Probabilités et Statistique, module M3201
DUT Informatique

Pierre-Cyrille Héam

18 décembre 2014

Chapitre 1

Introduction

La Statistique peut être définie comme la sciences des données, de leur collecte, de leur analyse et de leur présentation. S'il peut sembler facile de collecter quelques données et de calculer quelques valeurs, comme des moyennes, puis de les interpréter abusivement, il est en revanche beaucoup plus complexe de le faire dans un cadre scientifique rigoureux. Tout scientifique, qu'il soit informaticien, biologiste, chimiste, économiste, etc., et même tout citoyen, est confronté à des données et statistiques qu'il doit savoir analyser avec recul. L'étude en pratique (par des benchmarck en général) des performances de différents composants informatiques se fait par des études statistiques : comment collecter les données (quels tests faire), comment les analyser (que permettent-ils de conclure) et comment les présenter. Il est facile de se tromper sur l'un des trois points et d'arriver à des conclusions sans fondement.

Les probabilités sont un domaine des mathématiques ayant pour objet l'étude de l'incertitude et du hasard. Il ne faut pas confondre probabilités et statistique. Un processus stochastique (probabiliste) peut servir à produire des données que l'on peut analyser de façon statistique. De même, on cherche souvent, à partir d'un jeu de données, à modéliser leur apparition comme si elle venait d'un processus probabiliste, afin d'en déduire des modèles théoriques et prédicatifs. Les probabilités sont utilisées dans de nombreuses sciences (physique, biologie, économie,...) et aussi en informatique : la sécurité des communications (génération des clés, chiffrage, challenge de non-rejeu,...), les réseaux et systèmes distribués (gestion des collisions, élection de leader, routage, gestion des files d'attentes, ...), l'algorithmique (algorithmes de type Monte-Carlo ou Las Vegas, optimisation des tris), la simulation (systèmes multi-agents, simulation numérique,...) utilisent de façon importante et/ou cruciale les probabilités.

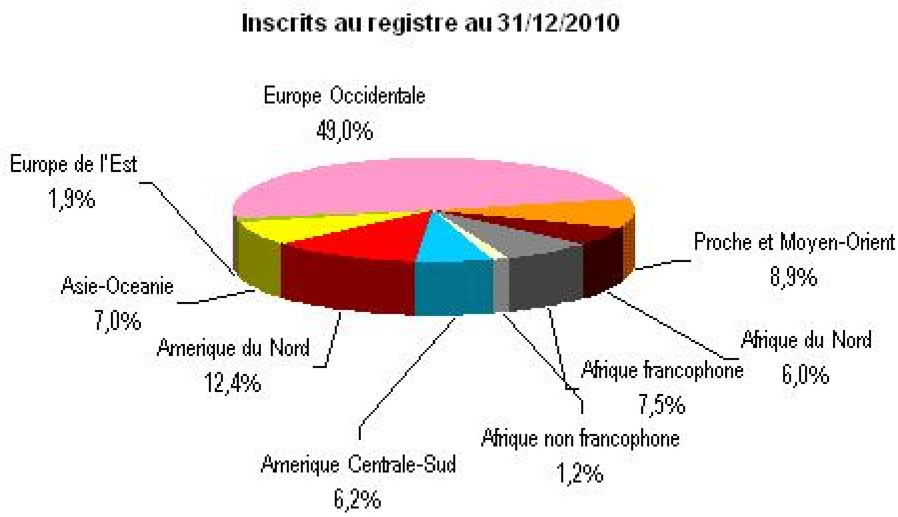


Figure 1.1 – Camembert en relief

1.1 Activité 1 : statistiques descriptives

Lorsque l'on présente des résultats statistique pour comparer des proportions, il est fréquent d'utiliser des camemberts ou des histogrammes. Un exemple est donné dans la figure 1.1 [1]. Cette présentation, en relief, est fréquemment utilisée car *plus jolie*. Cependant, elle est tendancieuse, car ce que l'oeil compare les aires des différents secteurs, et

1. http://france3-regions.blog.francetvinfo.fr/ftv-expats/2011/12/04/qui-sont-les-nouveaux-expatries-francais.html

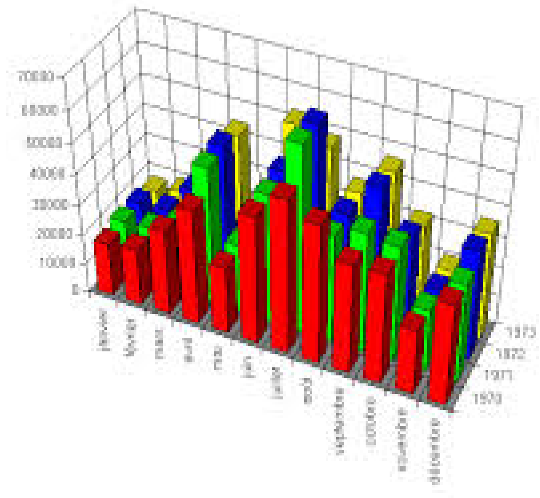
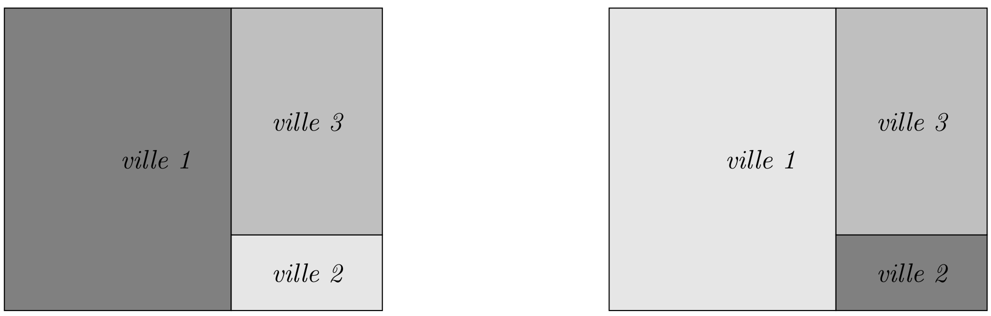


FIGURE 1.2 – Camembert en relief

l'utilisation de tranches (pour le relief) donne visuellement un poids plus fort aux données placées devant. Par exemple, sur la figure 1.1, l'Europe Occidentale occupe 50% de la surface de l'ellipse marquant le camenbert, mais bien moins de 50% de la surface totale du dessin, alors que cela devrait être le cas. De même, on peut remarque que l'Europe de l'Est paraît occuper une part moins importante que l'Afrique non francophone, ce qui n'est pas le cas. De même l'Amérique Centrale-Sud occupe plus place que le Proche et Moyen-Orient sur le dessin, alors que cela ne devrait pas être le cas. Dans un camembert en relief, les données placées sur l'avant on tendance à être surestimées. En pratique il convient donc de ne pas utiliser les camembert en relief, ou, lorsque cela est fait, il faut choisir une très faible épaisseur relativement au rayon du disque et une inclinaison par trop importante, afin de minimiser le biais. De même tout découpage en tranche (en faisant ressortir une tranche) ajoute de l'épaisseur visuelle et accentue l'effet visuel. Il est aussi important de faire attention aux couleurs : des couleurs vives, comme le rouges, attirent plus l'oeil.

