T.D. de Probabilités et Statistiques

Licence Informatique

Gilles HUNAULT, 1998

T.D. 1 : Probabilités élémentaires

1. Rappels de la définition d'une probabilité
2. Calcul de $p(A \cap \overline{B}) \cup (B \cap \overline{A})$ en fonction de $\alpha = p(A \cap B)$.
3. Détermination de A, B, p_1, p_2 tels que A et B sont indépendants pour p_1 mais pas pour p_2.
4. Etude de l'obtention d'un as lorsqu'on tire n cartes.
5. Etude des $p_i = C_n^i 0.2^i 0.8^{n-i}$ pour i de 1 à n.
6. Comptage des fichiers en $ghFs$.
7. Comparaison des chances de gain au tiercé avec n chevaux et au quarté avec $n+1$ chevaux.
8. Formule de *Bayes* pour les variables typées en $ghProlog$.

Exercice 1

Une probabilité sur E est une **fonction** définie sur $\mathcal{P}(E)$, à valeurs dans $[0,1]$ et **additive** pour des ensembles disjoints. Si (E_i) est un système complet d'évènements de E, donner chacun des $p(E_i)$ suffit à définir p. On doit alors avoir $p(E_i) > 0$ et $\sum p(E_i) = 1$.

Par exemple, pour i de 1 à n, la valeur $p_i = p(E_i) = i/(\ n(n+1)(n+3)\ )$ ne définit pas une probabilité car pour $n = 2$, $p_1 + p_2 = 3/10 < 1$.

Exercice 2

L'expression $p(A \cap \overline{B}) \cup (B \cap \overline{A})$ n'a aucun sens, donc il n'y a aucun calcul à faire ! Plus raisonnablement, essayons de calculer $p(\ (A \cap \overline{B}) \cup (B \cap \overline{A})\ )$.

A partir de $A = A \cap E = A \cap (B \sqcup \overline{B}) = (A \cap B) \sqcup (A \cap \overline{B})$ on déduit $p(A \cap \overline{B}) = p(A) - p(A \cap B)$ et de façon symétrique $p(B \cap \overline{A}) = p(B) - p(B \cap A)$ donc $p(\ (A \cap \overline{B}) \cup (B \cap \overline{A})\ ) = p(A \cap \overline{B}) + p(B \cap \overline{A}) = p(A) + p(B) - 2p(A \cap B)$ car $A \cap \overline{B}$ et $B \cap \overline{A}$ sont disjoints.

Avec des valeurs numériques, la probabilité α cherchée doit donc vérifier $p = p(A) + p(B) - 2p(\alpha)$ ce qui impose des conditions sur α pour avoir $0 \leq p \leq 1$. Ainsi, pour $p(A) = 0.2$, $p(B) = 0.5$, on en déduit, toutes inégalités confondues, que $0 \leq \alpha \leq 0.2$.

Exercice 3

Prenons le jet d'un dé avec $A =$"pair" et $B = \{5,6\}$. Puisque $A \cap B = \{6\}$, A et B ne sont pas incompatibles. Soit p_1 la probabilité équirépartie, alors $p_1(A \cap B) = 1/6$ et comme $p_1(A) = 1/2$, $p_1(B) = 1/3$, $p_1(A \cap B) = p_1(A).p_1(B) = 1/3$ donc A et B sont indépendants pour p_1.

Définissons maintenant p_2 par $p_2(6) = 1/12$, $p_2(2) = p_2(4) = p_2(5) = 1/6$. Alors $p_2(A \cap B) = 1/12$ et $p_2(A) = 5/12$, $p_2(B) = 3/12$, $p_2(A \cap B) \neq p_2(A).p_1(B) = 1/3$ donc A et B ne sont pas indépendants pour p_2.

Exercice 4

Traitons deux cas usuels avant de voir les formules générales. A la belote, $n = 32$, au bridge, $n = 52$. On pourra remarquer qu'il y a 4 "couleurs" (comme 4 saisons) et "comme par hasard" il y a 52 semaines dans une année (365=7x52+1).

Une structuration possible des cartes consiste à les noter H_i^j avec i une "couleur" de 1 à 4 et j une valeur de 7 à 13 (ou de 1 à 13), 13 étant l'as, 12 le roi etc. On peut aussi noter (i, j) une telle carte. Un as correspond donc aux couples $(i, 13)$ et il y en 4 soit la probabilité $4/n$ dans le cas général (respectivement $1/8 \simeq 0.125$ soit 13 % et $1/13 \simeq 0.0769$ soit 8%). Si on pose la question "quelle est la probabilité d'obtenir au moins 1 as si on tire deux cartes", il y a ambiguité puisqu'il peut s'agir de tirer deux cartes simultanément (c'est à dire **sans** remise) ou il peut s'agir de tirer une carte puis de la remettre et de retirer une carte qui peut être éventuellement la même (tirage **avec** remise).

