Probabilités discrètes et statistique descriptive

DUT Informatique, semestre 2

Version 1.0

29 mars 2010

Ph. Roux

2009-2010

Table des matières

Avertissement

Pour bien utiliser ce polycopié, il faut le lire au fur et à mesure de l'avancement du cours magistral, et prendre le temps de refaire les exercices types qui y sont proposés.

- Les définitions et théorèmes sont numérotés suivant le même ordre que dans le cours magistral. Ils apparaissent dans un cadre grisé et sont en général suivit de leur démonstration, signalée par une barre dans la marge et un $\square$ à la fin.
- La table des matières et l'index (à la fin du document) permettent de retrouver une notion précise dans ce polycopié.
- Les méthodes, techniques et algorithmes qui seront approfondies en TD/TP sont signalées par un cadre (sans couleurs)
- Des exercices types corrigés, et surtout rédigés comme vous devriez le faire en DS, sont signalés par le symbole :

  ✎

- Les erreurs et les confusions les plus fréquentes sont signalées dans des cadres rouges avec le symbole :

  ☡

1 Analyse combinatoire

En mathématiques l'analyse combinatoire étudie les configurations de collections finies d'objets ou les combinaisons d'ensembles finis, et leur dénombrements. Initialement, cette théorie avait pour objet la résolution des problèmes de dénombrement, provenant de l'étude des jeux de hasard, et se développa de façon significative à partir du $XVII^{\text{ième}}$ siècle, en même temps que le calcul des probabilités.

1.1 Factoriel, arrangements, combinaisons

Nous allons commencer par introduire (ou rappeler) un certains nombres d'outils de calcul qui permettent de résoudre les problèmes de dénombrement les plus simples.

Définition 1.1 (factoriel)
Pour $n \in \mathbb{N}$ on appelle « factorielle n »la quantité définie par $0! = 1$ et :

$$n! = n \times (n-1) \times \cdots \times 2 \times 1 \quad si \quad n \geq 1$$

La factorielle va intervenir dans de nombreuses situations (arrangements et combinaisons en particulier) car elle permet de dénombrer facilement des classements.

Théorème 1.2 *Il y a $n!$ manières d'ordonner n objets différents.*

Preuve : Faire un classement c'est remplir une liste, on va donc compter le nombre de possibilités de remplir chaque "case" de la liste en partant de la gauche :

1^{er}	$2^{\text{ième}}$	$3^{\text{ième}}$	...	dernier
$\downarrow$	$\downarrow$	$\downarrow$	...	$\downarrow$
n	$(n-1)$	$(n-2)$	...	1

le nombre total de possibilités est donc bien : $n \times (n-1) \times \cdots \times 2 \times 1 = n!$ $\square$

L'évaluation de $n!$ (ou de son ordre de grandeur) est problème difficile quand n devient très grand car la fonction $n!$ croît très vite. Il existe pourtant un équivalent assez simple de $n!$ permettant de l'exprimer à partir des fonctions usuelles.

Proposition 1.3 (formule de Stirling)

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \quad quand \quad n \to \infty$$

La formule de Stirling est assez difficile à démontrer, mais il faut comprendre son intérêt. Comme la fonction $n!$ croît très vite, cela pose certains problèmes pour les calculs, mais la formule de Stirling permet d'évaluer précisément l'ordre de grandeur de $n!$ sans en faire le calcul comme on peut le voir sur la figure FIG.1.

Définition 1.4 (Arrangement)
Pour $n,p \in \mathbb{N}$ avec $p \leq n$ on appelle « Arrangement de p parmi n »la quantité définie par :

$$A_n^p = \frac{n!}{(n-p)!} = n \times (n-1) \times \cdots \times (n-p+1)$$

On verra qu'il est souvent utile de développer les factoriels dans la définition 1, pour faire apparaître des simplifications. Par exemple on peut récrire la définition

n	$n!$	$\left(\frac{n}{e}\right)^n \sqrt{2\pi n}$
5	120	118.019168
10	5628800	5598695.619
15	1.50767 10^{12}	1.50045 10^{12}
20	2.45290 10^{18}	2.42279 10^{18}
25	1.55112 10^{25}	1.54596 10^{25}

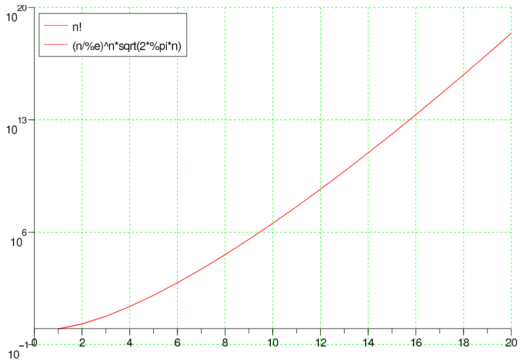


FIG. 1 – *$n!$ et la formule de Stirling*

de l'arrangement A_n^p sans fraction :

$$\begin{aligned}\frac{n!}{(n-p)!} &= \frac{n \times (n-1) \times \cdots \times (n-p+1) \times (n-p) \times \cdots \times 2 \times 1}{(n-p) \times \cdots \times 2 \times 1} \\ &= n \times (n-1) \times \cdots \times (n-p+1)\end{aligned}$$

Ceci permet de voir facilement que, comme le factoriel, l'arrangement permet de dénombrer des classements.

