En psychologie, on s'intéresse à la façon dont un individu est amené à sélectionner une action quand un choix se présente entre différentes actions possibles. Ce choix peut être influencé par un grand nombre de facteurs impondérables, ce qui fait qu'il est légitime de le modéliser à l'aide de variables aléatoires. L'objet du problème est de présenter quelques éléments simples de la théorie des modèles de choix discret. Dans le modèle binaire le plus simple, le choix se fait en fonction de la réaction à un stimulus. Dans une première partie, on étudie la modélisation élémentaire de la réponse à un stimulus. Dans une deuxième partie, on considère une importante modélisation de choix dépendant du hasard, dit *modèle de Luce*, et on étudie ses propriétés. Enfin, dans une troisième partie, on regarde le cas où les différents choix possibles engendrent des réactions aléatoires et on étudie des propriétés de la réaction optimale. **Les trois parties sont indépendantes**.

Toutes les variables aléatoires sont définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$. Si elles existent, on note $E(T)$ et $V(T)$ l'espérance et la variance d'une variable aléatoire T.

I Modèles avec réponse discrète

Soit α un réel (positif ou négatif) représentant un niveau de stimulus. On considère une variable aléatoire réelle X à valeurs dans $\mathbb{R}$ représentant la tolérance de l'individu au stimulus en question. On considère donc que l'individu réagit si $X \leq \alpha$ et ne réagit pas si $X > \alpha$. On considère la variable aléatoire Y indicatrice de la réaction définie par

$$Y = \begin{cases} 1 & \text{si} \quad X \leq \alpha \\ 0 & \text{si} \quad X > \alpha \end{cases} .$$

Soit F la fonction de répartition de la variable aléatoire X.

1) Déterminer la loi de Y, son espérance θ et sa variance.

2) On considère n individus dont on observe la réaction au stimulus. La tolérance de l'individu i est une variable aléatoire X_i dont on suppose qu'elle suit la même loi que X. En outre, les tolérances pour les différents individus sont supposées indépendantes.

(a) Soit N la variable aléatoire égale au nombre d'individus réagissant au stimulus. Déterminer la loi de N, son espérance et sa variance.

(b) Construire à l'aide de N un estimateur sans biais de θ.

3) Soient m un réel et σ un réel strictement positif.

On suppose que la tolérance X est obtenue comme résultante d'un “grand nombre” n de facteurs indépendants de petite taille c'est-à-dire $X = \sum_{i=1}^{n} X_i$, où les X_i sont supposées être des variables aléatoires de même loi d'espérance $\frac{m}{n}$ et de variance $\frac{\sigma^2}{n}$.

(a) Déterminer l'espérance et l'écart-type de X.

(b) Montrer que pour tout réel a,

$$\lim_{n\to+\infty} P\Big(\frac{X-m}{\sigma} \leq a\Big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{u^2}{2}} du.$$

(c) Le résultat précédent justifie que pour n grand on peut considérer que la variable aléatoire $\frac{X-m}{\sigma}$ suit la loi normale centrée réduite. Trouver dans ce cas l'expression de θ en fonction de α, m et σ.

(d) Déterminer $\lim\limits_{\sigma\to+\infty} \theta$ et interpréter le résultat.

4) Plutôt que d'utiliser la loi normale, on préfère souvent une loi plus simple dont on étudie dans cette question quelques propriétés.

(a) Soit F la fonction définie sur $\mathbb{R}$ par

$$\forall y \in \mathbb{R}, F(y) = \frac{1}{1+e^{-y}}.$$

Montrer que F est la fonction de répartition d'une variable aléatoire réelle. On dit alors que cette variable aléatoire suit la *loi logistique*.

(b) On suppose que X suit une loi logistique. Déterminer θ.

On considère Z une variable aléatoire suivant la loi logistique.

(c) Déterminer une densité de probabilité de Z.

(d) Soit y un réel positif. Etablir une relation entre $F(y)$ et $F(-y)$.

(e) Montrer que Z admet une espérance et la déterminer.

(f) Soit U une variable aléatoire de loi uniforme sur $]0,1[$. Déterminer la loi de la variable aléatoire $\ln\left(\frac{U}{1-U}\right)$.

II Règles de décisions stochastiques : le modèle de Luce

On suppose maintenant que l'individu doit choisir une action dans un ensemble fini d'actions possibles A. On note $\mathcal{F} = \{S \subset A / \mid S \mid \geq 2\}$ où $\mid S \mid$ désigne le cardinal de l'ensemble S. Quand le nombre d'actions possibles est très grand, la procédure de choix se passe en deux temps : l'individu commence par sélectionner une partie S de $\mathcal{F}$ à laquelle il va restreindre son choix, puis choisit une action précise à l'intérieur de S.