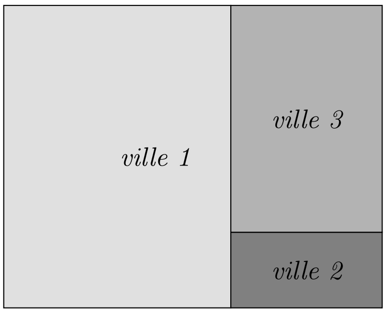
Exercice 1 *Que pensez vous de la représentation de la figure 1.2*[2] *?*

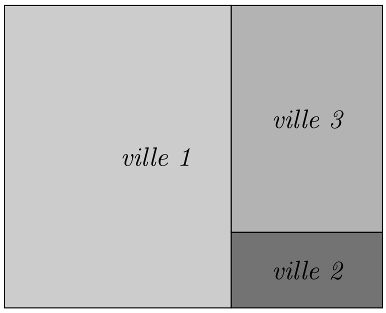
Exercice 2 *Ce problème de lien entre une valeur et l'aire représentée est aussi délicat à gérer sur les cartes. On considère par exemple trois communes disposées comme sur la carte ci-dessous et gérée par le même commissariat. Plus la couleur est foncée, plus le nombre de cambriolages est important. Le dessin de gauche représente la situation en 2000 et celle de droite en 2010.*



1. *Peut-on dire que la situation s'est améliorée ? Comment faudrait il griser la carte pour que cela soit visuellement pertinent ?*
2. *On considère maintenant la figure ci-dessous où plus une ville est foncée, plus il y a de cambriolage par hectare de la ville, encore une fois en 2000 et 2010. Cela permet il de savoir où il est préférable d'habiter si l'on craint les cambriolages ?*

2. `http://nte-serveur.univ-lyon1.fr/nte/immediato/math2002/mass11/seance06/EX00604.ht`





Exercice 3 *Que pensez vous de la courbe de la figure 3*[3].

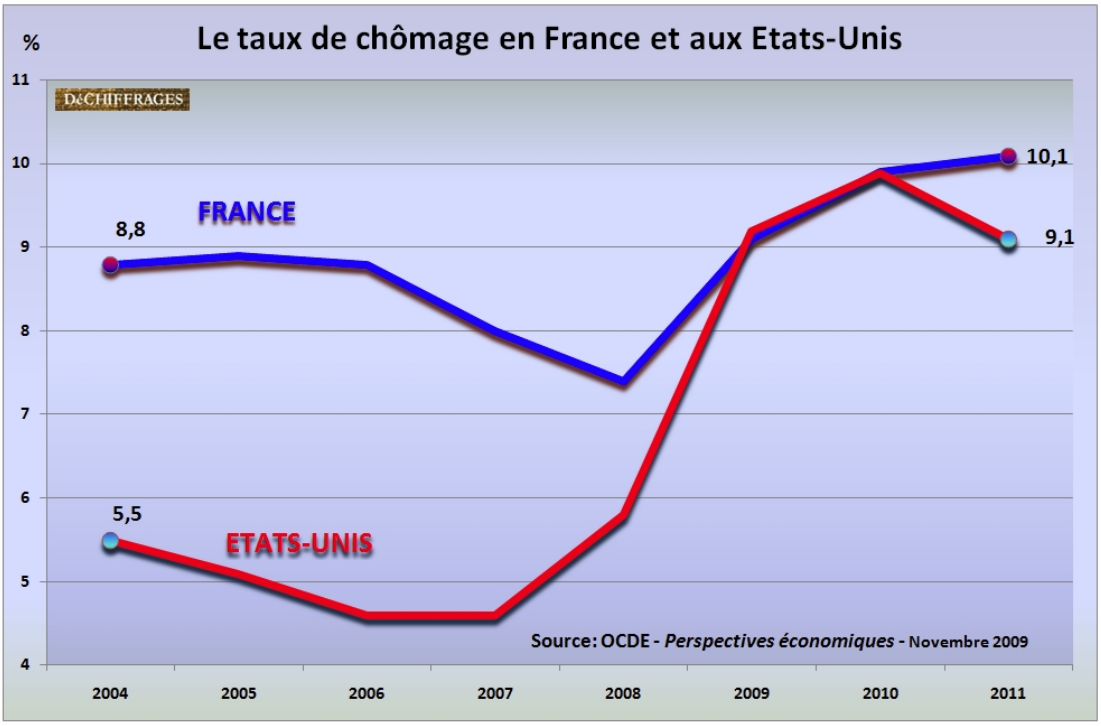


FIGURE 1.3 – Courbe du chômage

1.2 Activité 2 : paradoxe de Simpson et phénomène de Rogers

1.2.1 Paradoxe de Simpson

Exercice 4 *On considère deux lycées, notés A et B. On observe les taux de réussites suivants dans les deux lycées.*

	Lycée A	*Lycée B*
garçons	*70%*	*71.05%*
filles	*75%*	*80%*

1. *Pouvez vous dire quel lycée à le meilleur taux de réussite ?*
2. *On considère maintenant les effectifs des deux lycées donnés ci-dessous. Calculer le taux de réussite de chaque lycée. Qu'en déduire ?*

	Lycée A	*Lycée B*
garçons	*100*	*190*
filles	*100*	*10*

3. http://www.canooe.fr/blog/billet.php?url=2013-04-18-Les-objectifs-en-gestion-des-affaires-publiques#.U_NVDrsidXQ

Le texte ci-dessous vient de wikipedia [4] et illustre, sur un véritable exemple, le paradoxe se Simpson.

Un exemple réel provenant d'une étude médicale sur le succès de deux traitements contre les calculs rénaux permet de voir le paradoxe sous un autre angle.

La première table montre le succès global et le nombre de traitements pour chaque méthode.

Taux de succès (succès/total)

Traitement A	Traitement B
78 % (273/350)	83 % (289/350)

Cela semble révéler que le traitement B est plus efficace. Maintenant, en ajoutant des données concernant la taille des calculs rénaux, la comparaison prend une autre tournure :

Taux de succès (succès/total) (à gauche petits calculs, à droite gros calculs)

Traitement A	Traitement B
93 % (81/87)	87 % (234/270)

Traitement A	Traitement B
73 % (192/263)	69 % (55/80)

L'information au sujet de la taille des calculs a inversé les conclusions concernant l'efficacité de chaque traitement. Le traitement A est maintenant considéré comme plus efficace dans les deux cas. Le traitement le plus efficace peut être déterminé grâce à l'inégalité entre les deux rapports (succès/total). Le rebroussement de cette inégalité, qui conduit au paradoxe, se produit à cause de deux effets concurrents :

La variable supplémentaire (ici la taille) a un impact significatif sur les rapports. Les tailles des groupes qui sont combinés quand la variable supplémentaire est ignorée sont très différentes.

Retournons à la sécurité routière. Considérons la différence du nombre d'accidents sur un parcours Paris-Besançon entre rouler avec une petite citadine et une grosse berlines (les chiffres sont inventés).

	citadines	berlines
nombre de trajets	1300	2500
nombre d'accidents	10	11
taux d'accidents	0.76%	0.44%

Avec ces chiffres, la berlines semble beaucoup plus sûre que la citadine. Regardons cependant les chiffres ci-dessous.

	citadines	berlines
nombre de trajets sur autoroute	300	2000
nombre d'accidents sur autoroute	1	6
taux d'accidents sur autoroute	0.33%	0.3%
nombre de trajets sur nationale	1000	500
nombre d'accidents sur nationale	9	5
taux d'accidents sur nationale	0.9%	1%

En fait, les deux types de voitures semblent avoir a peu près le même taux d'accident, les premiers résultats sont biaisé par le fait que beaucoup plus de personnes roulant avec une petite citadine prennent plutôt la nationale. Mais il faudrait aussi prendre en compte les conditions de route (pluie, nuit, etc.) pour pouvoir conclure quelque chose. Connaître les variables significatives à prendre en compte pour faire des probabilité demande un expertise certaines, à la fois en statistique mais surtout sur le sujet d'étude.

