ENSIBS Cyber Data

Année universitaire 2022 / 2023

Probabilités

Responsable : A. Ridard

Avant-propos

Ce document est spécifiquement rédigé pour des séances de Cours/TD.

Il présente les éléments de cours habituels (définitions et propriétés) enrichis de remarques, indiquées par , donnant un certain éclairage pour mieux les comprendre, les retenir et les utiliser.

Ce cours est aussi ponctué d'exercices, indiqués par , qui seront traités en classe ou à la maison pour bien assimiler les différentes notions présentées. Grâce à ces exercices, vous allez fabriquer les exemples du cours ainsi que certaines preuves. C'est effectivement en étant acteur dans ses apprentissages que l'on profite au mieux des enseignements !

Ce document sera complété par des feuilles de TD ou TP pour s'entraîner d'avantage.

Les situations étudiées dans ce document sont plutôt issues des "jeux de hasard" (pièce, dé, urne, cartes). Elles constituent des modèles auxquels on peut se ramener par un travail de modélisation : transformation d'un problème concret (de la vraie vie) en un problème mathématique "type" que l'on sait résoudre. Les exercices de TD ou TP, quant à eux, seront plutôt issus de l'Informatique.

Bonne lecture, et bon travail...

Table des matières

I. Probabilités

1. Un peu de vocabulaire pour commencer

Une **expérience** est dite **aléatoire** [1] lorsque son résultat n'est pas connu à l'avance.
On note alors Ω l'ensemble des résultats possibles appelé **univers**.

Dans ce cours, sauf mention contraire, Ω désignera un ensemble fini $\{x_0, x_1, \ldots, x_N\}$.

La théorie des probabilités s'appuie sur la théorie des ensembles mais sa terminologie est spécifique :

- L'ensemble des parties de Ω est notée $\mathscr{P}(\Omega)$ et ses éléments sont appelés des **événements**.
  On dit alors que le couple $(\Omega, \mathscr{P}(\Omega))$ est un **espace probabilisable**
- Si A est un événement, $\overline{A}$ est l'**événement contraire** de A
- Ω est l'**événement certain** et $\varnothing$ l'**événement impossible**
- Deux **événements** A et B sont **incompatibles** si $A \cap B = \varnothing$
- Un **système complet d'événements** est une partition de Ω c'est à dire une famille [2] $(A_1, \ldots, A_n)$ d'événements :
  - possibles : $\forall i \in \{1, \ldots, n\},\ A_i \neq \varnothing$
  - incompatibles deux à deux : $\forall i, j \in \{1, \ldots, n\},\ i \neq j \Rightarrow A_i \cap A_j = \varnothing$
  - décrivant tout l'univers : $\Omega = \displaystyle\bigcup_{i=1}^{n} A_i$

2. Notion de probabilité

Définition (Probabilité)**.**

Une probabilité sur un espace probabilisable $(\Omega, \mathscr{P}(\Omega))$ est une application $\mathrm{P} : \mathscr{P}(\Omega) \to [0,1]$ vérifiant :

- $\mathrm{P}(\Omega) = 1$
- Si A et B sont des événements incompatibles, alors $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$ (additivité)

On dit alors que le triplet $(\Omega, \mathscr{P}(\Omega), \mathrm{P})$ est un espace probabilisé.

Définition (Probabilité uniforme)**.**

On appelle probabilité uniforme l'application :

$$\begin{array}{rccl} \mathrm{P}: & \mathscr{P}(\Omega) & \to & [0,1] \\ & A & \mapsto & \mathrm{P}(A) = \frac{Card(A)}{Card(\Omega)} \end{array}$$

Il existe d'autres probabilités

La probabilité uniforme est utilisée uniquement lorsque tous les résultats possibles sont équiprobables

Propriété (Propriétés élémentaires)**.**

- $\mathrm{P}(\varnothing) = 0$
- Si A est un événement, alors $\mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A)$
- Pour tout événement A et B, on a :
  - $A \subset B \Rightarrow \mathrm{P}(A) \leq \mathrm{P}(B)$
  - $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$
- Si $A_1, \ldots, A_n$ sont des événements incompatibles deux à deux, alors $\mathrm{P}\left(\displaystyle\bigcup_{i=1}^{n} A_i\right) = \displaystyle\sum_{i=1}^{n} \mathrm{P}(A_i)$

[1]. En opposition avec déterministe
[2]. En fait, la notion de famille est plus générale car elle autorise une infinité dénombrable d'événements, mais cela sort du cadre de ce cours

II. Probabilités conditionnelles et indépendance

1. Notion de probabilité conditionnelle

Définition (Probabilité conditionnelle)**.**

Soit B un événement réalisable c'est à dire vérifiant $\mathrm{P}(B) \neq 0$.
On appelle probabilité conditionnelle sachant B l'application :

$$\begin{array}{rccl} \mathrm{P}_B : & \mathscr{P}(\Omega) & \rightarrow & [0,1] \\ & A & \mapsto & \mathrm{P}_B(A) = \mathrm{P}(A|B) = \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)} \end{array}$$

- P_B est une probabilité !
- Si $\mathrm{P}(B) = 0$, on convient de poser $\mathrm{P}(A|B) = 0$

Propriété (Formule des probabilités composées)**.**

Si A et B sont des événements, alors $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B|A)$.

