i)}$ $(i=1,2,3,4)$

$$g_A^{(1)} \cup k_B^{(1)} \asymp g_A^{(2)} \cup k_B^{(2)}$$

$$h_A^{(1)} \cup g_B^{(1)} \asymp h_A^{(2)} \cup g_B^{(2)}$$

alors

$g_A^{(3)} \cup k_B^{(1)} \succsim g_A^{(4)} \cup k_B^{(2)}$ si et seulement si

$$h_A^{(1)} \cup g_B^{(3)} \succsim h_A^{(2)} \cup g_B^{(4)}$$

Axiome F : si $A \cap B = \emptyset$, $\{n\}$ étant une suite de nombre entiers consécutifs,
$g_B^{(1)} \not\asymp g_B^{(2)}$ et $g_A^{(i)} \cup g_B^{(1)} \asymp g_A^{(i+1)} \cup g_B^{(2)}$.
$\forall i$, $i+1 \in \{n\}$,
alors soit $\{n\}$ est finie
soit $\{g_A^{(i)}$ avec $i \in \{n\}\}$ est non borné.

Axiome G : (i) si $R \in N$ et $S \subset R$, alors $S \in N$
(ii) $R \in N$ si et seulement si, $\forall g_{A \cup R} \in G$, $g_{A \cup R} \asymp g_A$ où
$g_{A \quad R}$ est réduite sur A.

Axiome H : (i) $A_H - N$ contient au moins trois éléments 2 à 2 disjoints.
(ii) $G/\asymp$ contient au moins deux classes d'équivalence distincte.

Axiome I : (i) si A et g_B sont donnés, alors $\exists h_A \in G$ pour lequel
$h_A \asymp g_B$.
(ii) si $A \cap B = \emptyset$ et $h_A^{(1)} \cup g_B \succsim g_{A \cup B} \succsim h_A^{(2)} \cup g_B$ alors

$\exists h_A \in G$ tel que $h_A \cup g_B \asymp g_{A \cup B}$.

A partir de ces axiomes, le théorème central relatif à ce modèle est :

<u>*Théorème 0*</u>.- Si les axiomes précédents A à I sont vérifiés, alors il existe des fonctions à valeur réelle v sur G et une distribution de probabilité P sur $\mathcal{A}_H$ σ-finie telle que :

$\forall A,B \in \mathcal{A}_H - \mathcal{N}$, $\forall N \in \mathcal{A}_H$, $\forall g_A, g_B \in G$:

(i) $N \in \mathcal{N} \iff P(N) = 0$

(ii) $g_A \succcurlyeq g_B \iff v(g_A) \geqslant v(g_B)$

(iii) si $A \cap B = \emptyset$ alors :

$$v(g_A \cap g_B) = v(g_A)\ P(A|A \cup B) + v(g_B)\ P(B|A \cup B)$$

De plus P est unique et v aussi à une transformation linéaire positive près.

Nous renvoyons à (25) pour avoir la démonstration précise de ce théorème. Par ailleurs il existe une variante au théorème précédent basé sur une axiomatique similaire :

<u>Axiomes A', B', C', D', G'</u>, identiques à A,B,C,D,G

<u>Axiomes</u> E' : si $A,B,C,D \in \mathcal{A}_H - \mathcal{N}$, $A \cap B = C \cap D = \emptyset$,

$$g_A^{(i)} \asymp h_C^{(i)} \succ g_B^{(i)} \asymp k_B^{(i)} \qquad i=1,2$$

alors $g_A^{(1)} \cup g_B^{(1)} \succcurlyeq h_C^{(1)} \cup k_D^{(1)}$ si et seulement si

$$g_A^{(2)} \cup g_B^{(2)} \succcurlyeq h_C^{(2)} \cup k_D^{(2)}$$

<u>Axiome</u> F' : $\succcurlyeq$ est totale sur G

Axiome I' : (i) identique à I(i)

(ii) si $A \cap B = \emptyset$ et $g_{A \cup B}$ et g_B sont données alors

$\exists h_A \in \mathcal{G}$ tel que $h_A \cup g_B \asymp g_{A \cup B}$.