Ce qu'il est important de retenir, c'est qu'il est facile (et malheureusement usuel) d'utiliser des représentations graphiques trompeuses. Par ailleurs, il est parfois difficile d'obtenir les valeurs pertinentes permettant d'en tirer des conclusions utiles. Deux exemples : il y a en France plus de commotion cérébrale dû à des accidents de la route sur des piétons que sur des vélos. Certaines associations d'utilisateur de vélos l'utilise pour que le port du casque ne soit pas rendu obligatoire. Mais ces chiffres doivent être rapportés aux volume du trafic vélo et du trafic piétons. Se pose alors la question de choisir la mesure du trafic (en temps passé ou en kilomètres parcourus ?) et de savoir le mesurer (actuellement on ne sait le faire qu'avec un facteur 100 près). Toujours sur la sécurité routière, il est par exemple aussi difficile de comparer la dangerosité des routes entre deux pays comme la France et le Royaume-Uni. On peut bien entendu compter le nombre d'accidents mortels selon des critères identiques, mais doit on rapporter ce total au nombre de véhicule ? au nombre d'habitants ? au nombre de kilomètres parcourus ? au nombre de kilomètres de route ? Par ailleurs, le trafic n'est pas du tout le même en France (pays de transits entre l'Europe du Nord et du Sud) et le Royaume-Uni qui est une île. De même, on sait que les conditions météorologiques influencent sensiblement le nombre d'accidents. Comment le prendre en compte ? En conclusion, il faut en statistique avoir les bonnes données (cela demande une expertise métier en général), en nombre suffisant, et savoir restituer ces résultats convenablement.

4. `http://fr.wikipedia.org/wiki/Paradoxe_de_Simpson`

1.2.2 Phénomène de Rogers

Le phénomène de Rogers peut traduire qu'il est possible, avec de même données, de donner des résultats aux conclusion différentes. Cela montre qu'il est possible de truquer des résultat sans pour autant truquer des données, par un moyen purement mathématique. Imaginons un IUT dans lequel il y a deux groupes de niveaux A et B. Les étudiants du groupe A sont, en général, meilleur que ceux du groupe B. Une année, les groupes A a eu 14.2 de moyenne et le groupe B a eu 8 de moyenne.

La seconde année, les notes du groupe A sont : 16, 15, 15, 13, 11. les notes du groupe B sont 13, 8, 6 et 3. La moyenne du groupe A est de 14 et celle du groupe B de 7.5. Les résultats semblent donc moins bon que l'année précédente. L'enseignant décide alors de faire passer le plus mauvais étudiant du groupe A dans le groupe B. Les notes du groupe A deviennent alors 16, 15, 15 et 13, soit une moyenne de 14.75. Celle du groupe B sont 13, 11, 8, 6 et 3, soit une moyenne de 8.2. De cette façon, les moyennes des groupes A et B ont augmenté !

Cela montre qu'il faut faire attention avec les moyennes, et surtout avec les moyennes de moyennes.

1.3 Activité 3 : les dés non transitifs

Exercice 5 *On suppose que l'on a trois dés à 6 faces, appelés A, B et C. Lors d'un lancer, pour chaque dé, chaque face sort avec une probabilité $1/6$: les dés ne sont pas pipés.*

- *le dé A a sur ses faces 2,2,4,4,9,9,*
- *le dé B a pour sa part 1,1,6,6,8,8,*
- *le dé C a 3,3,5,5,7,7.*

On s'intéresse à un jeu à deux joueurs où chacun prend un dé, le lance et celui qui a le plus grand résultat gagne.

1. *Si j'ai le dé A et mon adversaire le B, quelle est la probabilité que je gagne ?*
2. *Si j'ai le dé B et mon adversaire le C, quelle est la probabilité que je gagne ?*
3. *Si j'ai le dé C et mon adversaire le A, quelle est la probabilité que je gagne ?*
4. *Quel est le dé le plus fort ?*

1.4 Activité 4 : le MontyHall

Exercice 6 *Un prisonnier est face à trois portes et il doit en choisir une. L'une mène à la liberté, les deux autres à la potence. Il y a à coté de la porte un juge. Le prisonnier lui indique la porte qu'il désire prendre. A ce moment, avant que le prisonnier ne bouge, le juge ouvre l'une des deux autres portes en montrant au prisonnier qu'elle conduit à la potence. Le prisonnier peut alors partir par la porte choisie au départ ou prendre l'autre porte encore fermée. Quelle est pour lui la meilleure stratégie pour retrouver la liberté ? Doit-il garder son choix initial, changer, ou peu importe ?*

Chapitre 2

Introduction et rappels sur les probabilités

2.1 Espace de probabilités

2.1.1 Univers et évènements

Dans tout ce chapitre Ω désigne un ensemble (fini ou infini) appelé **univers**. Les éléments de Ω sont appelés **issues** ou **éventualités** et les sous-ensembles de Ω des évènements.

Considérons par exemple $\Omega = \{\text{pile}, \text{face}\}$ est un univers à deux issues/éléments. Les évènements sont $\emptyset$, $\{\text{pile}\}$, $\{\text{face}\}$ et Ω.

On peut aussi prendre $\Omega = \{1, \ldots, 6\}$. Un évènement possible est alors *un nombre pair*, qui correspond au sous-ensemble $\{2, 4, 6\}$.

Comme les évènements sont des ensembles, on les combine souvent en utilisant les opérateurs booléens classiques sur les ensembles : union, intersection, complément.

Exercice 7 *On considère un univers Ω constitués d'étudiants de l'IUT. On considère les évènements suivants :*

— *A : Étudiants garçons.*
— *B : Étudiantes.*
— *C : En première année.*
— *D : En seconde année.*
— *E : En licence pro.*
— *F : Au département informatique.*
— *G : Au département carrières sociales.*

1. *Exprimer à l'aide des évènement ci-dessus et d'opération ensembliste, l'évènement* Une étudiante de licence pro.
2. *Même question avec* un ou une étudiante de première année.
3. *Même question avec* Une étudiante inscrite à la fois en première et seconde année.
4. *Même question avec* Une étudiante de DUT informatique ou de carrières sociales.

Exercice 8 *On considère un univers Ω constitués des 32 cartes classiques d'un jeu. On considère les évènements suivant :*

— *A : Rois*
— *B : Figures.*
— *C : Pics.*
— *D : Trèfle.*

Exprimer en français à quoi correspondent les évènements suivants :

1. $A \cup C$.
2. $A \cap C$.
3. $B \cap (C \cup D) \cap \overline{A}$.

2.1.2 Probabilités discrètes

Definition 9 *Une* **probabilité** *sur un univers Ω est une application $\mathbb{P}$ qui à chaque évènement associe un réel entre dans $[0, 1]$ telle que :*

— $\mathbb{P}(\Omega) = 1$,
— *Si $(A_i)_{i \in I}$ est une famille finie ou dénombrable (c'est-à-dire dont on peut compter les éléments) d'évènements deux à deux disjoints (c'est-à-dire que si $i \neq j$, alors $A_i \cap A_j = \emptyset$), alors*

$$\mathbb{P}(\cup_{i \in i} A_i) = \sum_{i \in I} \mathbb{P}(A_i).$$

Une probabilité est aussi appelée une **loi de probabilité** ou une **distribution de probabilité**, ou encore simplement une **distribution**.

Le couple $(\Omega, \mathbb{P})$ est appelé **espace de probabilités** ou **espace probabilisé**. Pour un évènement A, $\mathbb{P}(A)$ est appelée la probabilité que l'évènement A se produise.

