Maths et stats en Gestion

Chapitre V
Données et Statistiques, Distributions

où l'on étudie comment construire des statistiques à partir d'un ensemble de données, où, plus généralement, l'on introduit la notion de distribution des données, discrète et continue, et de statistiques, qui proposent des représentations à la fois partielles et plus synthétiques d'une distribution.

Données statistiques et variables

Définition

Les *données* sont la collection extensive des propriétés (ou caractéristiques) connues (ou mesurées) d'une population. On parle de *données statistiques* d'une population.

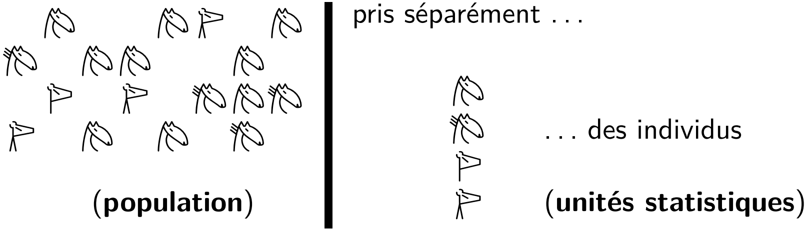
Définition

Les *variables* ou caractères statistiques sont les critères sur la base desquels est décrite la population observée

Population et unités statistiques

- La *population* : désigne l'ensemble des unités statistiques sur lequel portent les observations (exemple : la population étudiante de Tours ou les camemberts en vente un jour donné chez XTOK)

- Les *individus* ou unités statistiques : il s'agit des unités composant la population observée (dans les exemple précédents : les étudiants ou les différents types de camembert)



Objet et tâches de la statistique

- L'objet de la statistique est de représenter de manière simple les propriétés d'une population. À cet effet, il est essentiel de pouvoir les *agréger*, c'est-à-dire d'en présenter un résumé.

- Par exemple, si un fichier comporte 2000 données, impossible de toutes les présenter. L'objet de la statistique est alors de construire des indicateurs qui permettent de caractériser les spécificités de chaque série de données et/ou de les comparer.

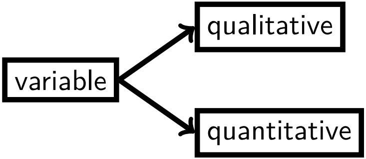
Tâches

- **1** Construire les données, cad constituer un ensemble d'unités statistiques, qui permettent de répondre à une question donnée

- **2** Traiter ces données. Nous nous concentrerons principalement sur ce second aspect.

Différents types de variables

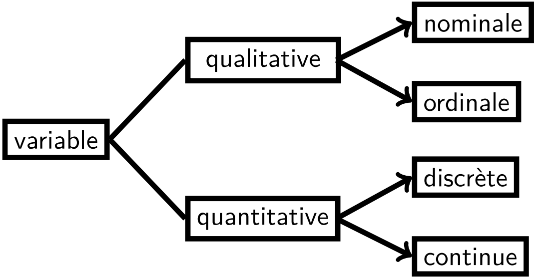
Chaque unité statistique est décrite à travers une ou plusieurs variables. Ces variables sont soit *qualitatives*, c'est-à-dire qu'elle ont des modalités qui ne sont pas mesurables, qui expriment un état, soit *quantitatives*.



- Les *modalités ou classes* sont les situations que présente la variable (exemple de la variable nationalité : française, allemande, exemple de la forme des fromages : rond, bûche, carré)

Typologies des variables

- On distingue les variables *qualitatives nominales* (il n'existe pas d'ordre entre les modalités, exemple : statut matrimonial) et les variables *qualitatives ordinales* (on peut ordonner les modalités, exemple : variable mention : P, AB, B, TB)



- On distingue les variables *quantitatives discrètes* ne pouvant prendre que certaines valeurs (exemple : variable nombre d'enfants (0,1,2,3,. . .mais pas 1.8 ni 1227) des variables *quantitatives continues* pouvant prendre n'importe quelle valeur

Représentation et Traitement des données

Pour représenter les valeurs prises par les variables dans une population, on a recours à des tableaux et graphiques qui permettent de regrouper les variables par modalités, et à des statistiques que l'on définit à partir de mesures numériques qui permettent de synthétiser les variables représentées.