Proposition 1.5 *Il y a A_n^p manières d'ordonner p objets différents pris parmi n.*

Preuve : On fait le même raisonnement qu'auparavant en comptant le nombre de possibilités de remplir chaque "case" de la liste en partant de la gauche :

1^{er}	$2^{\text{ième}}$	$3^{\text{ième}}$	...	$p^{\text{ième}}$
$\downarrow$	$\downarrow$	$\downarrow$	...	$\downarrow$
n	$(n-1)$	$(n-2)$	...	$(n-p+1)$

Le nombre totale de possibilités est donc bien : $n \times (n-1) \times \cdots \times (n-p+1) = A_n^p$
$\square$

Pour les situations ne faisant pas intervenir de notion d'ordre ou de classement, on aura besoin de la notion de combinaisons.

Définition 1.6 (Combinaisons)
Pour $n,p \in \mathbb{N}$ avec $p \leq n$ on appelle « Combinaisons de p parmi n »la quantité définie par :

$$C_n^p = \frac{n!}{(n-p)!p!} = \frac{n \times (n-1) \times \cdots \times (n-p+1)}{p \times (p-1) \times \cdots \times 2 \times 1}$$

En effet les combinaisons permettent de dénombrer le nombre de choix dans un ensemble.

Théorème 1.7 *Il y a C_n^p manières de choisir p objets différents pris parmi n.*

Preuve : Si on choisit p objets différents pris parmi n il y a ensuite $p!$ manières de les ranger, on a donc que

$$A_n^p = C_n^p \times p! \Longrightarrow C_n^p = \frac{A_n^p}{p!} = \frac{n!}{(n-p)!p!}$$

$\square$

Dès que n ou p deviennent grand il devient difficile de calculer C_n^p en utilisant la formule de la définition 1.6. Par exemple si on fait le calcul de C_{20}^3 en développant directement les factorielles on a :

$$C_{20}^3 = \frac{20!}{17!3!} = \frac{2432902008176640000}{355687428096000 \times 6} = 1140$$

alors qu'on aurait pu simplifier les calculs dès le départ :

$$C_{20}^3 = \frac{20!}{17!3!} = \frac{20 \times 18 \times 17}{6} = 1140$$

Il existe une manière plus simple (et plus efficace) de calculer les combinaisons, qui repose sur la formule suivante due à Pascal.

Proposition 1.8 (règle de Pascal)

$$\forall n,p \in \mathbb{N},\ \ p \leq n,\ \ C_{n-1}^{p-1} + C_{n-1}^{p} = C_n^p$$

Preuve : La démonstration est très simple, il suffit de partir du membre de gauche et de mettre au même dénominateur :

$$\begin{aligned}
C_{n-1}^{p-1} + C_{n-1}^{p} &= \frac{(n-1)!}{(p-1)!((n-1)-(p-1))!} + \frac{(n-1)!}{p!(n-1-p)!} \\
&= \frac{(n-1)!}{(p-1)!(n-p)!} + \frac{(n-1)!}{p!(n-1-p)!} \\
&= \frac{p(n-1)!}{p!(n-p)!} + \frac{(n-p)(n-1)!}{p!(n-p)!} \quad \text{(même dénominateur)} \\
&= \frac{p(n-1)! + (n-p)(n-1)!}{p!(n-p)!} \\
&= \frac{n(n-1)!}{p!(n-p)!} = \frac{n!}{p!(n-p)!} = C_n^p
\end{aligned}$$

□

Cette formule permet de calculer les combinaisons pour $n = 0$ puis 1 puis 2 ... en remplissant ligne par ligne un tableau qu'on appelle le **triangle de Pascal**. On peut formaliser cet algorithme comme suit :

```
procédure triangle de Pascal (n)
  pour chaque   ligne k = 0 jusqu'à nfaire
    on met un « 1 »dans la première case de la ligne k
    tant que la case « au dessus »est non-vide faire
             ajouter le C_n^p juste au dessus à son prédécesseur
             mettre le résultat dans la case courante
    fin faire
    on met un « 1 »dans la case suivante de la ligne
  fin faire
```

Pour $n = 7$ on verra qu'on obtient très facilement le tableau FIG.2 et surtout plus rapidement qu'en calculant chaque coefficient avec la définition 1.6.

