

Probabilités et tableaux croisés

Leçon

1. Modèle probabiliste (rappels et compléments)

Définition

- L'**univers** Ω est l'ensemble des issues d'une expérience aléatoire.
- Une **loi de probabilité** associe à chaque issue un nombre entre 0 et 1, la somme de ces nombres étant 1.
- Un **événement** A est une partie de Ω . Sa probabilité $P(A)$ est la somme des probabilités des issues qui le composent.

Remarque. Le choix d'une loi de probabilité est une **hypothèse de modélisation** : elle ne se démontre pas. « Dé équilibré », « tirage au hasard » signifient que l'on choisit l'**équiprobabilité**.

Propriété

- En situation d'équiprobabilité : $P(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}$.
- Événement contraire : $P(\overline{A}) = 1 - P(A)$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Exemple. Dé équilibré. A : « pair » = $\{2; 4; 6\}$, B : « au moins 5 » = $\{5; 6\}$, $A \cap B = \{6\}$. $P(A \cup B) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$.

Propriété – Loi des grands nombres (version vulgarisée)

Lorsque n est grand, sauf exception, la fréquence observée d'un événement sur n répétitions indépendantes d'une expérience est proche de sa probabilité.

Exemple de simulation en Python.

```
1 from random import randint
2
3 def frequence_six(n):
4     compteur = 0
5     for i in range(n):
6         if randint(1, 6) == 6:
7             compteur = compteur + 1
8     return compteur / n
```

Pour $n = 100\,000$, on obtient une valeur proche de $\frac{1}{6} \approx 0,167$.

2. Tableaux croisés et fréquences conditionnelles

Définition

Un **tableau croisé d'effectifs** répartit une population selon deux caractères qualitatifs.

- La **fréquence marginale** d'une modalité est son effectif total divisé par l'effectif de la population.
- La **fréquence conditionnelle** de B parmi A est $\frac{\text{effectif de } A \cap B}{\text{effectif de } A}$.

Exemple.

	Sport	Pas sport	Total
Filles	48	52	100
Garçons	60	40	100
Total	108	92	200

- Fréquence marginale des sportifs : $\frac{108}{200} = 0,54$.
- Parmi les filles, fréquence des sportives : $\frac{48}{100} = 0,48$.
- Parmi les sportifs, fréquence des filles : $\frac{48}{108} \approx 0,44$.

Remarque. « 48 % des filles font du sport » et « 44 % des sportifs sont des filles » ne disent pas la même chose : la population de référence change.

3. Probabilités conditionnelles

Définition

Soient A et B deux événements avec $P(A) \neq 0$. La **probabilité de B sachant A** est

$$P_A(B) = \frac{P(A \cap B)}{P(A)}.$$

Propriété

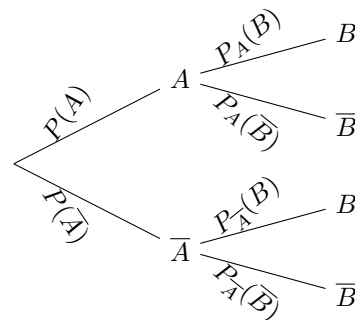
- Lorsqu'on tire au hasard (équiprobabilité) un individu d'une population : $P_A(B) = \frac{\text{Card}(A \cap B)}{\text{Card}(A)}$, qui est la fréquence conditionnelle lue dans le tableau croisé.
- $P(A \cap B) = P(A) \times P_A(B)$.

Exemple. On tire un élève au hasard dans le tableau précédent. F : « c'est une fille », S : « il fait du sport ». $P_F(S) = \frac{48}{100} = 0,48$ et $P_S(F) = \frac{48}{108} \approx 0,44$. En général, $P_A(B) \neq P_B(A)$.

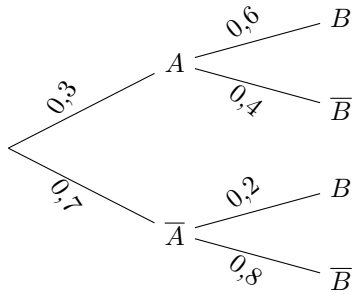
4. Arbres pondérés

Propriété – Règles de construction

- Sur les branches du premier niveau : $P(A)$ et $P(\bar{A})$.
- Sur les branches du second niveau : des probabilités **conditionnelles** ($P_A(B)$, ...).
- La somme des probabilités des branches issues d'un même nœud vaut 1.
- La probabilité d'un chemin est le **produit** des probabilités de ses branches : $P(A \cap B) = P(A) \times P_A(B)$.



Exemple. 30 % des clients d'un magasin ont la carte de fidélité (A). Parmi eux, 60 % achètent un produit en promotion (B), contre 20 % des autres clients.



$P(A \cap B) = 0,3 \times 0,6 = 0,18$: la probabilité qu'un client ait la carte *et* achète en promotion est 0,18.

$P(\bar{A} \cap B) = 0,7 \times 0,2 = 0,14$.

5. Tests de dépistage : faux positifs

Définition

On teste une maladie M ; T désigne l'événement « le test est positif ».

— La **sensibilité** du test est $P_M(T)$; la **spécificité** est $P_{\bar{M}}(\bar{T})$.

— Un **faux positif** est une personne non malade dont le test est positif ; un **faux négatif** une personne malade dont le test est négatif.

Exemple. Une maladie touche 1 % de la population ; sensibilité 95 % ; spécificité 98 %. Sur une population de 10 000 personnes :

	Test positif T	Test négatif $\bar{T}$	Total
Malades M	95	5	100
Non malades $\bar{M}$	198	9 702	9 900
Total	293	9 707	10 000

$P_M(T) = 0,95$, mais $P_T(M) = \frac{95}{293} \approx 0,32$: une personne testée positive n'a qu'environ 32 % de chances d'être malade !

Remarque. Il est essentiel de distinguer $P_M(T)$ (« le test détecte les malades ») et $P_T(M)$ (« un test positif signifie qu'on est malade ») : confondre les deux mène à de fausses affirmations.