- Tableaux : regroupement par modalités

Les tableaux permettent une représentation simplifiée de la population, par exemple en faisant un comptage des différentes modalités.

- Statistiques : différentes mesures des groupes

Les *variables* ou caractères statistiques sont les critères sur la base desquels est décrite la population observée

Exemple 1 : les salaires dans l'OCDE

Définition des bas salaires

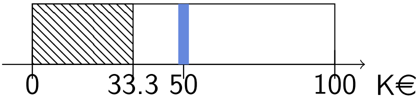
Dans une distribution donnée, un salaire est bas quand il est inférieur au 2/3 du salaire médian.
On appelle Fréquence des bas salaires (F_{BAS}), la proportion des bas salaires dans une distribution.

Exemple 1 : les salaires dans l'OCDE

Définition des bas salaires

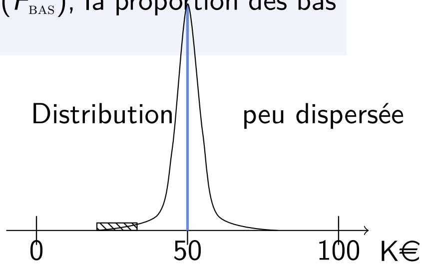
Dans une distribution donnée, un salaire est bas quand il est inférieur au 2/3 du salaire médian.
On appelle Fréquence des bas salaires (F_{BAS}), la proportion des bas salaires dans une distribution.

Distribution uniforme



$F_{\text{BAS}} = 33,33\%$

Distribution peu dispersée



$F_{\text{BAS}} \approx$ faible

▸ COMPARATIF PAR PAYS

Exemple 1 (suite)

- Ainsi, pour juger de l'inégalité des salaires, on utilise un seul indicateur, ici, la proportion des "salaires bas"

- Une bonne analyse descriptive revient à chercher l'indicateur le plus simple qui permet de voir une propriété des différentes modalités, qui soit robuste au choix de l'indicateur

Echantillon et population

Avant de clore cette introduction, concernant les données, on introduit souvent la différence entre population totale et Echantillon.

- Un *échantillon* : désigne une partie des unités statistiques sur lesquelles portent l'étude. C'est la partie qui est observée, connue de manière extensive. Si l'échantillon est de taille raisonnable, il représente la population. Il s'agit alors de décrire comment on peut déduire de l'échantillon des caractéristiques de la population.

- Un des objectifs des statistique est de décrire (suivant des critères très précis) comment la connaissance des caractéristiques d'un échantillon donne de l'information sur la connaissance de la population étudiée.

Plan du cours

1) Représentation objective des données
   - Distributions discrètes et continues
   - Statistiques de position et de dispersion

2) Évaluation subjective d'une distribution
   - Préférences moyenne - variance
   - Théorie de l'espérance d'utilité

1. Représentation objective des données

Probabilités et distributions

- Cardan (1501-1576) : « le joueur savant ».

- Probabilité d'un événement = $\sharp$ résultats favorables / $\sharp$ événements possibles.

- « Pile » a une probabilité de 1/2.
  Probabilité[obtenir un six au moins une fois en 3 lancés] $\geq 1/2$?
  c'est $= 1 - (5/6)^3 = 0,4213$ (de Méré).

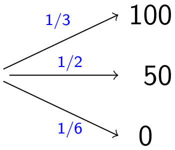
  Remarque : On aurait pu penser que comme en un lancé, la probabilité de voir apparaître 6 est de 1/6, en trois lancés, elle est trois fois plus grande (parce qu'on additionne la probabilité d'apparaitre au premier tour, puis la probabilité d'apparaitre au second tour et la proba d'apparaître au 3e tour. En fait, c'est méconnaître le fait que si 6 n'est pas apparu au premier tour, la probabilité qu'il apparaisse au second tour doit être déclassée du fait qu'il n'est pas apparu au premier tour. Puis que si le 6 n'est pas apparu au 1er et 2e tour, sa probabilité d'apparaître au troisième tour doit être déflatée du fait de ne pas être apparu au 1er et au 2e tour. En d'autres termes : La probabilité d'avoir 6 au moins une fois en trois lancés est : $\frac{1}{6} + \frac{5}{6} * \frac{1}{6} + \left(\frac{5}{6}\right)^2 * \frac{1}{6} < \frac{1}{2}$.

