

## MATHÉMATIQUES

## Devoir surveillé n°6

Durée : 4 heures

**L'usage d'une calculatrice est interdit pour cette épreuve. Si, au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il le signale sur sa copie et poursuit sa composition en expliquant les raisons des initiatives qu'il a été amené à prendre.**

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.

Toutes les variables aléatoires sont définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Si  $T$  est une variable aléatoire, on notera, sous réserve d'existence,  $E(T)$  l'espérance de  $T$  et  $V(T)$  sa variance.

Soit  $n$  un entier naturel non nul, on appellera  $n$ -échantillon de  $X$  toute suite  $(X_1, \dots, X_n)$  de variables aléatoires mutuellement indépendantes et suivant la même loi que  $X$ .

Si  $X$  est une variable aléatoire dont la loi dépend d'un paramètre  $a$ , et  $(X_1, \dots, X_n)$  un  $n$ -échantillon de  $X$ , une variable aléatoire  $T_n$  fonction de  $(X_1, \dots, X_n)$  est un estimateur sans biais et convergent de  $a$  si  $E(T_n) = a$  et  $\lim_{n \rightarrow +\infty} V(T_n) = 0$ . Le terme « estimateur » utilisé dans ce problème signifie par abus de langage que ces deux conditions sont vérifiées.

### Modèles avec réponse discrète

Soit  $\alpha$  un réel (positif ou négatif) représentant un niveau de stimulus. On considère une variable aléatoire réelle  $X$  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 :

$$Y = \begin{cases} 1 & \text{si } X \leq \alpha \\ 0 & \text{si } X > \alpha \end{cases} . \quad \text{En toute rigueur : } \forall \omega \in \Omega, Y(\omega) = \begin{cases} 1 & \text{si } X(\omega) \leq \alpha \\ 0 & \text{si } X(\omega) > \alpha \end{cases}$$

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

1. Déterminer la loi de  $Y$ , son espérance  $\theta$  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 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  $\theta$  et sa variance.
  - (b) Construire à l'aide de  $N$  un estimateur sans biais de  $\theta$ .
3. Soient  $m$  un réel et  $\sigma$  un réel strictement positif.
 

On suppose que la tolérance  $X$  est obtenue comme résultante d'un « grand nombre » de facteurs indépendants d'effet infinitésimal, autrement dit que la tolérance  $X$  est distribuée selon la Loi Normale  $\mathcal{N}(m, \sigma^2)$  d'espérance  $m$  et d'écart-type  $\sigma$ .

  - (a) Exprimer l'espérance  $\theta$  de la variable aléatoire  $Y$  en fonction de  $\alpha, m, \sigma$  à l'aide de la fonction de répartition  $\Phi$  de la Loi Normale  $\mathcal{N}(0, 1)$ .
  - (b) Préciser la limite à  $m$  fixé de cette espérance  $\theta$  lorsque  $\sigma$  tend vers  $+\infty$ .
4. Plutôt que d'utiliser la Loi Normale, on préfère souvent une Loi plus simple dont on étudie dans cette question les 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 la variable aléatoire  $X$  suit la *Loi Logistique*
- (b) On suppose que la tolérance  $X$  est distribuée selon cette Loi Logistique définie ci-dessus.  
 Déterminer alors l'espérance  $\theta$  de la variable aléatoire  $Y$ .
- (c) Soit  $Z$  une variable de Loi Logistique.
- Déterminer une densité  $f$  de  $Z$  qui soit continue sur  $\mathbb{R}$ .
  - Montrer que la courbe représentative de  $F$  admet pour centre de symétrie le point de coordonnées  $(0, \frac{1}{2})$ . Montrer que la courbe  $(C)$  représentative de  $f$  présente un axe de symétrie et deux points d'inflexion d'abscisse  $\pm \ln(2 + \sqrt{3})$ .
  - Montrer que  $Z$  admet une espérance nulle.
- (d) Soit  $U$  une variable aléatoire de Loi Uniforme sur  $]0, 1[$ .  
 Déterminer la loi de la variable aléatoire  $\ln\left(\frac{U}{1-U}\right)$ .

## Règles de décisions stochastiques : Le modèle de Luce

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

Pour chaque élément  $S$  de  $\mathcal{F}$ , on définit une probabilité  $P_S$  sur  $S$  :  $P_S(\{a\})$  représente la probabilité que l'individu ayant sélectionné  $S$  choisisse l'action  $a$ . Pour alléger l'écriture, on notera  $P_S(a)$  au lieu de  $P_S(\{a\})$ . En particulier,  $P_A(S) = \sum_{a \in S} P_A(a)$  est la probabilité que l'individu prenne dans  $S$  l'action qu'il choisit.

Pour  $a$  et  $b$  distincts de  $A$ , 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  $S$ ,  $P_S(a) \neq 0$ .

Dans le cas du modèle de Luce, le système de probabilités  $(P_S)_{S \in \mathcal{F}}$  satisfait à l'hypothèse (\*) si

$$\text{Pour tout couple } (S, T) \text{ éléments de } \mathcal{F} \text{ tels que } S \subset T : \forall a \in S, P_T(a) = P_T(S) \times P_S(a) \quad (*)$$

1. Interpréter le sens de cette condition (\*) en terme de probabilité conditionnelle.

2. Considérons un système de probabilités  $(P_S)_{S \in \mathcal{F}}$  vérifiant (\*).

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

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

- (b) Montrer que si  $w$  est une fonction réelle définie sur  $A$  satisfaisant (1), il existe un réel  $\mu$  strictement positif tel que  $w = \mu.v$ .

Une telle fonction  $v$  s'appelle une *utilité associée au système de probabilité*  $(P_S)_{S \in \mathcal{F}}$ .