Exercice 10 *Soit $(\Omega, \mathbb{P})$ un espace de probabilités.*

1. *Que vaut $\mathbb{P}(\Omega \cup \emptyset)$? en déduire la valeur de $\mathbb{P}(\emptyset)$.*
2. *Que vaut $\mathbb{P}(A \cup \overline{A})$, où A est un évènement quelconque ? En déduire une relation entre $\mathbb{P}(A)$ et $\mathbb{P}(\overline{A})$.*
3. *On suppose que $A \subseteq B$. Que peut on dire de $\mathbb{P}(A)$ et $\mathbb{P}(B)$?*

On a les propriété. suivantes

Proposition 11 *Soit $(\Omega, \mathbb{P})$ un espace de probabilités.*
- $\mathbb{P}(\emptyset) = 0$
- *Pour tout évènement A, $\mathbb{P}(A) = 1 - \mathbb{P}(\overline{A})$.*
- *Si $A \subseteq B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$.*
- *Pour tous évènements A et B, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.*

Prenons par exemple le jeu de cartes, avec Ω l'ensemble des 32 cartes. On définit (par exemple), $\mathbb{P}(A) = \frac{|A|}{32}$ (c'est ce qu'on appelle la distribution uniforme), chaque carte a la même probabilité de sortir.
- Considérons l'évènement *Trèfle*. Il contient 8 cartes, du 7 à l'as. Sa probabilité est donc $\frac{8}{32} = 0.25$; on a une chance sur 4 de tirer un trèfle.
- Considérons l'évènement *n'est pas un Trèfle*. C'est le complémentaire de l'évènement ci-dessus, sa probabilité est donc $1 - 0.25 = 0.75$.
- Considérons l'évènement *est un Roi*. Il contient 4 cartes, les quatre rois. Sa probabilité est donc $\frac{4}{32} = \frac{1}{8}$.
- Considérons l'évènement *Roi de Trèfle*. Il n'y a qu'une carte, la probabilité est donc de $\frac{1}{32}$.
- Si l'on considère l'évènement *Roi ou Trèfle*, en utilisant la formule $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, on obtient que sa probabilité vaut $\frac{1}{8} + \frac{1}{4} - \frac{1}{32} = \frac{11}{32}$.

Exercice 12
1. *Justifier pourquoi si l'on connaît $\mathbb{P}$ (avec les notations du cours) pour les singletons (les ensembles à un seul élément), alors on connaît $\mathbb{P}$ pour tout évènement.*
2. *On considère que $\Omega = \mathbb{N}$ et que $\mathbb{P}(n) = \frac{1}{2^{n+1}}$. En utilisant les résultats vu l'année précédente sur la somme des termes d'une suite géométrique, justifier que l'on a bien $\mathbb{P}(\Omega) = 1$.*
3. *Dans l'espace de la question précédente, quelle est la probabilité de l'évènement nombre pair ?*

Definition 13 *Soit Ω un ensemble fini de cardinal n. On appelle* **distribution uniforme** *sur Ω, la probabilité qui vérifie $\mathbb{P}(\{x\}) = \frac{1}{n}$ pour tout x dans Ω.*

Avec la distribution uniforme, lors d'une expérience, chaque issue à la même chance d'être tirée au sort. C'est le cas classique du dé ou de la pièce non truquée.

Soit $\mathbb{P}$ une loi uniforme sur Ω fini de cardinal n. Alors, pour tout évènement A, on a $\mathbb{P}(A) = \dfrac{|A|}{n}$.

2.2 Probabilités discrètes uniformes et combinatoire

Exercice 14 *On considère un sac contenant dix balles numérotées de 1 à 10, deux sont rouges, et huit sont bleues (les couleurs ne sont pas utilisées dans cette exercice, mais le seront plsu tard).*

1. *On considère l'expérience qui consiste à tirer au sort une balle sans regarder. Quel est l'univers ?*
2. *Cette fois ci l'expérience consiste à tirer une balle, puis la remettre dans le paquet, puis après avoir mélanger, à retirer une balle. Quel est l'univers ?*
3. *Maintenant, on tire une balle, puis une autre sans remettre la première dans le sac. Quel est l'univers ?*
4. *On tire maintenant deux balles à la fois. Quel est l'univers ?*
5. *Enfin, même question avec : on tire une balle, on la remet, puis on tire deux balles à là fois.*

On voit qu'il n'est pas toujours simple d'exprimer l'univers lorsque l'on fait plusieurs manipulations. Dans le cas de la loi uniforme, on peut utiliser la formule $\mathbb{P}(A) = \dfrac{|A|}{n}$ vue à la fin de la section précédente, pour transformer une problème de probabilité en un problème de comptage : il suffit ce compter le nombre d'éléments de l'univers et le nombre d'élément de l'univers. Pour cela, on utilise différentes techniques combinatoires dont les plus simples ont été vues au Lycée (et que nous allons revoir).

Definition 15 *Soit n et k deux entiers positifs vérifiant $k \leq n$. On a*
- *Factoriel n est défini par : $n! = n(n-1)(n-2)\dots 2.1$,*

— *Le nombre d'arrangements de k objets parmi n est*

$$A_n^k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}.$$

— *Le nombre de combinaisons de k objets parmi n est*

$$C_n^k = \frac{n(n-1)\dots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}.$$

Proposition 16 *Soit n et k deux entiers positifs vérifiant $k \leq n$.*
— *Il y a $n!$ façon d'ordonner les n objets.*
— *Il y a A_n^k façon d'ordonner k objets parmi les n objets.*
— *Il y a C_n^k façon de choisir k objets parmi les n objets.*
Par ailleurs, si A et B sont deux ensembles finis, alors $|A \times B| = |A|.|B|$.

Exercice 17 *Un classe comporte 14 étudiants.*

1. *On souhaite choisir deux délégués pour cette classe. Combien y'a-t-il de possibilités ?*
2. *On souhaite choisir un délégué et un délégué suppléant pour cette classe. Combien y'a-t-il de possibilités ?*
3. *Une salle machine comporte 14 places, combien y'a-t-il de façon d'y placer les étudiants ?*
4. *Une salle machine comporte 16 places, combien y'a-t-il de façon d'y placer les étudiants ?*
5. *On souhaite effectué des projets par binômes. Combien de façon y-a-t-il de constituer 7 binômes ?*

Exercice 18 *On considère un groupe de 23 personnes séléctionnée pour faire partie d'une équipe de football, qui contient 11 joureurs.*

1. *Combien d'équipes différentes peut-on faire ?*
2. *Sachant qu'il faut un gardien par équipe et qu'il y en a 3 parmi les 23 personnes, combien y'a-t-il de possibilités de constituer une équipe ?*
3. *Parmi les 23 joueurs, il y a 3 gardiens, 6 attaquants, 6 milieux de terrain et 8 arrières ? Un équipe est composée d'un gardien et 4 défenseurs. Il y a ensuite soit 3 milieux et 3 attaquants, soit 4 milieux et 2 attaquants. Combien y'a-t-il de possibilités de constituer une équipe ?*

Reprenons l'exemple de l'exercice 14 et calculons la taille de l'univers à chaque fois.

1. La taille de l'univers est 10.
2. Ici, on tire deux balles avec remises, la taille de l'univers est donc $10.10 = 100$.
3. Ici on choisit, de façon ordonné, deux balles parmi 10. Il y a donc $A_{10}^2 = 10.9 = 90$ possibilités.
4. Ici on choisit, sans ordre, deux éléments. Il y a donc $C_{10}^2 = (10.9)/2 = 45$ possibilités.
5. Ici on choisit d'abord une balle, puis on la remets, puis on en tire 2. Pour la première balle il y a 10 possibilités, puis pour les deux autres 45. Au total il y a 450 possibilités.