Définissons la relation $\gtrsim$ sur $\mathcal{A}_{\mathcal{H}} \times (\mathcal{A}_{\mathcal{H}} - \mathcal{N})$: pour $A, C \in \mathcal{A}_{\mathcal{H}}$ et $B, D \in \mathcal{A} - \mathcal{N}$, on a $A|B \gtrsim C|D$ si :

soit (i) $C \cap D \in \mathcal{N}$

soit (ii) $\tilde{A} \cap B \in \mathcal{N}$

soit (iii) $A \cap B$, $\tilde{A} \cap B$, $C \cap D$, $\tilde{C} \cap D \in \mathcal{A}_{\mathcal{X}} - \mathcal{N}$

$\exists g_{A \cap B}, g_{\tilde{A} \cap B}, h_{C \cap D}, k_{\tilde{C} \cap D} \in \mathcal{G}$ et si :

$$g_{A \cap B} \asymp h_{C \cap D} \succ g_{\tilde{A} \cap B} \asymp k_{\tilde{C} \cap D}$$

et

$$g_{A \cap B} \cup g_{\tilde{A} \cap B} \gtrsim h_{C \cap D} \cup k_{\tilde{C} \cap D}$$

De plus si $A \in \mathcal{A}_{\mathcal{H}}$ et si B est tel que $\tilde{B} \in \mathcal{A}_{\mathcal{H}} - \mathcal{N}$ (B étant alors appelé évènement propre) alors A est indépendant de B si $A|B \asymp A|\tilde{B}$.

Théorème 1.- Si les axiomes A' à I' sont vérifiés et si à chaque évènement propre il existe un évènement propre qui soit indépendant de lui, alors les propositions du théorème sont vraies et de plus : si $A, C \in \mathcal{A}_{\mathcal{H}}$, $B, D \in \mathcal{A}_{\mathcal{H}} - \mathcal{N}$ alors $A|B \gtrsim C|D$ si seulement si

$$\frac{P(A \cap B)}{P(B)} \geq \frac{P(C \cap D)}{P(D)} .$$

L'extension de la notion d'utilité introduite ci-dessus est possible dans certains cas au prix de quelques définitions supplémentaires. Un acte $g_A \in \mathcal{G}$ sera dit constant si pour $d \in \mathcal{D}$ et $\forall h \in A$, $g_A(h) = d$, décision notée d_A. Un acte conditionnel $g_A \in \mathcal{G}$ sera appelé loterie (toujours de l'anglais gamble) si g_A est fini - son image dans $\mathcal{D}$ étant un sous-ensemble fini de $\mathcal{D}$) et si $\forall d \in g_A(A)$, $g_A^{-1}(d) \in A_H - N$.

Un théorème correspondant est alors :

<u>*Théorème*</u> *2*.- supposons que les hypothèses des théorèmes 0 et 1 soient vérifiées.

Supposons de plus que :

(i) $\forall d \in \mathcal{D}$, $\exists A(d) \in A_H - N$ tel que $d_{A(d)} \in G$

(ii) si $d \in \mathcal{D}$, $A, B \in A_H - N$ et $d_{A_1} d_B \in G$

alors $d_A \asymp d_B$.

alors il existe une fonction unique u à valeur réelle sur $\mathcal{D}$ telle que pour chaque loterie $g_A \in G$ avec $A \in A_H - N$,

$$v(g_A) = \underset{P(.|A)}{E} u(g_A)$$

où u et P sont les fonctions et distributions construites dans le théorème III B.20.

Brièvement la comparaison de l'axiomatique ci-dessus avec celle de SAVAGE fait ressortir essentiellement trois points de divergence.

Premièrement dans la théorie conditionnelle, tous les axiomes ne sont pas exprimés en fonction des mêmes éléments, les éléments de la théorie conditionnelle étant plus primitifs. Cela est déjà moins vrai pour le théorème 1*.

Deuxièment et toujours dans la théorie conditionnelle l'ensemble des états de la nature n'est pas nécessairement infini alors que le caractère infini est obligatoire pour l'espace des états de la nature chez SAVAGE. Troisièmement, ce dernier utilise largement les décisions constantes dans ses démonstrations, ce qui n'a pas lieu dans la présentation conditionnelle. Par ailleurs, mais seulement dans un cas fini on peut montrer que le passage d'un modèle à l'autre est possible en utilisant une formalisation matricielle, mais cette équivalence ne peut que rester théorique et sans grande valeur pratique.