Pour chaque élément S de $\mathcal{F}$, on définit une probabilité P_S sur S : pour a un élément de S, $P_S(\{a\})$ représente la probabilité pour que l'individu ayant sélectionné S choisisse l'action a. Pour simplifier la notation, on notera

$P_S(a)$ pour $P_S(\{a\})$. En particulier, $P_A(S) = \sum_{a \in S} P_A(a)$ est la probabilité pour que l'individu prenne dans S l'action qu'il choisit.

Pour a et b distincts dans A on note $P(a,b) = P_{\{a,b\}}(\{a\})$; il s'agit donc de la probabilité de préférer l'action a à l'action b dans le cas d'un choix à faire entre a et b.

On suppose que pour tout S appartenant à $\mathcal{F}$ et tout a dans S, $P_S(a) \neq 0$.

On fait l'hypothèse suivante sur le modèle :

(*) Pour tout couple (S,T) d'éléments de $\mathcal{F}$ tel que S est inclus dans T, pour tout a élément de S,

$$P_T(a) = P_T(S)P_S(a).$$

5) Interpréter le sens de la condition (*) en termes de probabilités conditionnelles.

6)

(a) Soit k un réel strictement positif. On pose pour tout $a \in A$, $v(a) = kP_A(a)$. Montrer que pour tout S appartenant à $\mathcal{F}$ et pour tout a dans S,

$$P_S(a) = \frac{v(a)}{\sum_{b \in S} v(b)}. \tag{1}$$

(b) Montrer que si v et w sont deux fonctions réelles définies sur A satisfaisant (1), il existe un réel μ strictement positif tel que $v = \mu.w$. Une telle fonction v s'appelle une *utilité associée au système de probabilités* $(P_S)_{S \in \mathcal{F}}$.

7) Réciproquement, soit v une fonction réelle strictement positive sur A. On pose, pour tout S dans $\mathcal{F}$ et tout a appartenant à S,

$$Q_S(a) = \frac{v(a)}{\sum_{b \in S} v(b)}.$$

Montrer qu'on définit ainsi un système de probabilités vérifiant (*).

8)

(a) Soit v une utilité associée au système de probabilités $(P_S)_{S \in \mathcal{F}}$. Montrer que pour tout $S \in \mathcal{F}$, et pour tous a et b dans S,

$$v(a) \leq v(b) \Rightarrow P_S(a) \leq P_S(b).$$

La probabilité que a soit choisi augmente donc avec son utilité.

(b) Montrer qu'il existe une fonction ρ sur A telle que pour tous a et b distincts dans A,

$$P(a,b) = \frac{1}{1 + \exp(\rho(b) - \rho(a))}.$$

(c) Soit X une variable aléatoire suivant la loi logistique de fonction de répartition F définie par

$$F(y) = \frac{1}{1 + e^{-y}}.$$

Soient a et b distincts dans A. Trouver en fonction de ρ un réel $\alpha_{a,b}$ tel que $P(a,b) = P(X \leq \alpha_{a,b})$.

9)

(a) Montrer que pour tout couple (S,T) d'éléments de $\mathcal{F}$ tel que S est inclus dans T, et pour tous a et b dans S, on a

$$\frac{P_S(a)}{P_S(b)} = \frac{P_T(a)}{P_T(b)}.$$

Le rapport des probabilités de choix respectives de a et b est donc indépendant de la sélection de l'ensemble d'actions contenant a et b.

(b) On examine ici un cas concret. On suppose que l'individu devant se rendre de son domicile à son travail ait le choix entre utiliser sa voiture (symbolisée par V) ou le bus, dont deux lignes sont possibles : le bus rouge (symbolisé par R) ou le bus bleu (B). On a donc l'ensemble d'actions $A = \{V, R, B\}$. On suppose que l'individu est indifférent au fait de choisir sa voiture ou un bus, et est également indifférent à la couleur du bus. On définit ainsi un système de probabilités comme précédemment avec $P(V, R) = P(V, B) = \frac{1}{2}$ et de plus $P_A(R) = P_A(B)$. Montrer que $P_A(V) = \frac{1}{3}$. Ce résultat est-il satisfaisant ? Interpréter.

