GM – Probabilités et Statistique

http://moodle.epfl.ch/course/view.php?id=18431

Cours 3

- Révision probabilités (théorie fréquenciste, axiomes, éléments élémentaires équiprobables)
- Probabilité conditionnelle
- Formule de Bayes
- Indépendance
- Variables aléatoires : loi de probabilité, fonction de répartition
- VA de Bernoulli, binomiale

Révision : La probabilité

- Théorie fréquenciste des probabilités : $P(E) = \lim_{n\to\infty} \frac{n(E)}{n}$

Axiomes de probabilité

1. $0 \le P(E) \le 1$
2. $P(S) = 1$
3. Pour chaque séquence d'événements mutuellement exclusifs $E_1, E_2, \ldots,$ $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$.

- Événements élémentaires équiprobables : il y a un nombre fini d'éléments élémentaires de l'ensemble fondamental à la *même probabilité d'apparaître* : $P(E) = \dfrac{\text{nombre de points dans } E}{\text{nombre de points dans } S}$

Exemple 2.9 (encore une fois)

- Si deux dés sont lancés (et en supposant que les 36 issues possibles sont *équiprobables*, quelle est la probabilité que la somme des faces soit 8 ? ?

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Exemple 2.9, cont

- **Maintenant je donne l'information : le premier dé est** 
  $\Rightarrow$ *Sachant* cette information, P(somme $= 8$) $=$ *? ?*
- Il y a ____ issues dans ce nouvel ensemble fondamental
- **PARMI ces issues**, combien corréspond à l'événement : $\{$somme $= 8\}$ *? ?*

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	**(3,5)**	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Probabilité conditionnelle

- Autre façon de le dire : **étant donné** que le premier dé est ⚂ la probabilité *(conditionnelle)* que la somme soit 8 est *1/6*
- Soit $E =$ 'la somme vaut 8' ; $F =$ '1er est 3'
- On a $P(E \mid F) = 1/6$: '|' = '*étant donné*' ou '*sachant*'
- *Formellement*, si $P(F) > 0$, on a : $P(E \mid F) = \dfrac{P(E \text{ et } F)}{P(F)}$
- $P(E \mid F)$ **n'est pas définie** si $P(F) = 0$ (car la division par 0 est **illégale**)
- La notion de *probabilité conditionnelle* est importante pour deux raisons principales :
  - calcul des probabilités quand des *informations partielles* sur le résultat d'une expérience sont disponibles
  - utile comme *un outil* pour calculer plus facilement une certaine probabilité désirée

$P(E \text{ et } F)$

- On considère la *définition de la probabilité conditionnelle* :

  1. $P(E \mid F) = \dfrac{P(E \text{ et } F)}{P(F)}$

- On a également :

  2. $P(F \mid E) = \dfrac{P(E \text{ et } F)}{P(E)}$

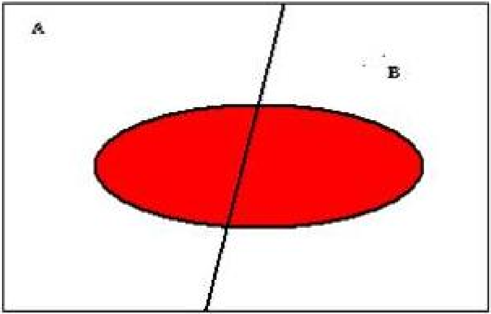
- Donc on a **deux façons** d'exprimer $P(E \text{ et } F)$:

  1. $P(E \text{ et } F) = P(F)P(E \mid F)$
  2. $P(E \text{ et } F) = P(E)P(F \mid E)$

- Ceci pourrait être généralisé :
  - $P(E_1 \text{ et } E_2 \text{ et } \ldots \text{ et } E_n)$
    $$= P(E_1) \times P(E_2 \mid E_1) \times P(E_3 \mid E_1, E_2) \times \cdots \times P(E_n \mid E_1, E_2, \ldots, E_{n-1})$$
  - *Pourquoi ? ?*

Partition

- Une **partition** (cloison) divise l'ensemble fondamental en des *sous-ensembles disjoints* :
  - sans trou (en anglais 'no gaps')
  - sans superflue (en anglais 'no overlaps')

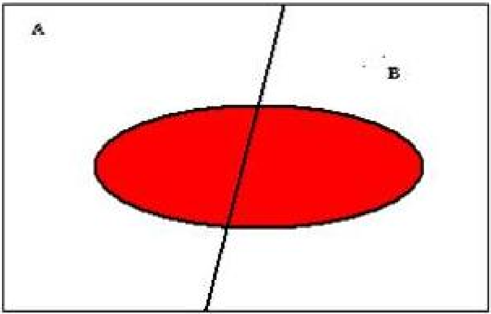


Formule des probabilités totales

- Soit A, B une partition de l'ensemble fondamental et soit R un événement ; alors,

$$\begin{aligned} P(R) &= P(R \text{ et } A) + P(R \text{ et } B) \\ &= P(R \mid A)P(A) + P(R \mid B)P(B) \end{aligned}$$