III C - Elargissement des théories d'utilité espérées.-

Nous avons vu ci-dessus diverses théories de décision dans lesquelles fonction d'utilité et distribution de probabilité subjectives étaient définis conjointement pour être compatiblesavec une relation de préférence définie. La liaison entre cette probabilité et cette fonction est réalisée par un opérateur de type espérance mathématique. De plus avec les distinctions faites en I, cette distribution subjective incorpore toute l'incertitude temporelle (c'est d'ailleurs celle-ci qui lui confère sa caractéristique essentielle de subjectivité).

Il est néanmoins souhaitable d'envisager d'autres types de liaison entre utilité et probabilité subjectives. C'est ce que nous allons introduire en nous reportant à nouveau à BERNARD (21)

En reprenant la question de l'utilité avec une optique voisine de celle adoptée en I, la question générale posée est : quelle est l'utilité d'une variable aléatoire pour le décideur ?

Par variable aléatoire on peut entendre (à partir de la définition de FRECHET) :

Soit $\mathcal{H}$ l'ensemble des états de la Nature et soit R un élément de cet ensemble obtenu par le hasard : une variable aléatoire sur $\mathcal{H}$ est un nombre H dont la valeur est parfaitement déterminée lorsque R arrive et est tel que pour tout nomb re certain h, il y ait une probabilité déterminée que $H < h$ probabilité notée $P(H < h)$.

Avec cette définition, on s'assure que H a une distribution ("objective", "axiomatique", "naturelle" ou "subjective"). De plus on supposera que $\mathcal{H}$ est fini et dénombrable ($\mathcal{H} = \{h_i \ ; \ i=1....n\}$). Le fait que $\mathcal{H}$ soit l'ensemble des états réels de la Nature implique que $P = \{p_i \ ; \ i=1,....n$ et $\sum_1^n p_i = 1\}$.

Soit $\mathcal{F} = \{f_j \ ; \ j=1....n\}$, l'ensemble des actes possibles, fini et dénombrable. Soit $\mathcal{D} = \{d_k, \ k=1,...\ell\}$ l'ensemble des décisions possibles, fini et dénombrable également. Par définition $f_j(h_i) \in \mathcal{D}$.

En définissant les termes à priori et à postériori par rapport à la réalisation de h_j, les questions sont alors formulées par :

(i) Quelle est à priori et à postériori, l'utilité $v(f_j(h_i))$ d'une décision incertaine par le décideur ?

(ii) Quelle est à priori et à postériori, l'utilité $u(f_j) = v(f_j(h_i))$ d'un choix f_j pour le décideur, et cela en présence de l'incertitude sur les résultats ?

Ci-dessus le terme à postériori fait supposer que l'incertitude n'existe plus. Nous supposons alors qu'il existe des utilités acceptables $v''(f_j(h_i))$ et $u''(f_j)$ qui seront des utilités d'actes certains.

Examinons alors l'utilité à priori d'actes incertains, soit $u'(f_j)$, les décisions $f_j(h_i)$ ayant une distribution de probabilités donnée à priori. f_j est vecteur dont les éléments $f_j(h_i)$ sont des valeurs diverses mais de même dimension. De plus ces éléments $f_j(h_i)$ ne coexistent pas et forment un "vecteur stochastique" tant que H_i n'est pas réalisé (H_i étant défini plus haut comme H. Nous en sommes réduits au véritable problème : quelle est l'utilité à priori d'un tel vecteur de valeurs incertaines ?

L'école "anglo-saxone" de l'utilité adoptant le critère "d'espérance d'utilité" conduit à prendre

$$u(f_j) = \sum_{i=1}^{n} v(f_j(h_i))\, p_i$$

Sans être limitée aux seuls cas finis, une telle forme a été rencontrée ci-dessus : D'autres formes sont envisageables. Postulons que le décideur peut estimer subjectivement, et cela conformément à certains critères de rationalité (au sens de critères universellement adoptés, voir (18)), l'utilité à priori d'une décision incertaine de probabilité donnée. En d'autres termes, on suppose que le décideur sait sortir de l'alternative suivante :

- disposer d'une valeur A disponible avec certitude

- disposer d'une valeur B incertaine dont la probabilité à priori est p

Ainsi on écrira $v\, f_j(h_i) = v(f_j(h_i),p_i)$.