III Utilités aléatoires

Dans cette partie, on aborde la question du choix sous un autre aspect. A chaque action i de l'ensemble d'actions $A = \{1, 2, \ldots, n\}$ est associée une variable aléatoire U_i représentant l'utilité de l'action i. L'individu est alors amené à choisir l'action qui maximise ces utilités. On suppose que les variables U_i sont indépendantes et que la loi de U_i est donnée par la fonction de répartition F_i. On s'intéresse dans cette partie à la valeur U de l'utilité maximale, c'est à dire à $U = \max(U_1, \ldots, U_n)$.

10)

(a) Déterminer la fonction de répartition G_n de U. Que vaut G_n dans le cas particulier où les U_i suivent la même loi de fonction de répartition F ?

On suppose désormais que les U_i ont même loi.

(b) Pour x réel donné, étudier $\lim\limits_{n\to+\infty} G_n(x)$.

(c) Montrer que les seules lois pour lesquelles on a $G_n = F$ pour tout $n \geq 1$ sont les lois de variables aléatoires constantes.

11) Pour obtenir un type de loi plus intéressant pour U, on va chercher des lois admettant une densité strictement positive sur $\mathbb{R}$ et dont la fonction de répartition F vérifie que pour tout $n \geq 1$ il existe $b_n \leq 0$ tel que pour tout x réel, $(F(x))^n = F(x + b_n)$.

On suppose qu'une telle loi existe et on cherche des conditions qu'elle vérifie.

(a) Montrer que F est une fonction continue et strictement croissante telle que $\lim\limits_{x\to-\infty} F(x) = 0$ et $\lim\limits_{x\to+\infty} F(x) = 1$. F définit donc une bijection de $\mathbb{R}$ sur]0,1[.

(b) Montrer que la suite $(b_n)_{n\geq 1}$ est décroissante.

(c) Soit (n, N) un couple d'entiers strictement positifs. On considère $U_1, \ldots, U_{nN}$, nN variables aléatoires indépendantes de même loi F, et on pose pour j tel que $1 \leq j \leq n$,

$$Y_j = \max(U_{(j-1)N+1}, \ldots, U_{jN}).$$

Montrer que les variables Y_j sont indépendantes.

(d) Quelle est la fonction de répartition de Y_j ?

(e) En remarquant que $\max(Y_1, \ldots, Y_n) = \max(U_1, \ldots, U_{nN})$, montrer que pour tout x réel,

$$F(x)^{nN} = F(x + b_n + b_N) = F(x + b_{nN}).$$

Déduire que pour tout couple (n, N) d'entiers strictement positifs , $b_{nN} = b_n + b_N$.

(f) Montrer que pour tout entier n strictement positif et tout $k \in \mathbb{N}$, $b_{n^k} = kb_n$.

(g) Soient p et m deux entiers strictement positifs. Montrer qu'il existe un unique $k_m \in \mathbb{N}$ tel que $2^{k_m} \leq p^m < 2^{k_m+1}$ et que $\lim\limits_{m\to+\infty} \frac{k_m}{m} = \frac{\ln p}{\ln 2}$.

(h) En déduire qu'il existe un réel γ tel que pour tout entier p strictement positif, $b_p = \gamma \ln p$.

(i) Montrer que la fonction $F(x) = \exp(-e^{-x})$ satisfait aux conditions cherchées. La loi ainsi définie est dite loi de Gumbel.

12) Dans cette section, on étudie un certain nombre de propriétés de la loi de Gumbel. Soit X une variable aléatoire de loi de Gumbel c'est-à-dire de fonction de répartition F telle que $F(x) = \exp(-\mathrm{e}^{-x})$.

(a) Déterminer une densité de probabilité de X.

(b) On pose $Z = e^{-X}$. Déterminer la loi de la variable aléatoire Z.

(c) Soient x et y deux réels strictement positifs.
Etablir une relation entre $P_{[X \leq -\ln x]}(X \leq -\ln(x+y))$ et $P(X \leq -\ln y)$.

(d) On considère $(Y_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes de même loi exponentielle de paramètre 1. Soit L une variable aléatoire suivant une loi de Poisson de paramètre 1 indépendante de $(Y_i)_{i \geq 1}$. On considère la variable aléatoire $W = \max(Y_1, \ldots, Y_L)$ telle que pour tout $k \geq 1$, et tout $\omega \in [L = k]$, $W(\omega) = \max(Y_1(\omega), \ldots, Y_k(\omega))$ et $W(\omega) = 0$ si $L(\omega) = 0$.
Montrer que pour tous réels a et b tels que $0 < a < b$, on a $P(a \leq W \leq b) = P(a \leq X \leq b)$. Que vaut $P(W = 0)$?