Calcul de $C_5^2 = C_4^1 + C_4^2$

$n \backslash p$	0	1	2	5	4	5	6	7
0	1							
1	1	1						
2	1	2	1					
5	1	5	5	1				
4	1	4	6	4	1			
5	1	5	10	10	5	1		
6	1	6	15	20	15	6	1	
7	1	7	21	55	55	21	7	1

FIG. 2 – *Remplir le triangle Pascal pour $n = 7$*

Une application très utile des combinaisons est le développement d'une puissance.

Théorème 1.9 (formule du binôme de Newton) $\forall a,b \in \mathbb{R}$ *et* $\forall n \in \mathbb{N}$ *on a*

$$(a+b)^n = \sum_{k=0}^{n} C_n^k a^{n-k} b^k = C_n^0 a^n + C_n^1 a^{n-1} b + \cdots + C_n^{n-1} a b^{n-1} + C_n^n b^n$$

Preuve : Pour comprendre comment vont apparaître les différents termes de cette somme il faut raisonner sur le développement du produit :

$$(a+b)^n = (a+b) \times (a+b) \times \cdots \times (a+b) \times (a+b)$$

pour former le terme $a^{n-k}b^k$ il faut choisir les k facteurs ou l'on va prendre le terme b et l'on prendra le terme a dans les $n-k$ autres facteurs. Il y a donc C_n^k choix possibles chacun produisant un seul terme de la forme $a^{n-k}b^k$, quand on additionne ses termes on obtient le terme final $C_n^k a^{n-k} b^k$ qui apparaît dans la somme. $\square$

✎ **1.1** il suffit donc d'utiliser les coefficients sur la ligne n du triangle de Pascal pour obtenir le développement de $(a+b)^n$. Par exemple pour $n = 4$ on a :

$$\begin{aligned} (1+x)^4 &= \mathbf{1}x^0 + \mathbf{4}x^1 + \mathbf{6}x^2 + \mathbf{4}x^5 + \mathbf{1}x^4 \\ &= 1 + 4x + 6x^2 + 4x^5 + x^4 \end{aligned}$$

$$\begin{aligned} (2+5x)^4 &= 1 \times 16(5x)^0 + \mathbf{4} \times 8 \times 5x + \mathbf{6} \times 4 \times 9x^2 + \mathbf{4} \times 2 \times 27x^5 + 81x^4 \\ &= 16 + 96x + 216x^2 + 216x^5 + 81x^4 \end{aligned}$$

C'est quand même beaucoup plus rapide que d'effectuer les 4 produits (et toutes les simplifications qui vont avec).

1.2 Dénombrement

Le calcul des probabilités discrètes repose sur la capacité à dénombrer le nombre de cas favorables et le nombre total d'éventualités possibles dans une expérience donnée. C'est la source même de la définition du dénombrement :

Définition 1.10
Le dénombrement correspond au calcul du cardinal d'ensembles finis.

Il va donc falloir représenter les résultats d'expériences aléatoires par des ensembles assez simples, pour être dénombrés facilement, et utiliser les outils de la section précédente.

✎ **1.2 Représenter différentes épreuves avec des ensembles**
Prenons l'expérience décrite par la phrase suivante :

« lancer un dé et noter la face supérieure »

on représente les résultats de cette expérience par l'ensemble

$$E = \{1; 2; 3; 4; 5; 6\}$$

sur cette base on peut alors décrire certains types de résultats, définis par des contraintes logiques, en utilisant des sous-ensembles de E et les opérations de base sur les ensembles

- $A =$« les résultats pairs »$\Longrightarrow A = \{2; 4; 6\}$
- $B =$« les résultats supérieurs ou égaux à trois »$\Longrightarrow B = \{3; 4; 5; 6\}$
- $C =$« les résultats pairs **et** supérieurs et égaux à trois »$\Longrightarrow C = A \cap B = \{4; 6\}$
- $D =$« les résultats pairs **ou** supérieurs ou égaux à trois »
  $\Longrightarrow D = A \cup B = \{2; 3; 4; 5; 6\}$
- $N =$« les résultats impairs **ou** strictement inférieurs à trois »
  $\Longrightarrow N = \overline{A \cap B} = \overline{A} \cup \overline{B} = \{1; 2; 3; 5\}$
- $P =$« les résultats impairs **et** strictement inférieurs à trois »
  $\Longrightarrow A \setminus B = A \cap \overline{B} = \{2\}$

Il faudra pouvoir représenter des situations plus complexes en utilisant des représentations graphiques comme les diagrammes de Venn *cf.* FIG.3.

On va donc retrouver les mêmes liens entre le vocabulaire du dénombrement et des probabilités que ceux qui existent entre la logique propositionnelle et la théorie des ensembles. On trouvera un petit récapitulatif des parallèles qu'on peut établir entre ces trois théories dans le tableau FIG.4.