Distributions (ou risque)

- **Distributions discrètes**

Il y a un nombre fini d'évènements possibles $i \in \mathcal{I}$, chacun avec probabilité p_i. Cette association à chaque évènement de sa probabilité, c'est ce qu'on appelle la **distribution** des risque. Cette distribution satisfait toujours la contrainte

$$\sum_{i \in \mathcal{I}} p_i = 1$$

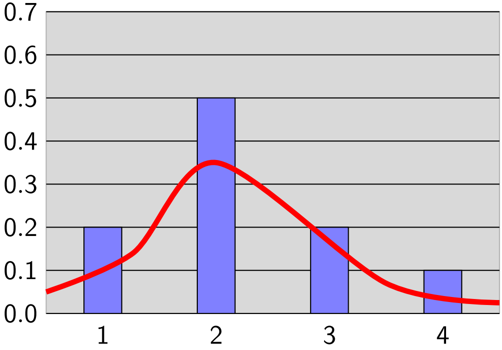


- **Distributions continues**

Il y a un nombre infini, voire continu d'évènements possibles : chacun, pris isolément apparaît avec une probabilité nulle. Ces distributions sont représentées par des **distributions de probabilités** et/ou par des **fonctions de repartition**

Représentation de Distributions de probabilités

- en bleu, diagramme en tuyau d'orgue représentant une distribution discrète
- en rouge, une distribution d'une variable continue sous-jacente.



▶ dans sa version continue, la surface comprise sous la distribution, entre deux valeurs a et b représente la probabilité d'être entre a et b :

$$Prob(a \leq X \leq b)$$

Fonctions de répartition

Dans le cas d'une distribution continue, le poids relatif des évènements de faible gain par rapport aux évènements de gains plus élevés.

$$F(x) = Prob(X \leq x)$$

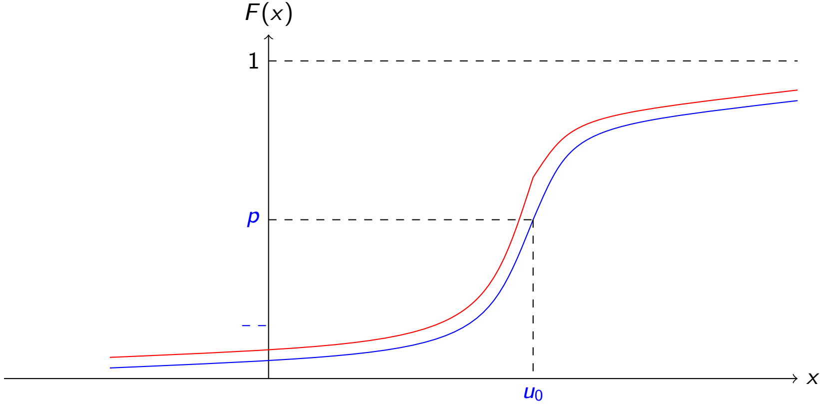
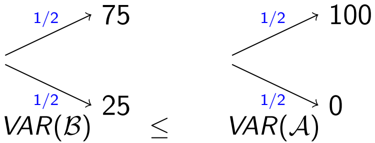


Figure – deux fonctions de répartitions : F et G

- **Moyenne** $\sum_i$ probabilités * richesses
  dans l'exemple précédent, moyenne=50
- **Variance** une mesure de la distance à la moyenne.
  exemple : la distribution $\mathcal{A}$ a une plus grande variance que la distribution $\mathcal{B}$.



- **Modes** représente le/les évènements avec la plus grande probabilié
- **Fractiles** Divise la population en classes égales, représentées par une richesse pivot.

Statistiques - Pour aller plus loin

Il y a en fait deux familles de statistiques :

les *statistiques de position* dont l'objectif est de donner un ordre de grandeur des valeurs observées

les *statistiques de dispersion* qui évaluent le niveau d'étalement de la série autour de la valeur centrale.