3. Réciproquement, considérons une fonction réelle  $v$  strictement positive sur  $A$ .

On pose, pour tout  $S$  dans  $\mathcal{F}$  et tout  $a$  appartenant à  $A$

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

Montrer qu'on définit ainsi un système  $(Q_S)_{S \in \mathcal{F}}$  de probabilités vérifiant (\*).

4. Nous supposons que le système de probabilités  $(P_S)_{S \in \mathcal{F}}$  satisfait à l'hypothèse (\*)

- (a) Soit  $v$  une utilité associée à ce système de probabilités.

Montrer que pour tout  $S \in \mathcal{F}$  et pour tous  $a$  et  $b$  dans  $S$

$$v(a) \leq v(b) \implies P_S(a) \leq P_S(b)$$

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

- (b) Montrer qu'il existe une fonction  $\rho$  définie 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 distribuée selon la Loi logistique dont la fonction de répartition  $F$  est définie par :

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

Soient  $a$  et  $b$  deux actions distinctes de  $A$ .

Trouver en fonction de  $\rho$  un réel  $\alpha_{a,b}$  tel que :  $P(a, b) = P(X \leq \alpha_{a,b})$

5. Nous supposons que le système de probabilités  $(P_S)_{S \in \mathcal{F}}$  satisfait à l'hypothèse (\*)

- (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  $b$  distincts 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 le modèle de Luce dans un cas concret. on suppose que l'individu devant se rendre de son domicile à son travail a le choix entre utiliser sa voiture (symbolisée par V) et prendre le bus, dont deux lignes sont possibles : le bus rouge (symbolisé par R) ou le bus bleu (B). On définit donc l'ensemble des actions  $A = \{V, R, B\}$ . On suppose que l'individu n'a pas de préférence entre 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 :

$$P(V, R) = P(V, B) = \frac{1}{2} \quad \text{et} \quad P_A(R) = P_A(B)$$

Montrer que  $P_A(V) = \frac{1}{3}$ . Ce résultat est-il satisfaisant ? Interprétez !

## Utilités aléatoires

Dans cette partie, on aborde la question du choix sous un autre aspect. À chaque action  $i$  de l'ensemble d'actions  $A = \{1, 2, \dots, n\}$  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  $U = \max(U_1, \dots, U_n)$ . La fonction de répartition de cette variable aléatoire  $U$  est alors notée  $G_n$ .

1. Que vaut  $G_n$  dans le cas particulier où les  $U_i$  suivent la même loi de fonction de répartition  $F$  ?
2. On suppose désormais que les  $U_i$  ont même loi.
  - (a) Pour  $x$  un réel donné, étudier  $\lim_{n \rightarrow +\infty} G_n(x)$ .
  - (b) Montrer que les seules lois pour lesquelles on a  $G_n = F$  pour tout  $n \geq 1$  sont les lois de variables aléatoires presque sûrement constantes.
3. 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 :

Pour tout entier naturel non nul, il existe un réel  $b_n \leq 0$  tel que :  $\forall x \in \mathbb{R}, (F(x))^n = F(x + b_n)$

On suppose qu'une telle loi existe et on cherche les propriétés qu'elle vérifie.

- (a) Montrer que  $F$  est une fonction continue et strictement croissante telle que :

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

$F$  définit donc une bijection de  $\mathbb{R}$  sur  $]0, 1[$ .

- (b) Montrer que la suite  $(b_n)_{n \in \mathbb{N}}$  est décroissante.
- (c) Soit  $(n, N)$  un couple d'entiers strictement positifs. On considère  $U_1, \dots, U_{nN}$ ,  $nN$  variables aléatoires indépendantes de même loi  $F$ , et on pose pour  $j \in \llbracket 1, n \rrbracket$  :  $Y_j = \max(U_{(j-1)N+1}, \dots, U_{jN})$ .

Montrer que les variables  $Y_j$  sont indépendantes.

- (d) Quelle est la fonction de répartition de  $Y_j$  ?
- (e) En remarquant que  $\max(Y_1, \dots, Y_n) = \max(U_1, \dots, U_{nN})$ , montrer que :

$$\forall x \in \mathbb{R}, F(x)^{nN} = F(x + b_n + b_N) = F(x + b_{nN})$$

En 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  $k \in \mathbb{N}$ ,  $b_{nk} = kb_n$ .
- (g) Soient  $p$  et  $m$  deux entiers strictement positifs.  
Montrer qu'il existe un unique  $k_m \in \mathbb{N}$  tel que :  $2^{k_m} \leq p^m < 2^{k_m+1}$ . Montrer que  $\lim_{m \rightarrow +\infty} \frac{k_m}{m} = \frac{\ln p}{\ln 2}$ .
- (h) En déduire qu'il existe un réel  $\gamma$  tel que pour tout entier  $p$  strictement positif,  $b_p = \gamma \ln p$ .
- (i) Montrer que la fonction  $F : x \mapsto \exp(-e^{-x})$  satisfait aux conditions cherchées.  
La loi ainsi définie est dite *Loi de Gumbel*.

4. 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 dont la fonction de répartition  $F$  est  $x \mapsto \exp(-e^{-x})$

- (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.  
Établir une relation entre  $P_{\lfloor X \leq -\ln x \rfloor}(X \leq -\ln(x+y))$  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 des  $Y_i$ . On considère la variable aléatoire  $W = \max(Y_1, \dots, Y_L)$  telle que pour tout  $k \geq 1$  et tout  $\omega \in [L = k]$ ,  $W(\omega) = \max(Y_1(\omega), \dots, Y_L(\omega))$  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  $P(W = 0)$  ?