On fera particulièrement attention au dénombrement d'expériences faisant intervenir la disjonction ($\cup$ qui correspond au **ou** logique) pour lesquels il faut utiliser la loi de la mesure.

Théorème 1.11 (loi de la mesure)
Soit $A,B \subset E$ *alors* $\mathrm{Card}(A \cup B) = \mathrm{Card}(A) + \mathrm{Card}(B) - \mathrm{Card}(A \cap B)$*, en particulier*

$$\mathrm{Card}(A \cup B) = \mathrm{Card}(A) + \mathrm{Card}(B) \Longleftrightarrow A \cap B = \emptyset$$

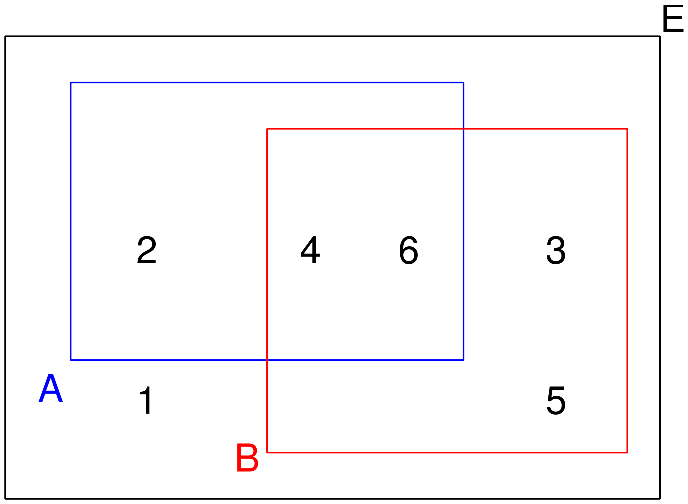


FIG. 3 – *Diagramme de Venn*

Dénombrement	Théorie des ensembles	Logique
événement	$A \in \mathscr{T} = \mathscr{P}(E)$	proposition A
addition	$\cup$	$\vee$
multiplication	$\cap$	$\wedge$
contraire	$\bar{A}$	$\neg A$
impossible	$\emptyset$	F
univers	E	V

FIG. 4 – *Liens entre dénombrement, logique et théorie des ensembles*

Si on reprend l'exemple 3 on a bien :

$$\text{Card}(D) = 5 = 3 + 4 - 2 = \text{Card}(A) + \text{Card}(B) - \text{Card}(C)$$

En général il est plus facile d'utiliser la loi de la mesure pour des ensembles disjoints (cas $A \cap B = \emptyset$), par contre on évitera d'utiliser la différence d'ensemble qui conduit souvent à des erreurs. Sur l'exemple 3 on a :

$$\text{Card}(A \setminus B) = \text{Card}(A \cap \overline{B}) = 1 \neq \text{Card}(A) - \text{Card}(B) = 3 - 4 = -1$$

Pour des tirages multiples selon qu'on ait une notion d'ordre ou pas on pourra utiliser des ensembles de parties d'ensembles $\mathscr{P}(E) = \{A | A \subset E\}$ ou des ensembles de p-liste qui correspondent au produit cartésien.

Définition 1.12 (p-listes ou p-uplets)
On appelle produit cartésien des ensembles $A_1, \dots A_p$ l'ensemble :

$$A_1 \times \cdots \times A_p = \{(x_1, \dots, x_p) | \forall i = 1, \dots p, \;\; x_i \in A_i\}$$

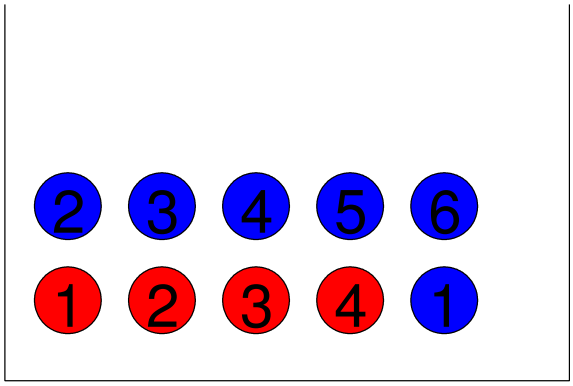
les éléments $(x_1, \dots, x_p)$ de cet ensemble sont appelé des p-listes ou p-uplets

Le dénombrement des p-listes repose sur le théorème suivant vu au premier semestre.