Généralisation :

- Si A, B, C sont des événements, alors $\mathrm{P}(A \cap B \cap C) = \mathrm{P}(A)\mathrm{P}(B|A)\mathrm{P}(C|A \cap B)$
- Si $A_1, \ldots, A_n$ sont des événements, alors $\mathrm{P}(A_1 \cap \cdots \cap A_n) = \mathrm{P}(A_1)\mathrm{P}(A_2|A_1)\mathrm{P}(A_3|A_1 \cap A_2)\ldots\mathrm{P}(A_n|A_1 \cap \cdots \cap A_{n-1})$

Une urne contient n boules blanches et n boules rouges.
On tire successivement et sans remise n boules dans cette urne.
On note A_k l'événement « la boule obtenue lors du k-ième tirage est blanche ».

Déterminer la probabilité qu'une boule rouge figure dans ce tirage en mesurant l'événement contraire.
On pourra étudier les cas $n = 2$ et $n = 3$ avant de répondre dans le cas général.

Propriété (Formule des probabilités totales)**.**

Soit A un événement de sorte que la famille $\left(A, \overline{A}\right)$ soit un système complet d'événements [a].
Si B est un événement [b], alors $\mathrm{P}(B) = P(B|A)\mathrm{P}(A) + P(B|\overline{A})\mathrm{P}(\overline{A})$.

Généralisation :

- Soit (A_1, A_2, A_3) un système complet d'événements.
  Si B est un événement, alors $\mathrm{P}(B) = P(B|A_1)\mathrm{P}(A_1) + P(B|A_2)\mathrm{P}(A_2) + P(B|A_3)\mathrm{P}(A_3)$
- Soit $(A_1, \ldots, A_n)$ un système complet d'événements.
  Si B est un événement, alors $\mathrm{P}(B) = \sum_{i=1}^{n} \mathrm{P}(B|A_i)\mathrm{P}(A_i)$

a. Si ce n'est pas le cas (A impossible ou certain), ce conditionnement n'aurait pas d'intérêt !
b. dont la réalisation "dépend" de celle de A

Problème de Monty Hall

Le jeu oppose un présentateur à un candidat (le joueur). Ce joueur est placé devant trois portes fermées. Derrière l'une d'elles se trouve une voiture (ou tout autre prix magnifique) et derrière chacune des deux autres se trouve une chèvre (ou tout autre prix sans importance). Il doit tout d'abord désigner une porte. Puis le présentateur doit ouvrir une porte qui n'est ni celle choisie par le candidat, ni celle cachant la voiture (le présentateur sait quelle est la bonne porte dès le début). Le candidat a alors le droit ou bien d'ouvrir la porte qu'il a choisie initialement, ou bien d'ouvrir la troisième porte.

Que doit faire le candidat ? Quelles sont ses chances de gagner la voiture en agissant au mieux ?
Indice : on pourra considérer G l'événement « le joueur gagne la voiture » et B celui « le joueur avait choisi la bonne porte », puis calculer $\mathrm{P}(G)$ si le joueur change de porte, et sinon.

On dispose de six urnes numérotées de 1 à 6.
L'urne numéro k comporte k boules blanches et une boule rouge.
Un joueur lance un dé équilibré puis choisit une boule dans l'urne correspondant au résultat du dé.
On note A_k l'événement « le dé donne la valeur k » et B l'événement « la boule tirée est blanche ».

Déterminer la probabilité que la boule tirée soit blanche.

Propriété (Formule de Bayes)**.**

Soit A un événement de sorte que la famille $\left(A, \overline{A}\right)$ soit un système complet d'événements.
Si B est un événement réalisable, alors $\mathrm{P}(A|B) = \dfrac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(B)} = \dfrac{\mathrm{P}(B|A)\mathrm{P}(A)}{P(B|A)\mathrm{P}(A) + P(B|\overline{A})\mathrm{P}(\overline{A})}$.

Généralisation :
Soit $(A_1, \ldots, A_n)$ un système complet d'événements.
Si B est un événement réalisable, alors pour tout $k \in \{1, \ldots, n\}$, $\mathrm{P}(A_k|B) = \dfrac{\mathrm{P}(B|A_k)\mathrm{P}(A_k)}{\sum\limits_{i=1}^{n} \mathrm{P}(B|A_i)\mathrm{P}(A_i)}$.

La formule de Bayes est utile pour les raisonnements « rétroactifs ». Si on sait mesurer la conséquence B d'un événement A et que l'on sait l'événement B réalisé, la formule de Bayes permet de savoir si l'événement A l'a été. On parle parfois de la formule de probabilité des causes.

Une urne contient deux dés : l'un est équilibré et l'autre donne systématiquement un 6.
On choisit un dé dans l'urne et on le lance.
On note A l'événement « le dé choisi est équilibré » et B l'événement « le dé lancé donne un 6 ».

En supposant que le dé lancé donne un 6, déterminer la probabilité que ce dé soit équilibré.

Problème de Monty Hall

On se propose ici de résoudre le problème de Monty Hall d'une autre manière.
On considère F_i l'événement « la voiture est derrière la porte i » et O_j celui « le présentateur ouvre la porte j ».
Pour fixer les idées, sans perdre de généralité [a], on peut supposer que le joueur ait choisi la porte 1, et le présentateur ouvert la porte 2.
Indice : Calculer $\mathrm{P}(F_1|O_2)$ et $\mathrm{P}(F_3|O_2)$, puis conclure [b].

a. Les calculs seraient les mêmes dans les autres cas
b. Le joueur peut évaluer les probabilités de F_1 et F_3 avec une information supplémentaire

2. Événements indépendants

Définition (Indépendance de deux événements)**.**

On dit que deux événements A et B sont indépendants si $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$

- Soit B un événement réalisable. Alors,
  A et B sont indépendants si et seulement si $\mathrm{P}(A|B) = \mathrm{P}(A)$ autrement dit, savoir que B est réalisé n'apporte rien !
- Si A et B sont des événements indépendants, alors A et $\overline{B}$ le sont aussi, tout comme $\overline{A}$ et $\overline{B}$.