Les paramètres de position (ou valeurs centrales) sont des valeurs numériques qui « résument » une série statistique en caractérisant l'ordre de grandeur des observations. Ils s'expriment dans la même unité que les observations. Les paramètres de position permettent de situer la position de plusieurs séries comparables. Lorsque la distribution est parfaitement symétrique, mode, moyenne et médiane sont confondues.

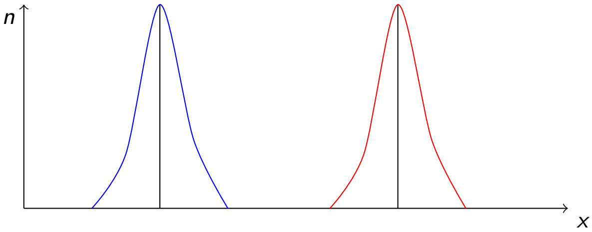


Figure – Deux statistiques semblables à des positions différentes

Les deux courbes ont la même allure, mais ne se positionnent pas du tout au même endroit sur l'axe des valeurs (des modalités). Les paramètres de position le mettent clairement en évidence.

Moyenne arithmétique d'un ensemble de N nombres

Définition

La moyenne arithmétique de N nombres est égale à la somme de ces nombres divisée par leur nombre.

$$\bar{x} = \frac{1}{N} \sum_i x_i$$

- Exemple simple :

3 individus, gagnent respectivement 10.000 euros, 20.000 euros et 30.000 euros. La moyenne de leur revenu est 20.000 euros.

- Remarque :

La moyenne arithmétique est exactement la quantité qui pourrait être identiquement distribuée à chaque individu. En effet, la conséquence directe de la définition de $\bar{x}$ est : $N\ \bar{x} = \frac{1}{N} \sum_i x_i$.

Moyenne arithmétique d'une distribution

Dans le cas d'une distribution, il faut prendre en compte la fréquence d'apparition de chacune des réalisations.

- **Cas discret : à partir du tableau de fréquences**

Une variable X prend les valeurs x_i avec la fréquence f_i pour $i = 1, \ldots, N$. La moyenne de cette variable est

$$\bar{X} = \sum_i f_i \; x_i$$

  - la comparaison avec la formule du transparent précédent est immédiate. $\frac{1}{N}$ est remplacé par la fréquence (individualisée) de chaque réalisation f_i.

- **Cas continu : à partir de la fonction de distribution** Un variable X est définie par sa fonction de distribution $f(x)$, sa moyenne est

$$\bar{X} = \int_{-\infty}^{+\infty} x \; f(x) dx$$

Distribution représentée comme un bruit blanc autour d'une moyenne

Quand il n'y a pas trop de dispersion autour de la moyenne, il est assez naturel de représenter une distribution comme étant une valeur certaine autour de laquelle il y a un bruit blanc.

Définition Un bruit *blanc* est une variable aléatoire $\tilde{\varepsilon}$ dont la moyenne est nulle ($E(\tilde{\varepsilon}) = 0$) dont les réalisations sont faibles en regard de la valeur (de position) x.

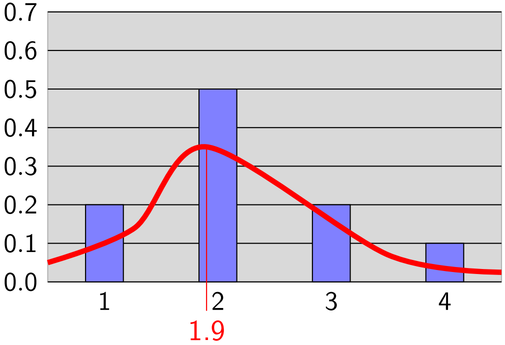
Exemple Soit la variable aléatoire $\mathcal{A}$ suivante, on peut la représenter comme la somme de sa moyenne et du bruit blanc $\tilde{\varepsilon} = \mathcal{A} - E[\mathcal{A}]$:

$$\mathcal{A}\ \begin{cases} \xrightarrow{1/2} 50,3 \\ \xrightarrow{1/2} 50,1 \end{cases} \quad = \quad 50,2 \quad + \quad \tilde{\varepsilon}\ \begin{cases} \xrightarrow{1/2} 0,1 \\ \xrightarrow{1/2} -0,1 \end{cases}$$

Tout se passe comme si un agent qui était exposé au risque représenté par $\mathcal{A}$ recevait la valeur sûre 50,2, dans un premier temps, cad la moyenne, et qu'avec égale probabilité, il perde (ou il gagne) à partir de cette valeur sûre -0,1 (ou +0,1).