- La partition pourrait être composée de plus que deux événements / sous-ensembles



Exemple

Exemple 3.1 On admet que 5% des hommes et 0,25% des femmes sont daltonien/nes. On sélectionne *une personne* au hasard. Quelle est la probabilité qu'il s'agisse d'*un/e daltonien/ne* en supposant :

(a) que les hommes sont *aussi nombreux* que les femmes *? ?*

(b) qu'il y a *deux fois plus des hommes que des femmes* dans la population *? ?* **[exercice pour vous]**

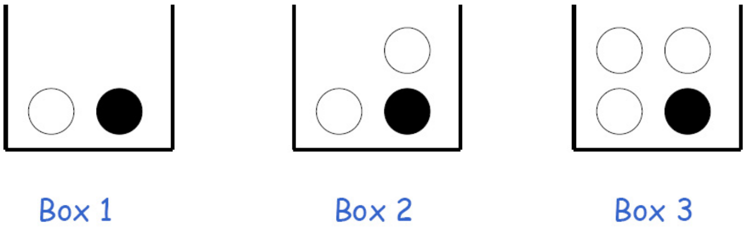
- *Astuce :* **expliciter** toutes les informations données

Quelle boîte ?

- Il y a 3 boîtes semblables : 1, 2, 3
- La boîte i contient i boules blanches et 1 boule noire
- Je choisis une boîte au hasard, puis je choisis une boule au hasard dans cette boîte (et je vous montre la boule) – elle est blanche
- **Deviner :** j'ai choisi quelle boîte *? ?* [pourquoi ?]

- **Quelle est la probabilité** que vous ayez eu bien deviné *? ?*

Quelle boîte ?



Formule de Bayes

- Pour une *partition* $F_1, F_2, \ldots, F_n$ de l'espace fondamental S :

$$\underline{\boldsymbol{P(F_j \mid E)}} = \frac{P(E \text{ et } F_j)}{P(E)}$$

$$= \frac{P(E \mid F_j)P(F_j)}{\sum_{i=1}^{n} P(E \mid F_i)P(F_i)}$$

$$= \boxed{\frac{P(E \mid F_j)P(F_j)}{P(E \mid F_1)P(F_1) + P(E \mid F_2)P(F_2) + \cdots + P(E \mid F_n)P(F_n)}}$$

- La formule de Bayes utilise les deux expressions pour $P(E \text{ et } F)$

Exemple 3.1 (cont.)

Supposons que 5% des hommes et 0,25% des femmes sont daltonien/nes. Une *personne daltonienne* est choisi au hasard. Quelle est la probabilité que la personne soit *de sexe masculin* en supposant :

(a) les hommes sont *aussi nombreux* que les femmes ? ?

(b) qu'il y a *deux fois plus des hommes que des femmes* dans la population ? ? **[exercice pour vous]**

Indépendance

- En générale, $P(E \mid F) \neq P(E)$; c.-à-d., le fait de savoir que F est survenu *influe sur* **la probabilité de E**
- Dans le **cas particulier** où $P(E \mid F) = P(E)$, on dit que les événements E et F sont **indépendants**
- Les événements sont indépendants lorsque le fait de savoir que l'un est survenu *ne modifie pas la probabilité* que l'autre se produit
- Pour les événements *indépendants* (mais **PAS** en général !),

$$P(E \text{ et } F) = P(E) \times P(F)$$

- Deux événements sont *dépendants* s'ils ne sont pas indépendants

Exemple 3.2 On lance 2 dès (bleu et rouge), et on suppose que les 36 événements élémentaires sont *équiprobables*. Soit E l'événement que la somme égale 8, et F l'événement que le bleu montre 3.

(a) E et F, sont-ils indépendants *? ?*

(b) Même question si E dénote l'événement que la somme égale 7 *? ?*

L'ensemble fondamental conditionnel ; $\sum = 8$

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	**(3,5)**	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

L'ensemble fondamental conditionnel ; $\sum = 7$

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	**(3,4)**	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Indépendance totale de trois événements

- Trois événements E, F et G sont dits **totalement indépendants** si :

$$\begin{aligned} P(E \text{ et } F \text{ et } G) &= P(E)P(F)P(G) \\ P(E \text{ et } F) &= P(E)P(F) \\ P(E \text{ et } G) &= P(E)P(G) \\ P(F \text{ et } G) &= P(F)P(G) \end{aligned}$$

- Indépendance totale de n événements : un ensemble d'événements $E_1, E_2, \ldots, E_n$ est dit **totalement indépendant** si pour tout sous-ensemble $E_{1'}, E_{2'}, \ldots, E_{r'}$, $r \leq n$,

$$P(E_{1'} \text{ et } E_{2'} \text{ et } \cdots \text{ et } E_{r'}) = P(E_{1'})P(E_{2'})\cdots P(E_{r'})$$

PAUSE

Epreuves (aléatoires)