1. On lance deux fois un dé équilibré.
   Les événements A : « le premier lancer donne un six » et B : « le second lancer donne un six » sont-ils indépendants ?
2. On tire successivement sans remise deux boules dans une urne contenant 5 boules blanches et 2 boules rouges.
   Les événements A : « la première boule tirée est blanche » et B : « la seconde boule tirée est blanche » sont-ils indépendants ?
3. Même question avec un tirage successif avec remise.

Ne pas confondre indépendance et incompatibilité

D'après le tableau suivant, dans quel(s) cas les événements A et B sont-ils indépendants ?

	$\mathrm{P}(A)$	$\mathrm{P}(B)$	$\mathrm{P}(A \cup B)$
cas 1	0.1	0.9	0.91
cas 2	0.4	0.6	0.76
cas 3	0.5	0.3	0.8

Définition (Indépendance mutuelle d'événements)**.**

On dit que les événements $A_1, \ldots, A_n$ sont mutuellement indépendants si pour tout J finie [a] $\subset \{1, \ldots, n\}$, on a :

$$\mathrm{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathrm{P}(A_j)$$

a. Inutile ici, mais permet de généraliser la définition à une infinité dénombrable d'événements

Ne pas confondre indépendance deux à deux et indépendance mutuelle

On lance deux dés discernables et l'on considère les événements :
- A « le premier dé lancé donne un résultat pair »
- B « le second dé lancé donne un résultat pair »
- C « la somme des deux dés est paire »

Montrer que les événements A, B et C sont deux à deux indépendants, mais pas mutuellement indépendants.

III. Variables aléatoires finies

1. Loi de probabilité

Définition (Variable aléatoire)**.**

Une variable aléatoire (v.a.) est une application mesurable [a] $X:\Omega \to X(\Omega)$
où $X(\Omega)=\{x_0,x_1,\dots,x_N\}$ désigne l'ensemble des valeurs prises par X.

a. Cette notion de la théorie de la mesure ne sera évidemment pas détaillée ici. En gros, elle assure de pouvoir mesurer la probabilité de tout événement sur $X(\Omega)$ à l'aide de la probabilité P définie sur Ω

- L'appellation variable aléatoire est usuelle bien que malheureuse. En effet, X n'est pas une variable, mais bien une fonction et celle-ci n'est pas aléatoire, mais plutôt déterministe. Ce sont les valeurs prises par X qui correspondent à des quantités qui vont varier selon le résultat de l'expérience aléatoire.
- Il est tout à fait possible d'utiliser une v.a. $X:\Omega \to X(\Omega)$ sans préciser l'espace probabilisé $(\Omega,\mathscr{P}(\Omega),\mathrm{P})$.
  C'est même ce qui fait la force de cet objet en le rendant très pratique !

Somme de deux dés équilibrés

On note X la v.a. représentant la somme de deux dés équilibrés.

Compléter la définition suivante :

$$X: \begin{array}{ccc} \dots & \longrightarrow & \dots \\ \dots & \longmapsto & \dots \end{array}$$

Définition (Loi de probabilité)**.**

On appelle loi d'une v.a. $X:\Omega \to X(\Omega)$, l'application :

$$\mathrm{P}_X: \begin{array}{ccl} \mathscr{P}(X(\Omega)) & \to & [0,1] \\ A & \mapsto & \mathrm{P}_X(A)=\mathrm{P}\big(X^{-1}(A)\big)=\mathrm{P}\big(\{\omega\in\Omega | X(\omega)\in A\}\big)=\mathrm{P}(X\in A) \end{array}$$

- P_X est une probabilité !
- L'existence de $\mathrm{P}\big(X^{-1}(A)\big)$ est assurée par la mesurabilité de X
- La dernière égalité fournit une notation très pratique

Somme de deux dés équilibrés - suite

Calculer la probabilité que la somme soit inférieure ou égale à 3 c'est à dire $\mathrm{P}_X\big(\{2,3\}\big)=\mathrm{P}\big(X\in\{2,3\}\big)=\mathrm{P}(X\leq 3)$.

Propriété (Détermination d'une loi de probabilité)**.**

La loi d'une v.a. $X:\Omega\to X(\Omega)$ est entièrement déterminée par :
- les valeurs possibles : $X(\Omega)=\{x_0,x_1,\dots,x_N\}$
- pour chacune des valeurs possibles, la probabilité associée : $\forall i\in\{0,\dots,N\},\ p_i=\mathrm{P}(X=x_i)$

Une telle loi est souvent présentée à l'aide d'un tableau :

Valeurs possibles x_i	x_0	x_1	$\dots$	x_N
Probabilités associées p_i	p_0	p_1	$\dots$	p_N

Bien entendu, on a : $\sum_{i=0}^{N} p_i = 1$.

Elle est représentée graphiquement par un diagramme en bâtons

Somme de deux dés équilibrés - suite

1. Déterminer la loi de X à l'aide d'un tableau.
2. Représenter graphiquement cette loi.
3. Calculer $\mathrm{P}(X \leq 3)$.