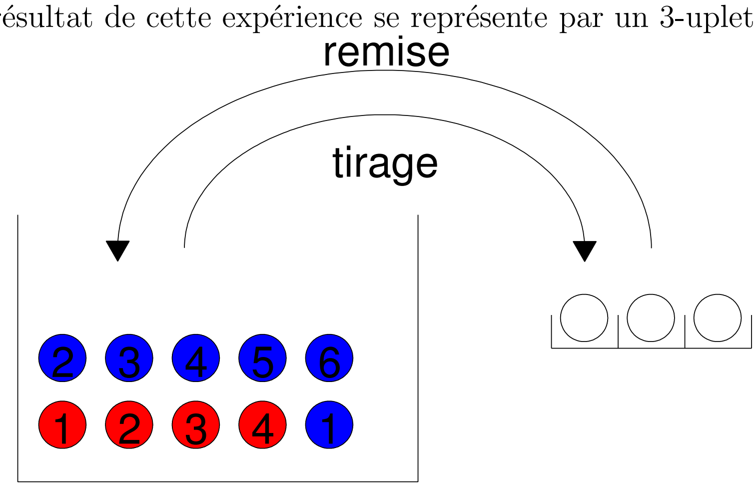
Théorème 1.13 (cardinal d'un produit d'ensemble)
Soit $A,B \subset E$ alors $\text{Card}(A \times B) = \text{Card}(A) \times \text{Card}(B)$

1.3 Application a des tirages répétitifs dans une urne

on va prendre une urne contenant 4 boules rouges (numérotées de 1 à 4) et 6 boules bleus (numérotées de 1 à 6) et examiner différents types de tirages dans cette urne.

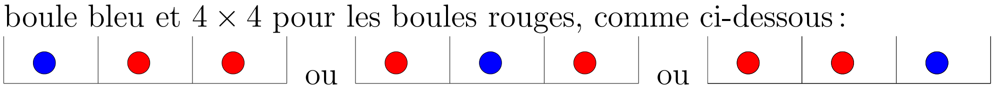


Expérience 1: « tirer 3 boules successivement avec remise » comme il y a un ordre un résultat de cette expérience se représente par un 3-uplet :



quelques exemples de dénombrements pour cette expérience :

- nombre total de tirages est donc $10^3 = 1000$
- nombre de tirages avec la première boule bleu $6 \times 10^2 = 600$ (attention on autorise que d'autres boules soient bleus ensuite)
- R_0 :« tirages avec 0 boule rouge »$\Longrightarrow$ $\text{Card}(R_0) = 6^3 = 216$
- R_3 :« tirages avec 3 boules rouges »$\Longrightarrow$ $\text{Card}(R_3) = 4^3 = 64$
- R_2 :« tirages avec 2 boule rouges »$\Longrightarrow$ $\text{Card}(R_2) = 3 \times (6 \times 4^2) = 288$ car pour chacune des 3 configurations possibles il y a 6 choix pour la boule bleu et 4×4 pour les boules rouges, comme ci-dessous :

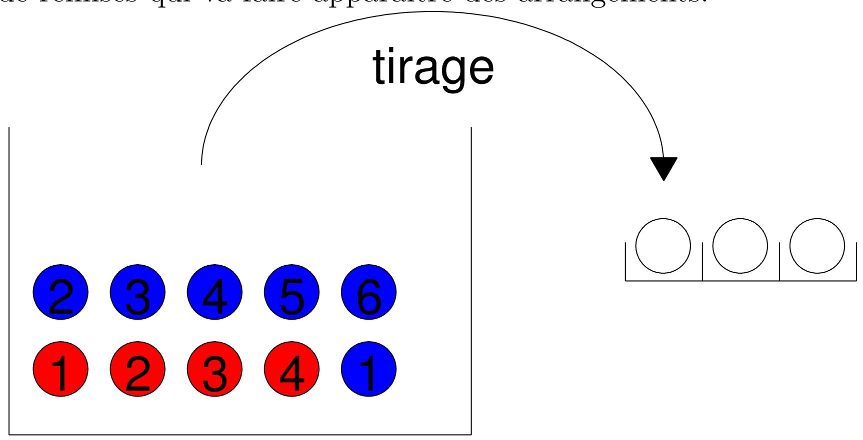


$6 \times 4 \times 4 \quad + \quad 4 \times 6 \times 4 \quad + \quad 4 \times 4 \times 6$

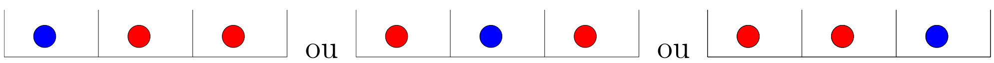
- R_1 =« tirages avec une boule rouge »$\Longrightarrow$ $\text{Card}(R_1) = 3 \times (4 \times 6^2) = 432$
- A:« tirages avec au moins une boule rouge »
$\Longrightarrow \text{Card}(A) = 10^3 - \text{Card}(R_0) = 784 = \text{Card}(R_1) + \text{Card}(R_2) + \text{Card}(R_3)$

On peut en plus vérifier que $64 + 216 + 288 + 432 = 1000$ ce qui correspond au fait que $\{R_0, R_2; R_2; R_3\}$ est une partition de l'ensemble des résultats.