Exercice 5

Si on connait la formule du binome $(a+b)^n = \sum_{i=0}^{n} C_n^i a^i b^{n-i}$, les p_i correspondent aux termes pour i de 1 à n seulement. Ce sont donc bien des nombres positifs inférieurs à 1 mais leur somme fait seulement $1 - 0.8^n$ donc ces nombres ne définissent pas une probabilité.

Exercice 6

Enoncé

Un enseignant à forte tendance pédagogique veut imposer un nouveau système de fichiers nommé *ghFs*. Dans un tel système, un identificateur de fichier se compose d'un nom et d'une extension reliés par un tiret. Un nom de fichier comporte de 1 à 5 caractères dont le premier est une lettre, les caractères suivants sont soit une lettre soit un chiffre. Une extension comporte de 1 à 3 caractères, le premier est une lettre, les suivants une lettre ou un chiffre.

De façon à ne pas avoir des noms trop imprononçables, on n'utilise que 20 lettres (mais on ne dit pas lesquelles, sauf la lettre P).

a) combien de noms de fichiers différents peut-on avoir, sachant qu'on ne distingue pas majuscule et minuscule ?

b) un fichier-programme est un fichier dont l'identificateur a une extension qui commence par la lettre P. Quelle est la proportion de fichiers-programmes dans ce système de fichiers ?

SOLUTION

Si L désigne une lettre et K une lettre ou un chiffre, un nom de fichier est représenté par L, LK, LKK, LKKK ou par LKKKK. L a 20 possibilités et K a 30 possibilités (les 20 de L plus 10 chiffres). De façon plus rigoureuse, pour L on veut un singleton dans un ensemble à 20 éléments et il y a donc $\mathrm{C}_{20}^{1} = 20$ possibilités. Chaque L et chaque K sont indépendants, donc on a $20 + 20.30 + 20.30^2 + 20.30^3 + 20.30^4$ noms de fichiers. Numériquement, on trouve donc $20.(30^5 - 1)/(30 - 1) = 16758620$ noms, soit à peu près 16,75 millions de noms différents.

Le nombre d'identificateurs (ce qui n'est pas demandé) est obtenu en ajoutant L, LK ou LKK à un nom de fichiers. Si N est le nombre de noms, le nombre d'identificateurs est donc $I = N.(20+20.30+20.30^2) = N.18620$. Numériquement $I = 312045504400$ (en gros 312 milliards d'identificateurs).

Un fichier programme a une extension qui commence par la lettre P. Comme il y a 20 lettres différentes, la proportion de fichiers-programmes est donc 1/20 si on suppose que seuls les fichiers programmes commencent par cette lettre. Par exemple, et cela ne contredit pas l'énoncé, un fichier de paramètres pourrait aussi commencer par P.

Exercice 7

Enoncé

On dispose de n boules numérotées de 1 à n, nommées $b(i)$, i variant de 1 à n. On dispose aussi de N boules de couleurs nommées $c(j,k)$. Il y a p couleurs différentes et $q(j)$ boules de couleur j, j variant de 1 à p. $c(j,k)$ est la $k$ième boule de couleur j, k variant de 1 à $q(j)$.

a) Au jeu de loto, sous controle d'un huissier, 6 numéros parmi 49 numéros de 1 à 49 sont tirés. Un joueur de loto est une personne qui achète un billet sur lequel 6 numéros sont inscrits (on ignore ici la notion de "numéro complémentaire"). On dit qu'on a gagné le gros lot si les numéros du billet acheté correspondent aux numéros tirés. Quel type de boules peut-on utiliser pour structurer les événements associés aux tirages du loto ? Quelle est la probabilité de gagner le gros lot ? Vous n'oublierez pas de détailler la structuration de votre espace probabilisé.

b) Au tiercé, une course met en présence n chevaux (classiquement une à deux dizaines, mais parfois moins). Chaque cheval comporte un numéro de 1 à n (on omet ici la notion de "non-partant"). A l'arrivée, on ne tient compte que des numéros des trois premiers chevaux. On dit qu'on a gagné "dans l'ordre" si on parié (et si on a acheté le ticket correspondant) sur ces trois numéros exactement dans l'ordre d'arrivée. On dit qu'on a gagné "dans le désordre" si on a parié sur ces trois numéros mais dans un ordre quelconque (et sauf l'ordre exact).

   Quel type de boules peut-on utiliser pour structurer les événements associés aux résultats du tiercé ? Quelle est la probabilité de gagner dans l'ordre et dans le désordre pour n=17 chevaux ? Reprendre avec n=17 chevaux au quarté (on s'intéresse alors aux quatre premiers chevaux).