2. Fonction de répartition

Définition (Fonction de répartition)**.**

La fonction de répartition d'une v.a. $X : \Omega \to X(\Omega)$ est définie par :

$$\begin{array}{rccl} F : & \mathbb{R} & \longrightarrow & [0,1] \\ & x & \longmapsto & \mathrm{P}(X \leq x) = \displaystyle\sum_{i | x_i \leq x} p_i \end{array}$$

- Elle est représentée graphiquement par une fonction en escalier où la marche x_i est de hauteur p_i
- Elle est croissante, tend vers 0 en $-\infty$ et vers 1 en $+\infty$
- Elle est continue à droite
- Cette notion est fondamentale pour simuler des lois à l'aide d'un ordinateur

Somme de deux dés équilibrés - suite

1. Représenter graphiquement la fonction de répartition de X.
2. Retrouver, graphiquement, $\mathrm{P}(X \leq 3)$.

Propriété (Caractérisation d'une loi de probabilité)**.**

La loi d'une v.a. est caractérisée par sa fonction de répartition :
- Les valeurs possibles x_i sont les points de discontinuité (à gauche) de F
- Les probabilités associées p_i sont déterminées par :

$$p_i = \mathrm{P}(X = x_i) = \mathrm{P}(X \leq x_i) - \mathrm{P}(X \leq x_{i-1}) = F(x_i) - F(x_{i-1}) \quad \text{pour } i \in \{1, \dots, N\} \quad \text{et } p_0 = F(x_0)$$

Somme de deux dés équilibrés - suite

A partir du graphique précédent, retrouver la loi de X.

3. Espérance

Définition (Espérance)**.**

On appelle espérance de X le réel défini par $E(X)=\sum_{i=0}^{N} x_i p_i$.

- C'est la moyenne des valeurs possibles pondérées par les probabilités associées
- Les différentes valeurs possibles se répartissent autour de $E(X)$ qui est un indicateur de tendance centrale
- Lorsque $E(X)=0$, on dit que X est centrée
- Si X est constante égale à c, alors $E(X)=c$

Somme de deux dés équilibrés - suite

Déterminer l'espérance de X.

Propriété (Linéarité)**.**

Pour tout $\lambda\in\mathbb{R}$, $E(\lambda X+Y)=\lambda E(X)+E(Y)$.
En particulier, pour tout $a,b\in\mathbb{R}$, $E(aX+b)=aE(X)+b$.

- L'espérance d'une combinaison linéaire est la combinaison linéaire des espérances
- La v.a. $X-E(X)$ est centrée

Somme de deux dés équilibrés - suite

On considère que la somme obtenue X permet, à chaque élève d'une classe de CE1 lançant les deux dés, de gagner Y Smarties, égal au double de la somme augmenté de un (petit jeu de fin d'année organisé par une maîtresse pour faire calculer mentalement ses élèves).

Déterminer le gain moyen qui permettra à cette maîtresse d'une classe de 30 élèves de prévoir [a] le bon nombre de Smarties.

a. cette prévision repose sur la loi forte des grands nombres qui sera vue dans un prochain module

Propriété (Théorème du transfert)**.**

Soit g une fonction réelle définie sur $X(\Omega)$.

L'espérance de la v.a. $g(X)$ est alors $E\big(g(X)\big)=\sum_{i=0}^{N} g(x_i)p_i$.

Somme de deux dés équilibrés - suite

Déterminer l'espérance de X^2.

4. Variance et écart-type

Définition (Variance et écart-type)**.**

On appelle variance de X le réel (positif) défini par $V(X) = E\left(\left(X - E(X)\right)^2\right) = \sum_{i=0}^{N} \left(x_i - E(X)\right)^2 p_i$.

On définit aussi son écart-type : $\sigma(X) = \sqrt{V(X)}$.

- La variance et l'écart-type permettent de mesurer la dispersion de X autour de sa moyenne
- Si X se comprend avec une unité, l'espérance et l'écart-type s'expriment avec la même unité
- Lorsque $V(X) = 1$, on dit que X est réduite
- Si X est constante égale à c, alors $V(X) = 0$
- Pour calculer une variance, on préférera la formule suivante

Pourquoi choisir de mesurer la dispersion de X autour de sa moyenne à l'aide de la moyenne des carrés des écarts à la moyenne (variance), plutôt qu'à l'aide de la moyenne des écarts à la moyenne ?

Propriété (Formule de Huygens)**.**

$V(X) = E\left(X^2\right) - E(X)^2$

Démontrer cette formule.

Somme de deux dés équilibrés - suite

Déterminer, à l'aide de cette formule, la variance de X.

Propriété (Changement affine)**.**

Pour tout $a, b \in \mathbb{R}$, $V(aX + b) = a^2 V(X)$

La v.a. $\frac{X - E(X)}{\sigma(X)}$ est centrée réduite

Somme de deux dés équilibrés - suite et fin

Déterminer l'écart-type de $Y = 2X + 1$.

5. Lois usuelles

a. Loi uniforme

Définition (Loi uniforme)**.**

On dit que X suit une loi uniforme sur $\{1,\dots,n\}$, et l'on note $X \sim \mathscr{U}_{\{1,\dots,n\}}$ si :

- $X(\Omega) = \{1,\dots,n\}$
- $\forall k \in \{1,\dots,n\},\ \mathrm{P}(X=k) = \frac{1}{n}$

La loi uniforme sur $\{1,\dots,n\}$ modélise une situation d'équiprobabilité.