Exercice 19 *On considère l'évènement A : avoir au moins une balle rouge. Pour chaque cas de l'exercice 14, calculer la probabilité de A (on pourra décomposer A en évènements disjoints). Même question avec l'évènement : avoir exactement une balle bleue.*

Exercice 20 *On considère un jeu de 32 cartes classique. Une main est un ensemble (non ordonné) de cinq cartes du jeu.*

1. *Combien y a-t-il de mains possibles ?*
2. *Quelle est la probabilité qu'une main contiennent les 4 as ?*
3. *Quelle est la probabilité qu'une main contiennent l'as de pique ?*
4. *Quelle est la probabilité qu'une main contiennent au moins un as ?*

2.3 Indépendance et probabilités conditionnelles

2.3.1 Évènements indépendants

Definition 21 *Deux évènements A et B sont* **indépendants** *si $\mathbb{P}(A \cap B) = P(A)P(B)$.*

Par exemple, si l'on lance deux fois de suite un dé à 6 faces et que l'on considère les trois évènements A : *le premier lancé est pair*, B : *le second lancé est pair* et C : *le premier lancé est 4*. On a $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(B) = \frac{1}{2}$ et $\mathbb{P}(C) = \frac{1}{6}$. Par ailleurs, $\mathbb{P}(A \cap B) = \frac{1}{4}$, donc A et B sont indépendants. Mais $\mathbb{P}(A \cap C) = \frac{1}{6}$, donc A et C ne sont pas indépendants.

Exercice 22 *En utilisant la proposition 11, si A et B sont indépendants, que vaut $\mathbb{P}(A \cup \mathcal{B})$? En lançant deux dés à la suite, quelle est la probabilité que l'un des deux au moins soit pair ?*

2.3.2 Probabilités conditionnelles et Formule de Bayes

Definition 23 *Soient A et B deux évènements tels que $\mathbb{P}(B) \neq 0$. On définit la probabilité de A sachant B, noté $\mathbb{P}(A \mid B)$, par :*

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

Prenons par exemple un dé à 6 face que l'on lance une fois, avec les évènements A : *nombre pair* et B : *on obtient un* 4. On a $\mathbb{P}(A \mid B)$ qui est la probabilité d'obtenir un nombre pair si l'on a obtenu un 4. Sans surprise $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{1/6}{1/6} = 1$. Maintenant $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B\cap A)}{\mathbb{P}(A)} = \frac{1/6}{1/2} = \frac{1}{3}$. La probabilité d'obtenir un 4 si l'on sait que l'on a obtenu un nombre pair est de une sur trois.

Exercice 24 *On considère une classe, dans laquelle il y a des filles et des garçons, avec ou sans lunettes. On sait que 20% de cette classe exactement sont des filles avec des lunettes. On sait aussi qu'une personne sur deux porte des lunettes. Si on tire au sort une personne qui a des lunettes, quelle est la probabilité que cela soit une fille ? (on pourra utiliser l'évènement A être une fille et l'évènement B avoir des lunettes).*

Exercice 25
1. *Soient A et B deux évènements tels que $\mathbb{P}(A) \neq 0$ et $\mathbb{P}(B) \neq 0$. Exprimer $\mathbb{P}(B \mid A)$ en fonction de $\mathbb{P}(A \mid B)$, de $\mathbb{P}(A)$ et de $\mathbb{P}(B)$.*
2. *On considère une salle contenant des individus, fille ou garçon, étudiant ou enseignant. Il y a 10% d'enseignants et, parmi les étudiants, 20% de filles. Si l'on choisit une fille au sort, quelle est la probabilité qu'elle soit une étudiante ?*

Exercice 26 *On considère un sac contenant des objets qui peuvent être bleus ou rouges, cubiques ou sphériques. On sait qu'il y a un tiers d'objets cubiques et un quart d'objets rouges. Par ailleurs, s'il l'on choisit un objet cubique, il y a une chance sur 3 qu'il soit rouge. Si l'on choisit un objet rouge, quelle est la probabilité qu'il soit cubique ?*

Exercice 27 *On suppose que la probabilité que l'anniversaire de quelqu'un tombe un mois donnée est de $\frac{1}{12}$.*
1. *Deux personnes se rencontrent, quelle est la probabilité qu'elles soient nées le même mois ?*
2. *Trois personnes se rencontrent, quelle est la probabilité qu'elles soient nées des mois tous différents ?*
3. *Même question avec cinq personnes.*

Definition 28 *Soient A_1, A_2, …, A_k des évènements. On dit qu'ils forment une partition de l'univers s'ils sont tous non vides et qu'ils sont deux-à-deux disjoints ($A_i \cap A_j = \emptyset$ si $i \neq j$) et qu'ils recouvrent tout l'univers ($A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$).*

Considérons le tirage au sort d'une carte dans un jeu classique de 32 cartes. Les évènements *tirer un coeur*, *tirer un trèfle*, *tirer un carreau* et *tirer un pique* forment une partition de l'univers : chaque carte à au moins une couleur (recouvrement), au plus une couleur (deux-à-deux disjoints) et il y a au moins une carte dans chaque couleur (évènement non vide). Notons qu'ici, avec une tirage non biaisé les évènements sont équiprobables, mais ça n'est pas demandé par la définition. Par exemple, les deux évènements *tirer une figure* et *tirer un nombre* forment aussi une partition de l'univers.

Exercice 29
On considère l'univers $\Omega = \{1, 2, 3, 4, 5\}$.
1. Est-ce que $\{1, 2\}$, $\{3, 4\}$ forme une partition ? Pourquoi ?
2. Est-ce que $\{1, 2, 3\}$, $\{3, 4, 5\}$ forme une partition ? Pourquoi ?
3. Est-ce que $\{1, 2, 3\}$, $\{4, 5\}$ forme une partition ? Pourquoi ?
4. Donner deux exemples de partitions.

Le résultat suivant s'appelle formule des probabilités totales.

Proposition 30 *Soient A_1, A_2, …, A_k des évènements formant une partition des évènements de l'univers. Pour tout événement B, on a*

$$\mathbb{P}(B) = \mathbb{P}(B \cap A_1) + \ldots \mathbb{P}(B \cap A_k),$$

ce qui s'écrit encore

$$\mathbb{P}(B) = \mathbb{P}(B \mid A_1)\mathbb{P}(A_1) + \ldots \mathbb{P}(B \mid A_k)\mathbb{P}(A_k).$$

Un frigidaire contient des yaourts qui sont à la fraise, à la poire ou à l'abricot. Certains yaourts sont mixés, les autres contiennent des morceaux de fruits. On sait qu'un tiers des yaourts à la fraise sont mixés, un quart de ceux à la poire le sont et la moitiés de ceux à l'abricot. La moitié des yaourts sont à la fraise et un quart à la poire. Quelle est la probabilité, si l'on prend un yaourt au hasard, d'avoir un yaourt mixé ? On pose F, P et A les évènements pour des yaourts d'être *à la fraise*, *à la poire* et *l'abricot*. On note M la probabilité qu'un yaourt soit mixé. On a, d'après la formule des probabilités totales :

$$\begin{aligned}\mathbb{P}(M) &= \mathbb{P}(M \mid F)\mathbb{P}(F) + \mathbb{P}(M \mid P)\mathbb{P}(P) + \mathbb{P}(M \mid A)\mathbb{P}(A) \\ &= \frac{1}{3}\frac{1}{2} + \frac{1}{4}\frac{1}{4} + \frac{1}{2}\frac{1}{4} = \frac{1}{6} + \frac{1}{16} + \frac{1}{8} = \frac{17}{48} \simeq 0.354\end{aligned}$$

Environ 35% des yaourts sont mixés.