Expérience 2« tirer 3 boules successivement et sans remise » là encore on va modéliser les résultats par un 3-uplet mais cette fois il faut tenir compte de l'absence de remises qui va faire apparaître des arrangements.



- nombre total de tirages :$10 \times 9 \times 8 = A_{10}^3 = 720 = \mathrm{Card}(\Omega)$
- nombre de tirages avec la première boule bleu $6 \times 9 \times 8 = 432$
- R_0 :« tirages avec 0 boule rouge »$\Longrightarrow \mathrm{Card}(R_0) = A_6^3 = 120$
- R_3 :« tirages avec 3 boules rouges »$\Longrightarrow \mathrm{Card}(R_3) = 4 \times 3 \times 2 = A_4^3 = 24$
- R_2 :« tirages avec 2 boule rouge »$\Longrightarrow \mathrm{Card}(R_2) = 3 \times (6 \times 4 \times 4) = 216$
  en faisant le même raisonnement que dans l'expérience 1 (mais en tenant compte de l'absence de répétitions) car pour chacune des 3 configurations on a 6 choix pour la boule bleu et 4×3 choix pour les rouges :

  $6 \times 4 \times 3 \quad + \quad 4 \times 6 \times 3 \quad + \quad 4 \times 3 \times 6$

  On peut aussi retrouver ce résultat par un autre raisonnement: pour faire un tirage il faut :

  « choisir deux boules rouges **et** choisir une boule bleu **et** ranger ces 3 boules »

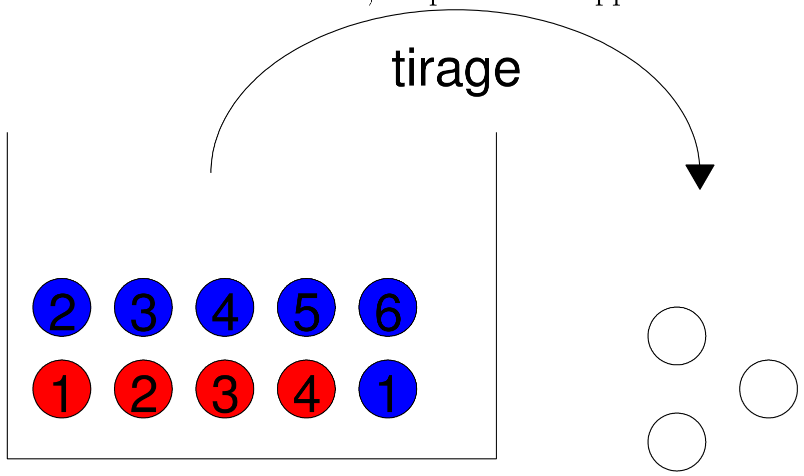
  $\Longrightarrow \mathrm{Card}(R_2) = C_4^2 \times C_6^1 \times 3! = 216$
- R_1 :« tirages avec 2 boule rouge »$\Longrightarrow \mathrm{Card}(R_1) = 3 \times (4 \times 6 \times 5) = 360$
  ou $\mathrm{Card}(R_1) = 3! \times C_6^2 \times C_4^1 = 360$ en suivant le raisonnement :

« choisir une boule rouge **et** choisir deux boules bleus **et** ranger ces 3 boules »

- A:« tirages avec au moins une boule rouge »
  $\Longrightarrow \mathrm{Card}(A) = 720 - \mathrm{Card}(R_0) = 600 = \mathrm{Card}(R_1) + \mathrm{Card}(R_2) + \mathrm{Card}(R_3)$

et comme $\{R_0, R_2; R_2; R_3\}$ est toujours une partition de l'ensemble des résultats on peut vérifier que : $24 + 120 + 216 + 360 = 720$.

Expérience 3« tirer 3 boules simultanément (donc sans remise) » cette fois on ne peut plus utiliser de 3-uplets, un résultat est un sous-ensemble à 3 éléments de l'ensemble des boules, ce qui va faire apparaître des combinaisons.



- nombre total de tirages :$C_{10}^3 = 120$
- cette fois la phrase « la première boule est bleu »n'a pas de sens dans!
- R_3 :« tirages avec 3 boules rouges »$\Longrightarrow$ $\text{Card}(R_0) = C_4^3 = 4$
- R_0 :« tirages avec 0 boule rouge »$\Longrightarrow$ $\text{Card}(R_0) = C_6^3 = 20$
- R_2 :« tirages avec 2 boules rouges »$\Longrightarrow$ $\text{Card}(R_2) = C_4^2 \times C_6^1 = 36$
- R_1 :« tirages avec 1 boule rouge »$\Longrightarrow$ $\text{Card}(R_1) = C_6^2 \times C_4^1 = 60$
- A:« tirages avec au moins une boule rouge »
  $\Longrightarrow \text{Card}(A) = 120 - \text{Card}(R_0) = 116 = \text{Card}(R_1) + \text{Card}(R_2) + \text{Card}(R_3)$

et on a bien $4 + 20 + 36 + 60 = 120$.