Propriété (Espérance et variance)**.**

Si $X \sim \mathscr{U}_{\{1,\dots,n\}}$, alors :

- $E(X) = \frac{n+1}{2}$
- $V(X) = \frac{n^2-1}{12}$

1. Démontrer la formule de l'espérance.
2. En admettant que $\displaystyle\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$, démontrer la formule de la variance.

On lance un dé équilibré et on considère X la v.a. égale au résultat obtenu.

1. Reconnaître la loi de X.
2. Représenter graphiquement la loi et la fonction de répartition de X.
3. Déterminer l'espérance et la variance de X.

b. Loi de Bernoulli

Définition (Loi de Bernoulli)**.**

On dit que X suit une loi de Bernoulli de paramètre p, et l'on note $X \sim \mathscr{B}(p)$ si :

- $X(\Omega) = \{0,1\}$
- $\mathrm{P}(X=1) = p$ et $\mathrm{P}(X=0) = 1-p = q$

La loi de Bernoulli de paramètre p modélise une expérience aléatoire à deux issues possibles (succès noté 1 et échec noté 0) de probabilité de succès p.

Propriété (Espérance et variance)**.**

Si $X \sim \mathscr{B}(p)$, alors :

- $E(X) = p$
- $V(X) = p(1-p)$

Démontrer ces deux formules.

On lance une pièce équilibrée et on considère X la v.a. égale à 1 si on obtient pile et 0 sinon.

1. Reconnaître la loi de X.
2. Représenter graphiquement la loi et la fonction de répartition de X.
3. Déterminer l'espérance et la variance de X.

c. Loi binomiale

Définition (Loi binomiale)**.**

On dit que X suit une loi binomiale de paramètres n et p, et l'on note $X \sim \mathscr{B}(n,p)$ si :

- $X(\Omega) = \{0,\dots,n\}$
- $\forall k \in \{0,\dots,n\},\ \mathrm{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

La loi binomiale de paramètres n et p modélise le nombre de succès à l'issue de n répétitions indépendantes d'une expérience aléatoire à deux issues possibles de probabilité de succès p.

Propriété (Espérance et variance)**.**

Si $X \sim \mathscr{B}(n,p)$, alors :

- $E(X) = np$
- $V(X) = np(1-p)$

On lance 10 fois une pièce équilibrée et on considère X la v.a. égale au nombre de piles obtenus.

1. Reconnaître la loi de X.
2. Représenter graphiquement la loi et la fonction de répartition de X.
3. Déterminer l'espérance et la variance de X.
4. Déterminer la probabilité d'obtenir au moins 2 piles.

6. Variables aléatoires indépendantes

Définition (v.a. indépendantes)**.**

Les v.a. X et Y sont indépendantes si pour tout $(i,j) \in \{0,\dots,N_X\} \times \{0,\dots,N_Y\}$, on a :

$$\mathrm{P}\big((X=x_i)\cap(Y=y_j)\big) = \mathrm{P}(X=x_i)\mathrm{P}(Y=y_j)$$

- En s'inspirant de l'indépendance mutuelle d'une famille d'événements, on définit l'indépendance mutuelle d'une famille[a] de v.a. : elle exprime alors le fait que la probabilité de n'importe quelle intersection construite à partir des v.a. coïncide avec le produit des probabilités
- Comme pour les événements, des v.a. deux à deux indépendantes ne sont pas, en général, mutuellement indépendantes

a. Cette notion sera surtout utilisée avec des suites de v.a.

Propriété (Variance d'une somme)**.**

Si X et Y sont indépendantes, alors $V(X+Y) = V(X)+V(Y)$.

Généralisation :

Si $X_1, \dots, X_n$ sont indépendantes deux à deux, alors $V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i)$

Faux en général

En fait, $V(X+Y) = V(X)+V(Y)+2Cov(X,Y)$ où $Cov(X,Y) = E\Big(\big(X-E(X)\big)\big(Y-E(Y)\big)\Big)$ désigne la covariance[a] de X et Y

a. cette notion sera vue dans un prochain module

On considère $X_1, \dots, X_n$ des variables aléatoires indépendantes et de même loi, la loi $\mathscr{B}(p)$.
On note $X = X_1 + \cdots + X_n$.

1. Déterminer l'espérance et la variance de X.
2. Reconnaître la loi de X.
3. Quel résultat du cours vient-on de démontrer ?

IV. Couples de variables aléatoires finies

On considère $X:\Omega\to E$ et $Y:\Omega\to F$ deux v.a. finies telles que $X(\Omega)=\{x_i\,|\,i\in I\}$ et $Y(\Omega)=\left\{y_j\,|\,j\in J\right\}$ avec I et J finis.

1. Loi conjointe

Définition (Couple de v.a.)**.**

Le couple (X,Y) représente la variable aléatoire (vectorielle), que l'on notera plus simplement $\mathbf{Z}$, définie par :

$$\mathbf{Z}=(X,Y):\begin{array}{ccc}\Omega & \to & E\times F\\ \omega & \mapsto & \mathbf{Z}(\omega)=\big(X(\omega),Y(\omega)\big)\end{array}$$

$\mathbf{Z}(\Omega)\subset X(\Omega)\times Y(\Omega)$ mais l'autre inclusion n'est pas toujours vraie.

Définition (Loi conjointe)**.**

On appelle loi conjointe de (X,Y) la loi de $\mathbf{Z}$.