On peut réécrire la forumle des probabilités totale sous une autre forme, appelée *formule de Bayes.*

Proposition 31 *Soient A_1, A_2, ..., A_k des évènements formant une partition des évènements de l'univers et B un évènement quelconque. On a*

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}{\mathbb{P}(B \mid A_1)\mathbb{P}(A_1) + \ldots \mathbb{P}(B \mid A_k)\mathbb{P}(A_k)}.$$

En reprenant l'exemple ci-dessus des yaourts, la probabilité qu'un yaourt soit à la fraise, sachant que c'est un yaourt mixé vaut

$$\mathbb{P}(F \mid M) = \frac{\mathbb{P}(M \mid F)\mathbb{P}(F)}{\frac{17}{48}} = \frac{\frac{1}{6}}{\frac{17}{48}} = \frac{8}{17}.$$

Exercice 32 *En reprenant les données de l'exemple des yaourts, calculer la probabilité qu'un yaourt soit à l'abricot sachant qu'il est mixé.*

Exercice 33 *On considère une population qui contient 20% d'enfants, 40% d'hommes adultes et 40% de femmes adultes. On observe que 10% des enfants sont malades, 15% des femmes adultes et 20% des hommes adultes.*

1. *On définit les évènements E être un enfant, F être une femme adulte, H être un homme adulte et M être malade. Lesquels forment une partition ?*
2. *Donner les valeurs de $\mathbb{P}(E)$, $\mathbb{P}(H)$, $\mathbb{P}(F)$, $\mathbb{P}(M \mid E)$, $\mathbb{P}(M \mid H)$, $\mathbb{P}(M \mid F)$.*
3. *Que vaut $\mathbb{P}(M)$?*
4. *Que valent $\mathbb{P}(E \mid M)$, $\mathbb{P}(F \mid M)$ et $\mathbb{P}(H \mid M)$?*

Exercice 34 *On considère un programme écrit en commun par Alice et Bob. Alice a écrit 60 % du programme et Bob le reste. De plus 5% des lignes écrites par Alice ont un beug, et 10% de celle écrites par Bob. Si l'on tire au sort un ligne contenant une erreure, quelle est la probabilité qu'elle ait été écrite par Bob ?*

Exercice 35 *On considère un groupe de sportifs, contenant des hommes (événement H) et des femmes (événement F), des joueurs/joueuses de tennis (événement T), des joueurs/joueuses de basket (événement B) et des pratiquant(e)s du judo (événement J). On suppose que chacun pratique un et un seul de ces trois sport. On choisit au sort uniformement une personne du groupe.*

1. *On suppose que $\mathbb{P}(J) = \frac{13}{36}$ et $\mathbb{P}(T) = \frac{5}{12}$. Que vaut $\mathbb{P}(B)$?*
2. *On sait que $\frac{3}{13}$-ième des judokas sont des hommes, un tiers des joueurs de tennis sont des hommes et qu'il y a autant de joueurs de basket que de joueuses de basket. Que vaut $\mathbb{P}(B \mid H)$?*

2.4 Variables aléatoires, espérance, variance

2.4.1 Variable aléatoire

Une **variable aléatoire** est une application de Ω dans $\mathbb{R}$.

Par exemple si Ω est un jeu de 32 cartes, on peut définir la variable aléatoire Y (des points à la belotte) qui

— Associe 20 au valet de pique,
— Associe 14 au 9 de pique,
— Associe 11 à chaque as et 10 à chaque 10,
— Associe 4 à chaque roi, 3 à chaque dame et 2 aux valets (autre que pique),
— Associe 0 aux autres cartes.

Soit $\omega \in \mathbb{R}$ et X une variable aléatoire. On définit $\mathbb{P}(X = \omega)$ comme étant $\mathbb{P}(\{x \mid X(x) = \omega\})$. On a donc

$$\mathbb{P}(X = \omega) = \mathbb{P}(\{x \mid X(x) = \omega\}) = \sum_{s \in \Omega, X(s) = \omega} \mathbb{P}(s).$$

Considérons par exemple $\Omega = \{pile, face\}$ et la variable aléatoire $X(pile) = 1$, $X(face) = 2$. On a $\mathbb{P}(X = 1) = \mathbb{P}(\{x \mid X(x) = 1\}) = \mathbb{P}(pile) = \frac{1}{2}$. On a aussi $\mathbb{P}(X = 2) = \mathbb{P}(\{x \mid X(x) = 2\}) = \mathbb{P}(face) = \frac{1}{2}$. Par ailleurs, $\mathbb{P}(X = 3) = \mathbb{P}(\{x \mid X(x) = 3\}) = \mathbb{P}(\emptyset) = 0$.

Exercice 36 *Sur le jeu de la belote, que valent* $\mathbb{P}(Y=20)$, $\mathbb{P}(Y=14)$, $\mathbb{P}(Y=10)$, $\mathbb{P}(Y=2)$ *et* $\mathbb{P}(Y=0)$ *?*

Deux variables aléatoires X et Y sont **indépendantes** si pour tout x et tout y,

$$\mathbb{P}((X=x)\cap(Y=y))=\mathbb{P}(X=x).\mathbb{P}(Y=y).$$

Considérons par exemple $\Omega=\{pile, face\}$ et la variable aléatoire $X(pile)=1$, $X(face)=2$ et la variable $Y(pile)=0$ et $Y(face)=1$. On a $\mathbb{P}((X=1)\cap(Y=1))=0$, car il faut à la fois que le résutat soit pile et face. Par ailleurs $\mathbb{P}((X=1)=\mathbb{P}((Y=1)=\frac{1}{2}$, les deux variables aléatoires ne sont pas indépendantes.

Considérons maintenant le lancer de deux dés, un bleu et un rouge et les variables aléatoires X et Y définies par : $X=1$ si le dé bleu à un résultat pair et 0 sinon; $Y=1$ si le dé rouge à un résultat pair et 0 sinon. On a $\mathbb{P}((X=1)\cap(Y=1))=\frac{1}{4}=\mathbb{P}(X=1)\mathbb{P}(Y=1)$, $\mathbb{P}((X=0)\cap(Y=1))=\frac{1}{4}=\mathbb{P}(X=0)\mathbb{P}(Y=1)$, $\mathbb{P}((X=0)\cap(Y=0))=\frac{1}{4}=\mathbb{P}(X=0)\mathbb{P}(Y=0)$, $\mathbb{P}((X=1)\cap(Y=0))=\frac{1}{4}=\mathbb{P}(X=1)\mathbb{P}(Y=0)$, les deux variables aléatoires sont donc indépendantes.

Intuitivement, deux variables aléatoires sont indépendantes si leur résultats dépendent du résultat de processus aléatoires indépendants : dans l'exemple ci-dessus, X dépend du dé bleu et Y du dé rouge, deux dés dont les résultats sont indépendants.

2.4.2 Espérance

Soit $\mathbb{P}$ une probabilité sur Ω. On définit l'*espérance* (qui est la notion duale de la moyenne en statistiques) relative à $\mathbb{P}$ d'une variable aléatoire X, notée $E_{\mathbb{P}}[X]$ comme

$$E_{\mathbb{P}}[X]=\sum_{\omega}\omega\mathbb{P}(X=\omega).$$

Considérons par exemple $\Omega=\{pile, face\}$ et la variable aléatoire $X(pile)=1$, $X(face)=2$. Avec la distribution uniforme, on a $E[X]=\frac{1}{2}.1+\frac{1}{2}2=\frac{3}{2}$.