On pourra retenir le tableau FIG.5 pour faire la différence entre arrangements, combinaisons et p-listes.

ordre / répétition	simultanés	successifs
avec remise	$\emptyset$	p-uplets n^p
sans remise	Combinaisons C_n^p	Arrangement A_n^p

FIG. 5 – *différents tirages de p éléments parmi n*

On remarquera enfin que pour chaque événement on a sous-entendu le mot « exactement »(exactement une boule rouge pour R_1,...) c'est l'interprétation qu'on adoptera par défaut. Dans le cas contraire on précisera le mot « au moins », mais dès que ce mot apparaît il y a un piège et il faut reformuler la question en utilisant la négation (qui se traduit par « aucun ») ou des « exactement ».

2 Statistiques descriptives

La statistique est la partie des mathématiques qui s'intéresse à la détermination des caractéristiques d'un ensemble de données (généralement très important). Ici nous ne nous intéresserons pas aux problèmes posés par la collecte d'un grand nombre de données, mais seulement aux deux aspects suivants :

- le traitement des données collectées qui est appelé *la statistique descriptive*
- l'interprétation des données est appelée *l'inférence statistique*

nous verrons à la fin du cours comment peut être faite l'interprétation en s'appuyant sur la théorie des sondages.

2.1 Représentation des données statistiques

Il existe deux types de données statistiques :

- les caractères qualitatifs (qui ne peuvent pas être mesurées) qui constitue une nomenclature, par exemple le sexe (Masculin/Féminin), les couleurs ...
- les caractères quantitatifs (qui peuvent être mesurées) et que l'on peut représenter par un nombre réel, par exemple taille, le poids, ...

Dans ce cours nous nous intéresserons aux caractères quantitatifs, qu'on va représenter par des variables statistiques.

Définition 2.1 (série statistique) *Soit P un ensemble, appelé « population », alors un variable statistique est une application :*

$$\begin{array}{rccc} X : & P & \longrightarrow & \mathbb{R} \\ & i & \longmapsto & X(i) \end{array}$$

$X(P)$ est appelé univers image de X et on dira que :

- *X est discrète si $X(P)$ est un ensemble discret (fini ou dénombrable)*
- *X est continue si $X(P)$ est un intervalle de $\mathbb{R}$*

$N = \mathrm{Card}(P)$ sera appelé l'effectif total

Dans la pratique on s'intéresse surtout à l'univers image $X(P)$[1] et on identifie souvent une variable statistique avec la liste des valeurs prises $X(i)$ prises par la variable. Dans ce cas on parle en général de *série statistique*.

✎ **2.1** Exemples de séries statistiques : on considère la population P composée des étudiants inscrits au semestre 1 du DUT INFORMATIQUE à Lannion (on admettra qu'on a un effectif total $N = 100$) et nous définissons deux variables statistiques :

- $X =$« nombre d'absence injustifiées (en maths) au semestre 1 »
- $Y =$« taille (exprimée en cm) de chaque individu »

ce sont bien des caractères quantitatifs qui peuvent donc être décrits par une série statistique, qui correspondent aux listes de nombres :
$X =$(2, 4, 2, 1, 0, 0, 3, 1, 3, 0, 2, 3, 0, 3, 1, 0, 2, 0, 2, 1, 0, 3, 1, 2, 2, 3, 1, 3, 2, 2, 0, 2, 0, 3, 2, 1, 2, 2, 3, 2, 0, 2, 0, 2, 0, 2, 1, 2, 0, 0, 3, 1, 0, 2, 0, 2, 1, 2, 2, 2, 1, 2, 0, 2,

1. pour des raisons de confidentialité on ne peut pas en général identifier chaque individu de la population P

1, 0, 0, 4, 1, 0, 3, 0, 2, 0, 0, 3, 1, 3, 0, 0, 1, 2, 1, 2, 6, 2, 2, 1, 1, 2, 3, 1, 0, 1, 1, 1, 3, 1, 3, 2)
et
Y =(188, 169, 181, 180, 159, 180, 164, 184, 177, 159, 187, 174, 170, 173, 175, 150, 171, 166, 173, 182, 167, 173, 174, 175, 166, 173, 180, 188, 165, 173, 178, 182, 175, 186, 162, 193, 173, 184, 184, 184, 182, 187, 193, 161, 193, 169, 163, 179, 179, 184, 182, 182, 172, 175, 193, 170, 176, 166, 177, 163, 191, 171, 189, 183, 178, 197, 166, 187, 180, 172, 181, 167, 177, 177, 186, 174, 168, 182, 183, 182, 170, 186, 193, 167, 184, 159, 169, 192, 172, 167, 178, 176, 177, 175, 172, 175, 182, 176, 179, 188)
c'est ce qu'on appelle des données brutes.