- Elle est entièrement déterminée par la connaissance de $\mathrm{P}\big((X=x_i)\cap(Y=y_j)\big)$ pour tout $\left(x_i,y_j\right)\in X(\Omega)\times Y(\Omega)$.
- Elle est souvent présentée à l'aide d'un tableau.

Une urne comporte 2 boules blanches, 1 rouge et 1 noire.
On tire simultanément deux boules de cette urne.
On note X le nombre de boules blanches et Y le nombre de boules noires tirées.

1. Compléter le tableau de la loi conjointe de (X,Y) :

	$X=0$	$X=1$	$X=2$
$Y=0$			
$Y=1$			

2. Que vaut la somme des cases ?
3. A-t-on $\mathbf{Z}(\Omega)=X(\Omega)\times Y(\Omega)$ où $\mathbf{Z}=(X,Y)$?

2. Lois marginales

Définition (Lois marginales)**.**

On appelle lois marginales de (X,Y) la loi de X et celle de Y.

Propriété (Détermination des lois marginales)**.**

La loi conjointe détermine entièrement les lois marginales :

- $\forall i\in I,\ \mathrm{P}(X=x_i)=\sum\limits_{j\in J}\mathrm{P}\big((X=x_i)\cap(Y=y_j)\big)$
- $\forall j\in J,\ \mathrm{P}(Y=y_j)=\sum\limits_{i\in I}\mathrm{P}\big((X=x_i)\cap(Y=y_j)\big)$

Suite

Augmenter le tableau de la loi conjointe avec les lois marginales :

	$X=0$	$X=1$	$X=2$	P_Y
$Y=0$	0	$\frac{1}{3}$	$\frac{1}{6}$	
$Y=1$	$\frac{1}{6}$	$\frac{1}{3}$	0	
P_X				

Les lois marginales ne suffisent pas à déterminer la loi conjointe

	$X=0$	$X=1$	P_Y
$Y=0$	0	$\frac{1}{2}$	$\frac{1}{2}$
$Y=1$	$\frac{1}{2}$	0	$\frac{1}{2}$
P_X	$\frac{1}{2}$	$\frac{1}{2}$	

	$X=0$	$X=1$	P_Y
$Y=0$	$\frac{1}{2}$	0	$\frac{1}{2}$
$Y=1$	0	$\frac{1}{2}$	$\frac{1}{2}$
P_X	$\frac{1}{2}$	$\frac{1}{2}$	

3. Lois conditionnelles

Définition (Lois conditionnelles)**.**

On appelle loi conditionnelle de Y sachant $X = x_i$, la loi de Y pour la probabilité conditionnelle $\mathrm{P}(.|X = x_i)$, autrement dit chaque valeur possible y_j de Y est associée à $\mathrm{P}\left(Y = y_j|X = x_i\right)$.

Suite

Compléter le tableau des lois conditionnelles de Y :

	$X=0$	$X=1$	$X=2$
$\mathrm{P}(Y=0\|X=x_i)$			
$\mathrm{P}(Y=1\|X=x_i)$			

4. Espérance et variance

Définition (Espérance de $\mathbf{Z}$)**.**

L'espérance du couple $\mathbf{Z} = (X, Y)$ est le couple des espérances :

$$E(\mathbf{Z}) = E\big((X,Y)\big) = \big(E(X), E(Y)\big)$$

- $E(\mathbf{Z})$ est un vecteur (couple) que l'on confond parfois avec sa matrice colonne (cf.ci-dessous)
- Plus généralement, l'espérance d'une matrice est la matrice des espérances (cf. ci-dessous)

Suite

Déterminer $E(\mathbf{Z})$.

Définition (Variance de $\mathbf{Z}$)**.**

La variance du couple $\mathbf{Z} = (X, Y)$ est définie par :

$$V(\mathbf{Z}) = E\Big(\big(\mathbf{Z} - E(\mathbf{Z})\big)\big(\mathbf{Z} - E(\mathbf{Z})\big)^t\Big) = \begin{pmatrix} E\Big(\big(X - E(X)\big)\big(X - E(X)\big)\Big) & E\Big(\big(X - E(X)\big)\big(Y - E(Y)\big)\Big) \\ E\Big(\big(Y - E(Y)\big)\big(X - E(X)\big)\Big) & E\Big(\big(Y - E(Y)\big)\big(Y - E(Y)\big)\Big) \end{pmatrix}$$

En définissant la covariance de X et Y par :

$$Cov(X, Y) = E\Big(\big(X - E(X)\big)\big(Y - E(Y)\big)\Big)$$

on obtient :

$$V(\mathbf{Z}) = \begin{pmatrix} V(X) & Cov(X, Y) \\ Cov(Y, X) & V(Y) \end{pmatrix}$$

$V(\mathbf{Z})$ est une matrice appelée, d'après la remarque précédente, matrice de variance-covariance.

- On a évidemment $Cov(X, Y) = Cov(Y, X) = E(XY) - E(X)E(Y)$ donc $V(\mathbf{Z})$ est une matrice symétrique
- Si X et Y sont indépendantes, alors $E(XY) = E(X)E(Y)$ donc $Cov(X, Y) = 0$ et donc $V(\mathbf{Z})$ est une matrice diagonale

Suite

Déterminer $V(\mathbf{Z})$.