- Il arrive parfois que l'expérience étudiée consiste à effectuer une suite d'*expériences partielles* – p. ex. plusieurs lancements d'une pièce
- Il est peut-être raisonnable d'admettre que l'issue de tout groupe d'expériences partielles sont *totalement indépendantes* – encore, lancements d'une pièce
- Si toutes ces expériences partielles sont *identiques* (le même ensemble fondamental, la même fonction de probabilité), elles sont appelées **épreuves**

Exemple : épreuves indépendantes

Exemple 3.3 On réalise une séquence infinie d'épreuves indépendantes. Chaque épreuve donne soit un succès, soit un échec avec probabilité p et $1 - p$ respectivement. Quelle est la probabilité qu'il y ait :

(a) *au moins 1* succès parmi les n premières épreuves *? ?*

(b) *exactement k* succès parmi les n premières épreuves *? ?*

Variables aléatoires

- Il arrive souvent qu'on s'intéresse plus à *une fonction* du résultat d'une expérience qu'au résultat lui-même
- Par exemple, en lançant 2 dès on pourrait s'intéresser à la *somme* des nombres au lieu des nombres eux-mêmes
- Une **variable aléatoire** (VA) est une fonction réelle définie sur l'ensemble fondamental
- *Variable aléatoires* : écrite en MAJUSCULES (X, Y, *etc.*)
- *Valeurs spécifiques* : écrite en miniscules (x, y, *etc.*)

Exemple

Exemple 3.4 Une expérience consiste à lancer (indépendamment) 3 pièces *équitables*. Soit Y le nombre de lancements donnant pile. Alors Y est une variable aléatoire et peut prendre les valeurs 0, 1, 2, 3 avec pour probabilité, respectivement :

$$
\begin{aligned}
P(Y=0) &= P(F,F,F) = \underline{\hspace{6cm}} \\
P(Y=1) &= P(\underline{\hspace{2cm}}) = \underline{\hspace{6cm}} \\
P(Y=2) &= P(\underline{\hspace{2cm}}) = \underline{\hspace{6cm}} \\
P(Y=3) &= P(\underline{\hspace{2cm}}) = \underline{\hspace{6cm}}
\end{aligned}
$$

[Exercice : Même question, mais cette fois $P(\text{pile}) = \frac{1}{3}$.]

VAs discrètes

- Une variable aléatoire ne pouvant prendre qu'une quantité *dénombrable* de valeurs est dite **discrète**
- Il y a *deux fonctions importantes* pour les VA discrètes :
- **Loi de probabilité** : $p(x) = P(X = x)$
  - $p(x) \geq 0$ pour chaque x
  - $\sum_x p(x) = 1$
- **Fonction de répartition** : $F(x) = P(X \leq x) = \sum_{i \leq x} p(i)$
  - $F(x)$ une fonction en escalier
  - $F(-\infty) = 0$
  - $F(\infty) = 1$
- En sachant la loi de X, *on pourrait répondre aux questions concernant les probabilités*

Exemple

Exemple 3.5 Soit X = la somme des nombres de deux dès lancés indépendamment.

- Trouver la loi de X ($p(x)$) et sa fonction de répartition ($F(x)$)

Solution

x	2	3	4	5	6	7	8	9	10	11	12	
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	
$F(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$	$[= 1]$

VA de Bernoulli

- Une VA de *Bernoulli* prend les valeurs 0 et 1
- Sa loi de probabilité est :

x	0	1
$p(x)$	$(1-p)$	p

- Utilisée dans la modelisation des problèmes ayant 2 résultats possibles : pile/face ; oui/non ; succès/échec ; *etc.*

La VA binomiale

- Les VAs de Bernoulli sont plus souvent considérées *n à la fois*
- Il existe *plusieurs situations concrètes* qui peuvent être considérées comme une séquence d'un nombre (fixe) n d'épreuves indépendantes, chacune ayant p pour probabilité de 'succès' et $(1-p)$ pour probabilité d''échec'
- (p. ex. lancements d'une pièce)
- La loi de probabilité de la *somme* d'un nombre n (fixe) épreuves de Bernoulli indépendantes, chacune avec probabilité de succès p, est la distribution **binomiale** de paramètres n et p
- On écrit $X \sim Bin(n, p)$: [X est distribuée comme Binomial de paramètres n et p, ou X est distribuée Binomial(n, p)]

La loi de probabilité d'une VA binomiale

- Facile à dériver en utilisant les principes fondamentaux (déjà fait en Exemple 3.4 ! !)
- Il y a **4 conditions** à satisfaire :
  1. nombre *fixe* (pas aléatoire) d'épreuves n
  2. 2 résultats possibles de chaque épreuve : *soit 1, soit 0*
  3. *la même probabilité* p pour chaque épreuve que le résultat soit 1
  4. les épreuves sont *indépendantes*.
- Donc si $X \sim Bin(n, p)$, la loi de probabilité est :

$$\boxed{P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n}$$

4 étapes pour résoudre les problèmes des VAs

1. *Identifier la VA*
2. Déterminer la *distribution* (loi) de la VA
3. *Traduire* la question
4. *Répondre* à la question

Exemples

Exemple 3.6 Quelle est la probabilité qu'en lançant 6 fois (indépendamment) une pièce équilibrée, on obtient au moins 4 piles ??