Le mode, défini pour toute variable aléatoire

Le mode d'une variable qualitative ou quantitative discrète : modalité dont la fréquence (absolue ou relative) est la plus élevée. Dans le cas où une variable continue a été regroupée en classes, le mode est la classe dont la fréquence est la plus élevée.

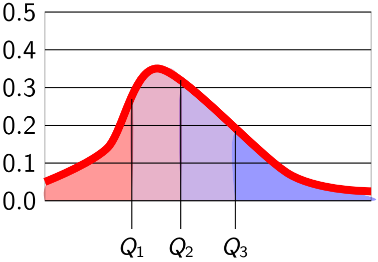


Dans l'exemple ci-dessus, le mode de la variable discrète est 2, celui de la variable continue, 1.9

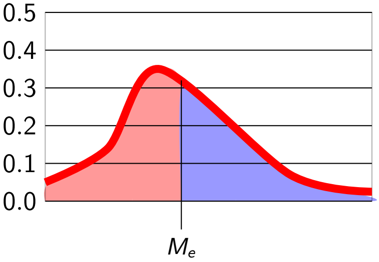
Les quantiles : séparer une distribution en parts égales

Lorsque la variable est ordonnée, si elle est continue, et parfois même quand elle est discrète ordonnée, on cherche à représenter les différentes parties d'une distribution. On nomme *quantiles* les valeurs qui permettent de séparer la distribution en parts égales. L'opération varie avec le nombre de parts.

Dans le cas d'une séparation en quatre, les *quartiles* sont les valeurs qui partagent la distibution en 4 parties de 25%.



Dans le cas d'une séparation en deux, la *médiane* est la valeur qui partagent la distibution en 2.

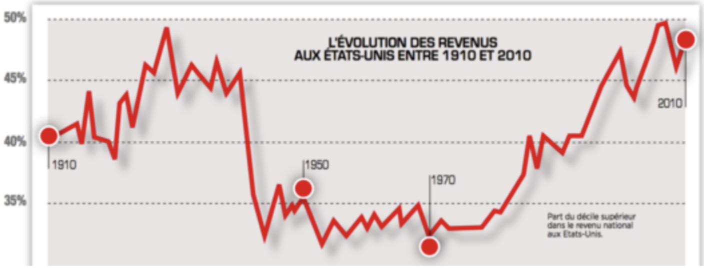


Le quantile, défini pour les variable ordonnées

Definition

les quantiles sont les valeurs de la variable partageant la série classée par ordre croissant de la variable en k sous-ensembles égaux.

- $k = 2$ c'est la *médiane* M_e
- $k = 4$ c'est les *quartiles* Q_1, Q_2, Q_3
- $k = 10$ c'est les *deciles* $D_1, D_2, \ldots, D_9$
- $k = 100$ c'est les *centiles* $C_1, C_2, \ldots, C_{99}$



Sur le graphique du transparent précédent (publié dans le livre de Thomas Piketti intitulé *Capital*), on constate que la part du décile supérieur dans le revenu national américain est passée de 45-50% dans les années 1910- 1920 à moins de 35% dans les années 1950, avant de remonter à son niveau initial dans les années 2000-2010.

C'est, d'après T. Piketti la conséquence mécanique du fait que, hors périodes exceptionnelles (les deux guerres puis les reconstructions), le rendement du capital est supérieur au taux de croissance.

2. Évaluation subjective des distributions

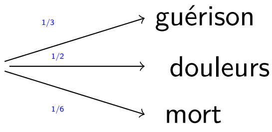
Applications au risque en santé et en économie

Incertain : Etats de santé, voire probabilités

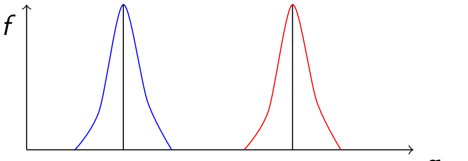
L'incertain concerne ce que sera l'état de santé futur. Un choix médical peut par exemple conduire à différents états de santé futurs possibles, $s_i \in \mathcal{S}$.