Définition (Espérance conditionnelle de Y sachant $X = x_i$)**.**

On appelle espérance conditionnelle de Y sachant $X = x_i$, l'espérance de la loi conditionnelle de Y sachant $X = x_i$:

$$E(Y|X = x_i) = \sum_{j \in J} y_j \mathrm{P}(Y = y_j | X = x_i)$$

Suite

Déterminer $E(Y|X = 0)$, $E(Y|X = 1)$ et $E(Y|X = 2)$.

Définition (Espérance conditionnelle de Y sachant X)**.**

On appelle espérance conditionnelle de Y sachant X, la v.a. $E(Y|X)$ dont les valeurs possibles sont les $E(Y|X = x_i)$ de probabilité associée $\mathrm{P}(X = x_i)$.

Suite

Déterminer la loi de la v.a. $E(Y|X)$.

$E(Y|X)$ est une variable aléatoire vérifiant :

- $E\Big(E(Y|X)\Big) = E(Y)$
- $V\Big(E(Y|X)\Big) \leq V(Y)$

Suite et fin

Observer ces deux résultats sur notre exemple.

Cette section peut se généraliser à des vecteurs $(X_1, \dots, X_n)$ de v.a. finies et même dénombrables ou réelles continues !

V. Chaînes de Markov sur un espace fini d'états

Dans cette section, on considère l'ensemble (des états) $E = \{1, \dots, n\}$**.**

1. Matrice stochastique

Définition (Matrice stochastique)**.**

On dit que la matrice $A = \left(a_{ij}\right)_{i,j\in E}$ est stochastique si :

- elle est positive : $\forall i, j \in E,\ a_{ij} \geq 0$
- la somme sur chaque colonne vaut 1 : $\forall j \in E,\ \sum_{i\in E} a_{ij} = 1$

- Le réel a_{ij} peut s'interpréter comme la probabilité de passer de l'état j à l'état i
- La matrice A peut se représenter à l'aide d'un graphe probabiliste (orienté et pondéré)

Dans la littérature, c'est plutôt la somme de chaque ligne qui vaut 1.
Mon choix[a] est moins cohérent avec la matrice d'un graphe, mais plus avec la matrice d'une application linéaire !

a. Pour suivre l'article « Comment fonctionne Google ? »

Représenter les graphes probabilistes pour les matrices stochastiques suivantes :

$$A_1 = \begin{pmatrix} 0.65 & 0.45 \\ 0.35 & 0.55 \end{pmatrix} \quad A_2 = \begin{pmatrix} 0.9 & 0.7 & 0.8 \\ 0.05 & 0 & 0 \\ 0.05 & 0.3 & 0.2 \end{pmatrix} \quad A_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad A_4 = \begin{pmatrix} 0.9 & 0.7 & 0.6 \\ 0.1 & 0.3 & 0.2 \\ 0 & 0 & 0.2 \end{pmatrix}.$$

Définition (Vecteur stochastique ou loi de probabilité)**.**

On dit qu'un vecteur $\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$ est stochastique si :

- il est positif : $\forall i \in E,\ \mu_i \geq 0$
- la somme de ses coordonnées vaut 1 : $\sum_{i\in E} \mu_i = 1$

Par abus, on confond le vecteur avec sa matrice colonne formée de ses coordonnées (dans la base canonique)

Dorénavant, sauf mention contraire, A **désignera une matrice stochastique et** μ **un vecteur stochastique.**

Définition (Loi de probabilité invariante)**.**

On dit que la loi de probabilité μ est invariante pour A si $A\mu = \mu$.

Autrement dit, si μ est un vecteur propre de A associé à la valeur propre 1

Définition (Matrice ergodique)**.**

On dit que la matrice A est ergodique s'il existe une puissance de A strictement positive[a] :

$$\exists k \geq 1,\ A^k > 0$$

a. Une matrice est dite strictement positive lorsque tous ses coefficients sont strictement positifs

- En notant $a_{ij}^{(k)}$ les coefficients de A^k, on a :

$$A^k > 0 \Leftrightarrow \forall i, j \in E,\ a_{ij}^{(k)} > 0$$

- Si A est strictement positive, alors elle est ergodique
- Si A est ergodique, alors elle est irréductible[a]

a. Matrice (positive) avec une seule classe d'équivalence pour la relation de communication :

$$i \leftrightarrow j \Leftrightarrow \exists k, k' \geq 1,\ a_{ij}^{(k)} > 0 \quad \text{et} \quad a_{ji}^{(k')} > 0$$

Les matrices A_1, A_2, A_3, A_4 sont-elles ergodiques?

Propriété (Existence et unicité de la loi de probabilité invariante)**.**

Si A est ergodique, alors il existe une unique loi de probabilité μ invariante pour A.

- De plus, $\mu > 0$
- Cette propriété reste vraie si A est seulement irréductible
- En fait, c'est une conséquence du théorème de Frobenius[a]

a. Le rayon spectral d'une matrice positive irréductible est une valeur propre simple associée à un vecteur propre strictement positif

Vérifier le deuxième point de la remarque pour la matrice A_3.

Vérifier la propriété pour les matrices A_1, A_2 (TP4).

Propriété (Théorème d'ergodicité)**.**

Si A est ergodique, alors pour tout $i, j \in E$:

$$a_{ij}^{(k)} \xrightarrow[k \to +\infty]{} \mu_i$$

avec μ l'unique loi de probabilité invariante pour A.