Exercice 37 *Calculer l'espérance des points à la belote pour la distribution uniforme.*

Exercice 38 *On considère* $\Omega=\{1,\ldots,6\}\times\{1,\ldots,6\}$ *avec equiprobabilité, et la variable aléatoire* $X((x,y))=x+y$ *(cela correspond au résultat du lancer de deux dés à six faces distincts). Calculer l'espérance de* X*. On considère maintenant la variable aléatoire* $Y((x,y))=\max\{x,y\}$ *(cela correspond à garder la valeur la plus forte des deux dés). Quelle l'espérance de* Y *?*

Proposition 39 *Soient* X *et* Y *deux variables aléatoires d'espérance finie, on a*

$$E[X+Y]=E[X]+E[Y].$$

Exercice 40 *En utilisant la proposition 39, reprendre la première question de l'exercice 38.*

Proposition 41 *Soit* X *une variable aléatoire et* $c\in\mathbb{R}$*, on a :* $E[cX]=cE[X]$.

Si on multiplie par 3 le résultats d'un dé, en moyenne on obtiendra trois fois la moyenne du dé, ce qui est assez intuitif.

Exercice 42 *On considère* $\Omega=\{1,\ldots,6\}$ *avec equiprobabilité et la variable aléatoire* $X(x)=x$ *(on prend la valeur du dé).*

1. *Que vaut* $E[X]^2$ *?*
2. *Que vaut* $E[X^2]$ *?*
3. *Comparer les deux.*

Proposition 43 *Soit* X *une variable aléatoire, on a :* $E[X^2]\geq(E[X])^2$.

Dans un cadre intuitif, l'inégalité ci-dessus peut se dire *la moyenne des carrés et supérieure au carré de la moyenne.* L'inégalité ci-dessous est connue sous le nom d'inégalité de Markov.

Proposition 44 *Soit* X *une variable aléatoire positive, pour tout* $a>0$*, on a*

$$\mathbb{P}(X\geq a)\leq\frac{E[X]}{a}.$$

Considérons par exemple un dé à 6 faces et la variable aléatoire associée (valeur de la face). L'inégalité de Markov donne, pour $a = 4$,

$$\mathbb{P}(X \geq 4) \leq \frac{E[X]}{4} = \frac{7/2}{4} = 7/8.$$

Considérons maintenant le lancé successif de n dés à 6 faces, chaque variable aléatoire associée est X_i. On pose $X = \sum X_i$, qui est la somme des résultats des dés. On a

$$\mathbb{P}(X \geq 5n) \leq \frac{E[X]}{5n} = \frac{E[\sum X_i]}{5n} = \frac{\sum E[X_i]}{5n} = \frac{7n/2}{5n} = \frac{7}{10}.$$

Il y a moins de sept chances sur 10 que la somme de n dé dépasse $5n$.

Exercice 45 *En reprenant l'exemple du jeu de Belote, appliquer l'inégalité de Markov pour majorer* $\mathbb{P}(X \geq 10)$*. Que vaut cette probabilité exactement ; comparer.*

Exercice 46 *On considère n lancés de dés (à 6 faces) et, à chaque fois une variable aléatoire X_i qui vaut* 1 *si le résultat du dé est 6 et* 0 *sinon. On considère aussi la variable aléatoire X du nombre de dés qui ont fait* 6.

1. *Exprimer X en fonction des X_i.*
2. *Que vaut $E[X_i]$? En déduire $E[X]$.*
3. *utiliser l'inégalité de Markov pour majorer la probabilité que 3/4 des dés aient fait un* 6.

Si X et Y sont deux variables aléatoires indépendantes, alors

$$E[XY] = E[X]E[Y].$$

2.4.3 Variance

Deux variables aléatoires de même espérance peuvent avoir des "comportements" très différents. Afin mieux comprendre et décrire une variable aléatoire nous allons introduire la notion de *variance*.

Definition 47 *Soit X une variable aléatoire et $(\Omega, \mathbb{P})$ un espace de probabilités. On définit la* variance *de X, notée* $\mathrm{var}_{\mathbb{P}}(X)$ *comme :*

$$\mathrm{var}_{\mathbb{P}}(X) = E_{\mathbb{P}}[(X - E_{\mathbb{P}}[X])^2 = E_{\mathbb{P}}[X^2] - (E_{\mathbb{P}}[X])^2.$$

La **déviation standard** *de X, notée $\sigma(X)$ est définie par*

$$\sigma(x) = \sqrt{\mathrm{var}(X)}.$$

On considère à nouveau un dé à 6 faces (numérotées de 1 à 6) avec un lancé equiprobable. On a $E[X] = \frac{7}{2}$. Par ailleurs $E[X^2] = \frac{1}{6}(1+4+9+16+26+36) = 15$. On a donc $\mathrm{var}(X) = 15 - \frac{49}{4} = \frac{11}{4} = 2.75$. Par ailleurs $\sigma(X) \simeq 1.66$.

Exercice 48 *Calculer la variance et la déviation standard dans le cas de l'exemple du jeu de Belote.*

Proposition 49 *Si X et Y sont deux variables aléatoires indépendantes, alors* $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)$.

Le résultat suivant est connu sous le nom d'*Inégalité de Chebyshev.*

Proposition 50 *Soit X une variable aléatoire. Pour tout $a > 0$,*

$$\mathbb{P}(|X - E[X]| \geq a) \leq \frac{\mathrm{var}(X)}{a^2}.$$

Cette inégalité permet de majorer la probabilité qu'une variable aléatoire s'écarte fortement de la moyenne.

Exercice 51 *On reprend les données de l'Exercice 46.*

1. *Que vaut* $\mathrm{var}(\mathrm{X_i})$ *?*
2. *Que vaut* $\mathrm{var}(\mathrm{X})$ *(le variables sont indépendantes) ?*
3. *Justifier que $X \geq \frac{3n}{4}$ ssi $|X - \frac{n}{6}| \geq \frac{7n}{12}$.*
4. *En utilisant l'inégalité de Chebychev montrer que $\mathbb{P}(X \geq 3n/4) \leq \frac{0.41}{n}$.*

En lançant 20 dés, la probabilité qu'il y en ait au moins 15 qui fassent un 6 *est inférieure à environ* 0.02.

Chapitre 3

Lois de probabilités classiques

Nous allons voir dans ce chapitre différentes lois de probabilités classiques. Pour le moment, la seul qui a été vue est la loi uniforme.

3.1 Loi de Bernouilli et loi Binomiale

Soit Ω un univers. On appelle **variable aléatoire de Bernouilli** une variable aélatoire X à valeur dans $\{0, 1\}$. La valeur $p = X^{-1}(\{1\})$, c'est-à-dire la probabilité que $X(s)$ vaille 1, est appellé **paramètre de la variable aléatoire de Bernouilli**.

Si on note Y la variable aléatoire qui vaut 1 si une pièce lancée tombe sur pile et 0 sinon est une variable aléatoire de Bernouilli, de paramètre $\frac{1}{2}$ si la pièce est non biaisée.

Les variables aléatoires de Bernouilli caractérisent le succès d'une expérience aléatoire.