La recherche aidant, ces états peuvent être associés à une probabilité p_i : Il s'agit alors d'une **distribution** des risques santé qui satisfait la contrainte $\sum_i \; p_i = 1$ (ou $\int f(x)dx = 1$)

Les distributions sont parfois discrètes, comme dans l'efficacité d'un traitement :



elles sont souvent continues, comme dans la distribution des globules blancs chez les patients sains ou malades :



Incertain et Incertitude : probabilités, ou non

Risky choice (ou incertain) : « each action leads to one of a set of possible specific outcomes, each outcome occurring with a known probability. The probabilities are assumed to be known to the decision maker » Luce and Raiffa, Games and Decision, 1957, p13

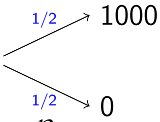
Uncertain choice (ou incertitude) : « each action has as its consequences a set of possible specific outcomes, but the probability of these outcomes are completely unknown or are not even meaningful » Luce et al.

Représenter un problème de décision en santé

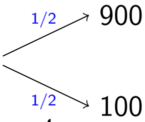
Si l'on a le choix entre deux stratégies médicales, chacune représentée par une distribution, laquelle devrait-on choisir ?

Choix A : Greffe du rein	Choix B : Dialise
$\xrightarrow{1-p}$ guérison $\xrightarrow{p}$ mort	$\xrightarrow{1}$ Etat dialisé

La littérature expérimentale cherche à comprendre comment les agents économiques évaluent différentes expositions au risque, au travers de choix entre plusieurs alternatives comme par exemple entre les loteries $\mathcal{B}$ et $\mathcal{A}$ suivantes

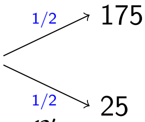


$\mathcal{B}$

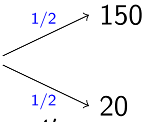


$\mathcal{A}$

ou entre les loteries $\mathcal{B}'$ et $\mathcal{A}'$ suivantes



$\mathcal{B}'$



$\mathcal{A}'$

Deux explications pour les cas présentés : le Spread et la FOSD

Souvent entre $\mathcal{A}$ et $\mathcal{B}$ les agents préfèrent $\mathcal{A}$.

On l'explique en remarquant que $\mathcal{B}$ est en *spread* de $\mathcal{A}$ à moyenne constante. C'est à dire qu'il s'agit de déplacer du poids du centre de la distribution vers les queues, comme par exemple dans les distribution suivantes f et g :

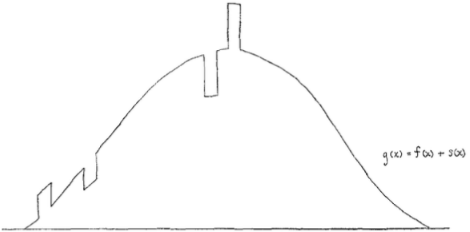


FIGURE 3

Toujours, entre $\mathcal{A}'$ et $\mathcal{B}'$ les agents préfèrent $\mathcal{B}'$.

On l'explique en remarquant que $\mathcal{B}'$ domine $\mathcal{A}'$ au sens de la First Order Stochastic Dominance, c'est-à-dire que la distribution $\mathcal{B}'$ rémunère plus tous les états de la nature que la distribution $\mathcal{A}'$.

Recherche d'un Critère de préférence

Dans le cas général, pour comprendre le comportement d'un agent, et plus précisément les choix qu'il fait lorsqu'il doit choisir entre plusieurs loteries $\mathcal{A}$ et $\mathcal{B}$, on essaye d'établir un *critère de notation* des différentes loteries.

Par ailleurs, si l'on revient à des applications comme la santé, les états de santé ne peuvent pas être remplacés par des conséquences monétaires, et l'on doit tout de même faire des choix.

Selon quels types de critères les agents rationnels classent-ils les loteries ?