A partir de cet axiome, on supposera de plus le critère d'additivité suivant :

$$u(f_j) = \sum_{i=1}^{n} v(f_j(h_i),p_i)\, p_i$$

Une telle formulation de l'utilité en tant qu'axiomatique générale de rationalité innove par rapport aux seules formulations d'utilité linéaire. Si certaines formes de telles utilités sont proposées pour répondre de manière plus précise à telle ou telle rationalité de comportement particulière, il ne semble pas exister, à ma connaissance, de théorie (axiomes de structures, conditions d'existence, propriétés diverses) par de telle fonction d'utilité.

* *

*

Annexe : Schéma de la théorie de SAVAGE

Axiomes

1) $\exists$ un préordre complet sur l'ensemble F des actes, $\leqslant$

2) si $f, f' \varepsilon F$ coïncident respectivement avec $g, g' \varepsilon F$ dans $\tilde{A}$
si f, g coïncident respectivement avec f, g' dans A
alors $f \leqslant g \Longleftrightarrow f' \leqslant g'$

3) si $f = d$, $f' = d'$ et B
alors
$f \leqslant f' | B \Longleftrightarrow d \leqslant d'$

4) si f_A, f_B sont des actes en escalier, soit :
$f_A(h) = d$ si $h \varepsilon \tilde{A}$
$f_A(h) = d'$ si $h \varepsilon \tilde{A}$,
respectivement B, alors
$f_A \leqslant f_B$ ne dépend que de d et d' pourvu que
$d < d'$

5) $\exists$ un couple de décisions d et d' : $d < d'$.

6) si $g < f$, une légère modification de g en g_* est possible avec encore
$g^* < f$

7) $(f \leqslant g(h)) | B \ \forall h \varepsilon B \Rightarrow (f \leqslant g) | B$

Définitions et théorèmes

Préférences conditionnelles

Définition de la préférence condition. sur les actes : $(f \leqslant g | B$

La préférence sur les décisions est intrinsèque

Théo : si $h \varepsilon B_i$, partition de B et si
$f(h) = d_i$
$g(h) = \delta_i$
alors
$d_i < \delta_i \Rightarrow (f \leqslant g) | B$

Probabilités qualitatives

$\leqslant\cdot$ est une proba. qualitative

Def. d'une relation d'ordre sur les évènements $\leqslant\cdot$

Probabilités quantitatives

La probabilité quali. est finie

$\exists$ une n-partition presque uniforme

$\exists$ une probabilité quantitative compatible.

La probabilité quanti est serrée.

Utilité

Définition des loteries $\underline{f} = \sum_i d_i f_i$

Si $\underline{f}_1 \leqslant g \leqslant f_2$
$\exists$ ρ unique :
$\rho \underline{f}_1 + (1-\rho) f_2 \asymp g$

Définition et propriétés de l'utilité

$\exists$ une fonction d'utilité sur les loteries

Extension de la fonct. d'utilité aux actes bornés

U est bornée $\Rightarrow$ tous les actes sont bornés.

BIBLIOGRAPHIE

[1] *K.J. ARROW* : Exposition of a theroy of choice under incertainty Synthese 16 (1966) p.253-269.

[2] *R.J. AUMANN, F.J. ANSCOMBE* : A definition of subjective probability. Ann. Math. Stat. 34 (1963) p. 199-205.

[3] *R.J. BERAN* : Upper and lower risks and minimax procedures. Proc. of the Sixth Berk. Symp. 1 (1972) p. 1-16.

[4] *G. BERNARD* : Valeur, temps, incertitude dans l'utilité CNRS - Sém. d'econometrie (1964)

[5] *P. BILLINGSLEY* : Convergence of probability measures Wiley (1968)

[6] *D. BLACWELL, M.A. GIRSHICK* : The theory of games ans statistional decisions Wiley (1954)

[7] *H. CHERNOFF, L.E. MOSES* : Elementry decision theory Wiley (1959)

[8] *H. CHERNOFF* : Rational selection of decisions functions Econometrica 22 (1954) p. 422-443.