On ne demande pour les deux questions suivantes aucun calcul numérique explicite (même s'ils sont effectués "au brouillon").

c) La société "La Française des Jeux" envisage de passer de 49 numéros à 60 numéros avec un tirage de 8 boules plutot que 6. Aura-t-on plus de chances de gagner le gros lot ? Comment les valeurs 6 et 49 ont-elles été choisies historiquement ?

d) Vaut-il mieux tenter de gagner (dans l'ordre, dans le désordre), au tiercé avec n chevaux ou au quarté avec $n+1$ chevaux ?

Solution

Le loto utilise 49 boules numérotées de 1 à 49. On utilisera donc $b(49)$. La probabilité de gagner au loto est celle de tirer le bon 6-uplet parmi tous les 6-uplets possibles (ce sont des 6-uplets car on tire 6 numéros sans ordre et sans répétition). A défaut d'autre hypothèse et ce qu'assure (sans démonstration) la Française des Jeux, chaque 6-uplet est équiprobable. On vient donc dénombrer des sous-ensembles de 6 éléments dans un ensemble de base à 49 éléments. La probabilité demandée est donc $1/C_{49}^{6}$ ce qui fait 1/13983816 soit en en gros $1/1,39.10^7$ donc $0,715.10^-7$.

Pour le tiercé dans le désordre, on peut utiliser la même structuration mais avec des triplets (3-uplets) où les $b(n)$ boules correspondent aux chevaux. Là encore, il faut faire abstraction de la réalité (météo, jockeys, etc.) pour attribuer une équirépartition de chances. La probabilité de gagner dans le désordre n'est pas $1/C_n^3$ soit 6 / 4080 car parmi les 6 triplets possibles il y a le "bon ordre" mais 5 / 4080. Numériquement on obtient 1/816 soit 0,001225.

Pour gagner dans l'ordre, il faut trouver le bon cheval arrivé en premier (probabilité 1/n) puis le second (1/(n-1)) et le troisième (1/(n-2)). On trouve alors $1/(n.(n-1).(n-2))$. Numériquement cela donne 1/4080 = 0,000245.

Si on joue au quarté, on obtient comme probabilités 1/57120 = 0.000175 de gagner dans l'ordre et 1/2380 = 0.0004201 de gagner dans le désordre.

Jouer avec 8 boules et 60 numéros diminuerait les chances de gagner car les coefficients du binome croissent très rapidement. De même, rajouter un cheval ne fait que diminuer la probabilité de gagner car on multiplie par 1/(n-3).

Numériquement, le rapport des chances est de 1/183 au loto si on passe de 49 et 6 à 60 et 8, pour le tiercé ce rapport est de 1/18 dans l'ordre et de 1/4.5 dans le désordre si on passe de 17 à 18 chevaux.

Quant à l'origine des nombres 6 et 49, je n'en ai aucune idée !

Exercice 8

Enoncé

En *ghProlog* (langage qui n'existe pas encore), une variable est soit libre, soit liée, soit indécise. On dit alors qu'elle est typée. Dans un programme de ce langage, la probabilité qu'une variable soit typée comme variable libre est de 0.45, la probabilité qu'une variable soit typée comme variable liée est de 0.45, la probabilité qu'une variable soit typée comme variable indécise est de 0.1. La probabilité qu'une variable soit correctement typée (par le programmeur) sachant qu'il s'agit d'une variable libre est de 0.90. La probabilité qu'une variable soit correctement typée (par le programmeur) sachant qu'il s'agit d'une variable liée est de 0.90. La probabilité qu'une variable soit correctement typée (par le programmeur) sachant qu'il s'agit d'une variable indécise est de 0.10. Quelle est la probabilité qu'une variable soit liée sachant qu'elle est correctement typée ?

Solution

Soit X l'évènement correspondant au typage d'une variable, soient A, B et C les évènements associés respectivement au typage libre, lié et indécis. Comme il n'y a pas d'autre typage possible : $X = A \cup B \cup C$ et cette union est disjointe. On sait par l'énoncé que $p(A) = 0{,}45$, $p(B) = 0{,}45$ et $p(C) = 0{,}10$. Soit maintenant Y l'évènement correspondant à un typage correct. L'énoncé dit aussi que $p(Y|A) = 0,9$, $p(Y|B) = 0,9$ et $p(Y|C) = 0,1$. La question posée est celle du calcul de $p(B|Y)$.

Décomposons Y en l'union disjointe $(Y \cap A) \cup (Y \cap B) \cup (Y \cap C)$. Comme $p(Y \cap A) = p(Y|A).p(A)$, $p(Y \cap B) = p(Y|B).p(B)$ et $p(Y \cap C) = p(Y|C).p(C)$, on en déduit $p(Y \cap B) = 0{,}9\text{x}0{,}45 = 0{,}405$. De même la valeur de $p(Y)$ est $0{,}9\text{x}0{,}45 + 0{,}9\text{x}0{,}45 + 0{,}1\text{x}0{,}1$ donc $p(Y) = 0{,}82$.

Finalement, $p(B|Y) = 0{,}405/0{,}820$ soit $p(B|Y) = 0{,}494$.