Espérance d'utilité

Définition

Plutôt que de prendre l'espérance de la lotterie, tout se passe comme si l'agent appréciait les différents revenus à travers un filtre. Ainsi, l'agent voit le revenu x à travers son utilité ressentie $u(x)$. *Son critère d'évaluation est l'espérance de ces utilités.*

$$U\left(\begin{array}{l} \nearrow^{1/3}\; x \\ \xrightarrow{1/2}\; y \\ \searrow_{1/6}\; z \end{array}\right) = \frac{1}{3}\ u(x) + \frac{1}{2}\ u(y) + \frac{1}{6}\ u(z)$$

▶ **Important :** Ce critère d'utilité s'applique soit à des loteries monétaires, soit à des loteries dans lesquels les réalisations x, y et z sont qualitatives, comme en santé.

L'espérance d'utilité, surtout en santé, peut être comprise comme une opération de cuisine. On ne sait pas trop comment faire avec ces états qualitatifs x, y et z, et on les remplace par des chiffres.

En fait, dans une approche axiomatique, Von Neuman Morgerstern a démontré en 1958 qu'étant donné des états de la nature $S = \{1, 2, \ldots, n\}$, si on suppose que le décideur a des préférences sur les distributions qui vérifient des axiomes assez simples (préordre partiel, continuité, indépendance et non trivialité), alors,

Théorème (de Von Neuman Morgerstern) : si la relation de préférence vérifie les 4 axiomes précédents, alors il existe une fonction d'utilité des états de la nature $u(\cdot)$ telle que :

$$x \succeq y \quad \Longleftrightarrow \quad \sum_s p_s u(x_s) \geq \sum_s p_s u(y_s)$$

(*La fonction* $u(\cdot)$ *est unique à une relation affine près*)

Décision dans un cas de décision thérapeutique

L'application du théorème VNM indique que le choix entre deux stratégies médicales peut se représenter comme le choix entre deux distributions d'utilités (QALY), dont on évalue l'espérance.

Il est fréquent dans la littérature médicale de normaliser le bon état de santé à 1, et le décès à 0 :

Choix A : Greffe du rein	Choix B : Dialise
$\xrightarrow{1-p}$ 1(guérison) $\xrightarrow{p}$ 0 (mort)	$\xrightarrow{1}$ d(Etat dialisé)

Ainsi, dans ce modèle, le choix d'opérer correspond à : $1 - p \geq d$, soit $p \leq 1 - d$.

EU : Interprétation pour les distributions monétaires

Le théorème de VNM s'applique aussi aux distributions monétaires, il souligne le rôle de l'utilité dans la décision : Plutôt que de considérer l'espérance d'une distribution, tout se passe comme si l'agent appréciait ses différents revenus possibles à travers un filtre, l'utilité ressentie $u(x)$ avant de calculer l'espérance de ces utilités.

$$U\left(\begin{array}{l} \xrightarrow{1/3} x \\ \xrightarrow{1/2} y \\ \xrightarrow{1/6} z \end{array} \right) = \frac{1}{3}\ u(x) + \frac{1}{2}\ u(y) + \frac{1}{6}\ u(z)$$

Remarque : La seule statistique utilisée est la moyenne. On n'a pas recours aux autres statistiques qui aident à appréhender les distributions monétaires : variance, min, max, mediane, [...]

un exemple

Lorsque les distributions concernent une richesse future, on a une compréhension plus immédiate de certain mécanismes de choix. Imaginons que lorsque vous avez commencé vos études, vous aviez le choix entre deux filières qui différaient en particulier sur le profil de revenus annuels futurs en début de carrière :

Filière A :		Filière B :	
$\xrightarrow{1/3}$	200K	$\xrightarrow{1/3}$	150K
$\xrightarrow{1/2}$	100K	$\xrightarrow{1/2}$	100K
$\xrightarrow{1/6}$	0K	$\xrightarrow{1/6}$	60K

Vous préférerez la filière B (qui conduit à un revenu moins dispersé) si $u(200K) - u(80K)$ n'est pas tellement plus élevée que $u(150k) - u(130K)$.

C'est en particulier le cas si la fonction $u(\cdot)$ est très concave. [On le vérifie ici avec $u(x) = \sqrt{x}$] - Alors que $E[A] > 116$ et $E[B] = 110$.