[9] *La décision* - Colloque CNRS Aix-en-Provence (1967)

[10] *J. DONID, E. HAYAT, M. TENENHAUS* : Le problème bayesien : présentation du modèle - Metra VIII - 2 (1969) ; application au problème de l'aide à l'action médicale Métra IX (1970).

[11] *K.A. DOKSUM* : Decision theory for some non parametric models Proc. of the sixth Berk. Symp. 1 (1972) p. 331-343.

[12] *D. DAVIDSON, P. SUPPES* : A finistic axiomatization of subjective probability and utility Econometrica 24 (1956) p. 264-275.

[13] *P.C. FISHBURN* : Decision and value theory. Wiley (1964)

[14] *P.C. FISHBURN* : A general theory of subjective probabilities and expected utilities. Ann. Math. Stat. 40-4 (1969) p. 1419-1429

[15] *P.C. FISHBURN* : Methods of estimating additives utilities Sciences 13-7 (1967)

[16] *P.C. FISHBURN* : Preference based definition of subjective probability Ann. Math. Stat. 38-6 (1967) p. 1605-1617.

[17] *T. FINE* : a note on the existence of quantitative probability

[17] *T. FINE* : A note on the existence of quantitative probability
Ann. Math. Stat. 42-4 (1971) p. 1182-1186.

[18] *FOURGEAUD, LENCLUD, SENTIS* : Critere de choix en avenir partiellement incertain.
R.I.R.O. 14 (1968) p. 9-20.

[19] *R. GUESNERIE, J.Y. JAFFRAY* : Probabilités subjectives et utilité de
Von Neumann et Morgenstern
Bull. de Math. Economiques

[20] *M. HAUSNER* : Multidimensional utilities and decision processes
Thrall, Cooms, Davis - Wiley (1954)

[21] *J. HAJEK* : On basic concepts of statistics
Proc. of the fifth Berk. Symp. 1 (1966) p. 139-162.

[22] *H. KUDO* : On partial prior information and the property of
parametric sufficiency.
Proc. of the fifth. Berk. Symp. 1 (1966)

[23] *N.E. JENSEN* : An introduction to Bernouillan utility theory : I
utility fonctions
Swodish jour. of Economics (1967)

[24] *E.L. LEHMANN* : Testing statistical Hypotheses
Wiley (1959)

[25] *R.D. LUCE* : Sufficient conditions for the existence of finitely
additive probability measures
Ann. Math. Stats. 35 (1967), p. 780-786.

[26] *R.D. LUCE, H. RAIFFA* : Games and decisions
Wiley (1957)

[27] *R.D. LUCE, D.H. KRANTZ* : Conditional expected utility
Econometrica 39-2 (1971)

[28] *MORINEAU* : Introduction à la théorie de l'utilité
Th. 3ème cycle - Paris (1971)

[29] *P. MALGRANGE* : Critères de choix en avenir incertain.
Bull. de Math. Economiques 4- (1970)

[30] *P. MASSE* : Le choix des investissements
Dunod (1959)

[31] *A. PLANT* : The individual decision problem
Non publié - Univ. de Montréal

[32] *J. PFANZAGL* : Theory of Measurement
Wiley (1968)

[33] *J.W. PRAT, H. RAIFFA, R. SCHLAIFER* : The fondation of decisions under incertainty : an elementary exposition
Jour. Ann. Stat. Ass. 59. (1964) p. 359-375.

[34] *H. RAIFFA, R. SCHLAIFER* : Applied Statistical decision theory
M.I.T. Press (1961).

[35] F.P. RAMSEY : The fondations of mathematics and other logical essays
Harcant, Brace and Co. N.Y. (1931)

[36] P. SUPPES : The role of subjective probability and utility in decision making
Proc. of the Third Berk. Symp. 1 (1956)

[37] L.J. SAVAGE : Fondations of statistics
Wiley (1954)

[38] C. VILLEGAS : On qualitative probability σ-algebras
Ann. Math. Stat. 35 (1964) p. 1787-1796.

[39] VON NEUMANN, O. MORGENSTERN : Theory of games and economic behaviour
Princeton Univ. Press. (1947).

[40] A. WALD : Statistical decision functions
Wiley (1950).

-=-=-=-=-=-=-