Si A est seulement irréductible, alors on a seulement la convergence en moyenne de Césaro :

$$\frac{a_{ij}^{(1)} + a_{ij}^{(2)} + \cdots + a_{ij}^{(k)}}{k} \xrightarrow[k\to+\infty]{} \mu_i$$

Vérifier la propriété pour les matrices A_1, A_2 et la remarque pour la matrice A_3 (TP5).

2. Chaîne de Markov

Définition (Chaîne de Markov)**.**

Une suite de v.a. $(X_k)_{k\in\mathbb{N}}$ est une chaîne de Markov si pour tout $k \in \mathbb{N}$ et tout $i, j, j_0, \ldots, j_{k-1} \in E$, on a :

$$\mathrm{P}\left(X_{k+1} = i \mid X_0 = j_0, \ldots, X_{k-1} = j_{k-1}, X_k = j\right) = \mathrm{P}\left(X_{k+1} = i \mid X_k = j\right)$$

« La prédiction du futur connaissant le présent et le passé n'est pas meilleure que celle connaissant uniquement le présent »

Définition (Chaîne de Markov homogène)**.**

Une chaîne de Markov $(X_k)_{k\in\mathbb{N}}$ est dite homogène si pour tout $k \in \mathbb{N}$, on a :

$$\mathrm{P}\left(X_{k+1} = i \mid X_k = j\right) = \mathrm{P}\left(X_1 = i \mid X_0 = j\right)$$

La matrice $A = \left(\mathrm{P}\left(X_1 = i \mid X_0 = j\right)\right)_{i,j\in E}$ est sa matrice de transition.
La loi de X_0 est sa loi initiale.

- On se limitera aux chaînes de Markov homogènes
- A est une matrice stochastique
- Une telle chaîne représente l'évolution d'un système dynamique à n états possibles, au cours du temps (discret), étant donnée une loi initiale et en supposant qu'à chaque instant le changement est régi par la matrice de transition A

Propriété (Loi conjointe)**.**

La loi conjointe de $(X_0, X_1, \ldots, X_k)$ est déterminée par la matrice de transition $A = \left(a_{ij}\right)_{i,j\in E}$ et la loi initiale :

$$\forall\, i_0, i_1, \ldots, i_k \in E,\ \mathrm{P}\big(X_0 = i_0, X_1 = i_1, \ldots, X_k = i_k\big) = \mathrm{P}\big(X_0 = i_0\big) a_{i_1 i_0} a_{i_2 i_1} \ldots a_{i_k i_{k-1}}$$

Évidemment, on a : $\mathrm{P}\big(X_0 = i_0, X_1 = i_1, \ldots, X_k = i_k\big) = \mathrm{P}\Big((X_0 = i_0) \cap (X_1 = i_1) \cap \cdots \cap (X_k = i_k)\Big)$

Démontrer la propriété.

Propriété (Lois conditionnelles)**.**

La matrice de transition en k pas $\left(\mathrm{P}\left(X_{m+k}=i \mid X_m=j\right)\right)_{i,j\in E}$ est égale à A^k :

$$\forall i,j\in E,\ \mathrm{P}\left(X_{m+k}=i \mid X_m=j\right)=a_{ij}^{(k)}$$

Démontrer la propriété.

Propriété (Lois marginales)**.**

La loi de X_{m+k} est donnée par :

$$\forall i\in E,\ \mathrm{P}\big(X_{m+k}=i\big)=\sum_{j\in E} a_{ij}^{(k)}\mathrm{P}\big(X_m=j\big)$$

En particulier, la loi de X_k est donnée par :

$$\forall i\in E,\ \mathrm{P}\big(X_k=i\big)=\sum_{j\in E} a_{ij}^{(k)}\mathrm{P}\big(X_0=j\big)$$

En notant $\mu^{(l)}$ la loi de X_l, on peut reformuler de la manière suivante :

$$\mu^{(m+k)}=A^k\mu^{(m)}\quad\text{et}\quad\mu^{(k)}=A^k\mu^{(0)}$$

ce qui n'est pas sans rappeler les formules sur les suites géométriques

Démontrer la propriété.

Propriété (Théorème d'ergodicité)**.**

Si A est ergodique, alors la loi de X_k tend vers l'unique loi de probabilité μ invariante pour A :

$$\mu^{(k)}\xrightarrow[k\to+\infty]{\mathscr{L}}\mu$$

Le comportement asymptotique ne dépend pas de la loi initiale!

Doudou le hamster ne connaît que trois endroits dans sa cage : les copeaux où il dort, la mangeoire où il mange et la roue où il fait de l'exercice. Ses journées sont assez semblables les unes aux autres, et son activité se représente aisément par une chaîne de Markov. Toutes les minutes, il peut soit changer d'activité, soit continuer celle qu'il était en train de faire. L'appellation processus sans mémoire n'est pas du tout exagérée pour parler de Doudou. Quand il dort, il a 9 chances sur 10 de ne pas se réveiller la minute suivante. Quand il se réveille, il y a 1 chance sur 2 qu'il aille manger et 1 chance sur 2 qu'il parte faire de l'exercice. Le repas ne dure qu'une minute, après il fait autre chose. Après avoir mangé, il y a 3 chances sur 10 qu'il parte courir dans sa roue, mais surtout 7 chances sur 10 qu'il retourne dormir. Courir est fatigant pour Doudou; il y a 8 chances sur 10 qu'il retourne dormir au bout d'une minute. Sinon il continue en oubliant qu'il est déjà un peu fatigué.

En supposant que Doudou soit en train de dormir (est-ce important?), prévoir son comportement dans deux heures (TP5).