Exercice 52
1. *Que vaut l'espérance d'une varaible aléatoire de Bernouilli de paramètre p ?*
2. *Que vaut la variance d'une variable aléatoire de Bernouilli de paramètre p ?*

Exercice 53 *On considère une pièce de manaie biasiée, c'est-à-dire qu'elle ressort pile avec une probabilité p. On suppose $p \neq 0$, mais que p est inconnue.*
1. *On considère l'expérience consistant à lancer deux fois de suite la pièce. Quelle est l'univers de cette expérience ?*
2. *Quelle est la probabilité, dans cette expérience, d'obtenir $(pile, pile)$? d'obtenir $(pile, face)$? d'obtenir $(face, pile)$? d'obtenir $(face, face)$?*
3. *On considère la procédure qui consiste à répéter cette expérience jusqu'à obtenir $(pile, face)$ ou $(face, pile)$. Dans le premier cas on retourne pile, dans le second, on retourne face. Si cette procédure s'arrete, qu'elle est la probabilité d'obtenir pile ? face ?*
4. *Quelle est la probabilité (en fonction de p) que la procédure ne s'arrete pas au bout de 10 tentatives ?*
5. *En vous aidant de votre calculatrice, quelle est la probabilité que la procédure ne se soient pas arrétée au bout de 10 coups si $p = 0.51$? si $p = 0,8$?*

Une **variable aléatoire binomiale** de paramètre p et n est variable aléatoire qui compte le nombre de succès sur n expériences indépendantes de Bernouilli de paramètre p. Par exemple, si on lance une pièce avec succès quand pile sort, $X = k$ si sur les n lancés, on a obtenus k fois pile. Plus formellement, la variable aléatoire de Bernouilli de paramètres p et n, notée $B(n, p)$ est définie par

$$\mathbb{P}(X = j) = C_n^j p^j (1 - p)^{n-j}.$$

Exercice 54 *On lance 20 fois une pièce de monaie équilibrée. Quelle est la probabilité d'obtenir exatement 1 fois pile ? 10 fois pile ?*

Exercice 55 *On considère un sac contenant 3 balles bleues et 2 balles rouges. On tire, avec remise et indépendamment 10 fois une balle du sac.*
1. *Quelle est la probabilité d'avoir tirer exactement 4 balle bleues ?*
2. *Quelle est la probabilité d'avoir tirer autant de balles bleues que de balles rouges ?*

En moyenne une expérience de Bernouilli de paramètre p exécutée n fois réussit np fois, ce qui s'écrit :

Proposition 56 *On a $E[B(n, p)] = np$.*

Par ailleurs, après quelques calculs, on peut montrer le résultat suivant :

Proposition 57 *On a $\mathrm{var}[B(n, p)] = np(1 - p)$.*

Exercice 58 *En utilisant la proposition 57 et l'inégalité de Tchebyshev, majorer la probabilité que lors de n lancés indépendants d'une pièce equilibrée, plus des 3/4 donnent pile.*

3.2 Loi géométrique

Une **variable aléatoire géometrique de paramètre** p est une variable qui compte le nombre d'essais avant de réussir une expérience de Bernouilli de parmaètre p. Plus formellement, c'est une variable aléatoire qui vérifie

$$\mathbb{P}(X = n) = (1 - p)^{n-1}p.$$

La probabilité qu'il faille n coups pour réussir l'expérience est la probabilité qu'elle est ratée $n - 1$ fois (ce qui donne le $(1 - p)^{n-1}$) fois la probabilité qu'elle réussisse au n-ième coup, ce qui donne le p. Par exemple, la loi géométrique de raison $\frac{1}{6}$ compte le nimbre de fois qu'il faut lancer un dé équilibré pour obtenir un 6.

Exercice 59 *On va vérifier dans cet exercice que la définition ci-dessus d'une variable aléatoire géométrique satisfait bien que la probabilité de l'univers est* 1*. On suppose* $0 < p < 1$*.*

1. *Ecire sous forme d'une somme* $\mathbb{P}(X \leq n)$*.*
2. *Dans la somme ci-dessus, factoriser par* p*.*
3. *En posant* $q = 1 - p$*, reconnaitre la somme des terme d'une suite géométrique dans la formule obtenue à la question précédente.*
4. *En déduire un formule pour* $\mathbb{P}(X \leq n)$*. Que devient-elle lorsque* n *tend vers l'infini.*

Proposition 60 *Soit* X *une variable aléatoire géométrique de paramètre* $p \neq 0$*. On a*

$$E[X] = \frac{1}{p}.$$

Exercice 61 *On lance plusieurs fois un dé à 6 face. Combien faut-il lancer de fois pour espérer obtenir un 6 ?*

Exercice 62 *On s'interesse ici au problème du collectionneur. Supposons qu'un enfant mange des céréales et que dans chaque paquet il y a une carte représentant un pokemon. On suppose qu'il n'est pas possible, en achetant une boite, de savoir quelle carte il y a dedans et que toutes les cartes apparaissent avec la même probabilité dans un paquet. La question est combien de paquet il devra acheter pour avoir les* n *cartes différentes.*

1. *On pose* X_i *(pour* $1 \leq i \leq n$*), la variable aléatoire qui compte le nombre de paquets achetés la première fois que l'enfant obtient exactement* i *cartes différentes. Par exemple* X_4 *est le nombre de paquets acheté le jour où il trouve sa quatrième carte différente. Que vaut* X_1 *?*
2. *Dans cette question on suppose qu'il y a 4 cartes différentes, notées* a*,* b*,* c *et* d*. L'enfants obtient les cartes suivantes dans l'ordre :*
$$a, a, b, c, a, c, b, d.$$
Sur cet exemple, quelles sont les valeurs de X_1*,* X_2*,* X_3 *et* X_4 *?*
3. *Justifier que la variable aléatoire qui nous interesse pour ce problème est* X_n*.*
4. *On pose* $Y_i = X_i - X_{i-1}$ *pour* i *de* 2 *à* n *et* $Y_1 = X_1$*. Justifier (c'est un simpel calcul qui ne fait pas intervenir les probabilités) que* $X_n = Y_1 + Y_2 + \ldots + Y_n$*.*
5. *Exprimer en français ce que compte* Y_i*.*
6. *Justifier que* Y_i *(pour* $i \geq 2$*) suit une loi géométrique de raison* $1 - \frac{i-1}{n}$*.*
7. *En utilisant la linéarité des espérances (proposition 39), montrer que*
$$E[X_n] = n \sum_{i=1}^{n} \frac{1}{k}.$$
8. *En admettant que* $\sum_{i=1}^{n} \leq 1 + \ln n$*, en déduire une majoration de* $E[X_n]$*. S'il y a 10 cartes, que donne cette majoration ? s'il en a 30 ?*

A titre culturel, on peut montrer que pour tout $c \geq 0$*,*

$$\mathbb{P}(X_n \geq n \ln n + cn + 1) \leq e^{-c}.$$

3.3 Loi Normale (de Gauss), Théorèmes limites

Une loi normale est une loi de probabilité continue, contrairement aux lois vues jusqu'à présent qui étaient discrètes. Sans rentrer dans les détails, une variable aléatoire à densité X est définie par une intégrale du type

$$\mathbb{P}(a < X \leq b) = \int_a^b f(x)dx,$$

où x s'appelle **densité de probabilités** de X.

Il existe de nombreuse lois continues, permettant de modéliser par exemple des températures, le temps, etc. Ici nous ne nous intéresserons qu'aux lois normales.

Une variable aléatoire à densité X suit une **loi normale** de moyenne μ et de variance σ^2 si

$$\mathbb{P}(a < X \leq b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Il n'est pas nécessaire de retenir la formule. La loi normale s'appelle aussi **loi de Gauss**. On la note $\mathcal{N}(\mu, \sigma^2)$. La loi $\mathcal{N}(0,1)$ s'appelle **loi normale centrée réduite**.

Nous n'étudierons pas plus en détail les lois normales, ilf aut juste retenir qu'elles ont des propriétés bien connues.

Le résultat suivant, connu sous le nom de **Théorème central limite**, explique l'importance des lois normales.

Théorème 63 *Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes suivants une même loi de probabilité d'espérance μ et de variance σ^2. Soit $S_n = X_1 + \ldots + X_n$. Alors*
- $E[S_n] = n\mu$,
- $\mathrm{var}(S_n) = n\sigma^2$,
- *Quand n grandit, $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ se comporte comme $\mathcal{N}(0,1)$.*

Nous ne définirons pas plus formellement ce que veut dire *se comporte comme*. Les deux premiers points sont des conséquences directes de la linéarité des espérances et des variances (pour les variables indépendantes). Le troisième point donne beaucoup plus d'information et permet de donner de quantifier le comportement.