EU : les QALYS pour les états de santé

L'espérance d'utilité n'est donc pas un bricolage qui tend à introduire des nombres à la place d'un état de santé, mais elle part du résultat selon lequel si un patient a des préférences, alors ces dernières s'expriment comme une espérance d'utilité : celà suppose

- de connaître la probabilité des différents états
- d'évaluer la QALY d'un état de santé.

La littérature sur les QALY est assez abondante depuis les années 80. [1] Des tables de QALY classifiant différents états de santé ont été publiées. On doit néamoins souligner que ce distingue les préférences, les choix, d'un individu à l'autre est justement ces valeur des QALY, par essence, personnelles.

1. *The Qaly Toolkit*, par Claire GUDEX et Paul KIND, University of York, 1990.

SEU : les QALYS et les croyances subjectives

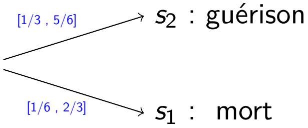
Cette théorie de l'espérance d'utilité subjective n'est pas non plus un bricolage, mais elle impose

- d'emettre des *croyances* sur l'occurence des différents états de la nature
- d'évaluer la QALY d'un état de santé.

Ambiguïté de l'information

Une piste plus récente est de considérer non pas que l'on a pas d'information (sur les probabilités des états), mais qu'on a une information ambigüe, dans le sens d'une *information qui est comprise de manière pauvre*[2].

Plus concrètement, on connait des distributions de manière imprécise, dans le sens où la vraie distribution se trouve dans un sous-ensemble de distributions.



2. Massimo Guidolin et Francesca Rinaldi, *Ambiguity in Asset Pricing and Portfolio Choice : a review of the litterature*, 2010.

MAXIMIN en cas d'ambiguité de l'information

On montre de manière axiomatique, que les préférences avec une information ambigue se représentent comme un maximin, les états de la nature étant représentés par des QALYS

$$U\left(\begin{array}{l} \xrightarrow{[1/3\ ,\ 5/6]} s_2 \\ \xrightarrow{[1/6\ ,\ 2/3]} s_1 \end{array}\right) = \min_{p_2\in[1/3\ ,\ 2/3]} \quad (1-p_2)u(s_1)+p_2u(s_2)$$

C'est à dire que l'on évalue une situation ambigue par la moindre valeur que pourrait prendre son espérance d'utilité pour les différentes probabilités possibles.
Exemple : si $s_1 = 0$ et $s_2 = 1$: $U = 1/3$: le maximin a évalué cette EU à partir de la distribution la plus défavorable au patient : une probabilité de décès de $2/3$.

Principe de précaution

En cas d’ambiguïté dans l’information : utiliser un maximin (l’interprétation est : comme si la nature jouait systématiquement contre nous)

Quand l'information sur les conséquences d'un choix thérapeutiques sont mal connues, on peut toujours avoir recours à des tests, qui permettent d'améliorer l'information.

Exemple Patients may be infected by a virus : not lethal but costs 10 000 Euros to cure. A prevention treatment (vaccine) exists : cost 1 000 Euros.

Il existe une croyance a priori sur la présence des virus dans la population : 5%. Ceci permet de calculer le coût moyen de ne pas traiter, comparé au coût du vaccin. [neutralité au risque] : ici, ne pas vacciner.

Un test est possible afin d'améliorer la connaissance de la pathologie du patient dont les probas ($a + b + c + d = 1$) sont les suivantes :

$$\begin{array}{cccc} Virus, + & a & Virus, - & b \\ Sain, + & c & Sain, - & d \end{array}$$

La sensibilité du test : $\frac{a}{a+c}$ (axe vertical : diagramme de ROC)
La spécificité du test : $\frac{d}{b+d}$ (axe horizontal : diagramme de ROC)
On peut alors calculer le coût moyen de la maladie, étant donné le test choisi et de choisir le meilleur test possibles

En conclusion, steps toward a decision

Étant donné un ensemble d'états de santé possibles et un ensemble de choix ou d'actions, que faire ?

- Form beliefs about the state of nature (Elire des croyances à propos des états de santé).
  - Perception as a Bayesian inference (Knill & Richards, 1996)
- Evaluate consequences of actions
  - Does the brain calculate value ? (Vlaev et al., 2011)
- Select the action that reaches the goal
  - Decision tools : decision theory, signal detection theory.