Évidement il n'est pas très parlant de représenter une série statistiques sous forme de données brutes c'est pourquoi on va utiliser diverses représentations permettant de faire ressortir la répartition des données.

Définition 2.2 (modalités)
Soit X une variable statistique alors on appelle modalités *de X :*

- *l'ensemble des valeurs différentes $\{x_1; x_2; \ldots; x_n\}$ prisent par X si X est discrète*
- *un ensemble d'intervalles $\{[x_1; x_2[; [x_2; x_3[; \ldots; [x_{n-1}; x_n]\}$ formant une partition de $X(P)$ si X est continue*

un tableau statistique est un tableau regroupant les caractéristiques d'une série statistiques par modalités :

série discrète

modalités	x_1	x_2	x_3	...
effectif n_i				
fréquence f_i				

série continue

modalités	$[x_1; x_2[$	$[x_2,x_3[$	...
effectif n_i			
fréquence f_i			

Si N est l'effectif total alors $f_i = \frac{n_i}{N}$ $\forall i = 1, \ldots, n$.

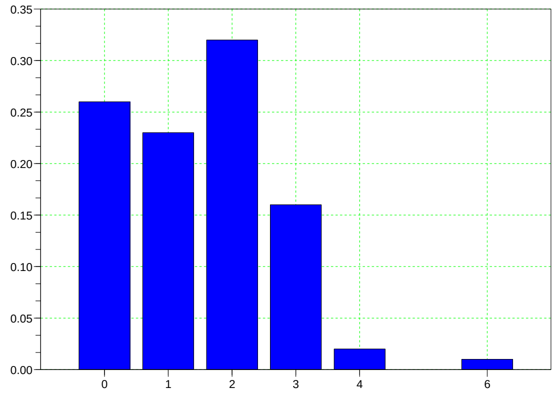
On représentera graphiquement les données d'une série discrète (resp. continue) en dessinant des colonnes de hauteurs (resp. surface) proportionnelle à l'effectif de chaque modalité, c'est ce qu'on appelle un diagramme en bâtons (resp. un histogramme).

✎ **2.2 tableau statistique d'une série statistique** Pour faire le tableau statistique de la série X, dont les modalités sont $\{0; 1; 2; 3; 4; 6\}$, il faudra à partir des données brutes construire le tableau :

$X_i =$	0	1	2	3	4	6	
n_i	26	23	32	16	2	1	$\to \sum n_i = N = 100$
f_i	0.26	0.23	0.32	0.16	0.02	0.01	$\to \sum f_i = 1$

(→ modalités)

On pourra aussi représenter graphiquement la série statistique à l'aide d'un diagramme en bâtons :

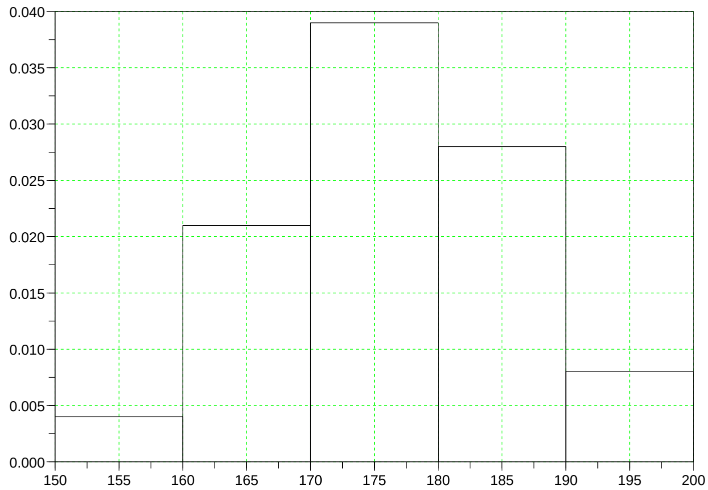


ou chaque colonne à une hauteur proportionnelle à l'effectif de la modalité.

Pour la série Y, même si les tailles ne prennent que des valeurs entières, il n'est pas opportun de la représenter comme une série discrète puisque dans ce cas on se retrouverait avec 34 modalités différentes! C'est pour cette raison qu'on va considérer cette série comme une série continue prenant ses valeurs dans l'intervalle $[150; 200]$. Si on découpe cet intervalle en modalités de même taille on obtient :

$Y_i =$	[150,160[	[160,170[	[170,180[	[180,190[	[190,200[
n_i	4	21	39	28	8
f_i	0.04	0.21	0.39	0.28	0.08

ce qui donne :



Proposition 2.3 (effectifs corrigés) *Soit X une variable statistique continue alors ayant des modalités $[x_1; x_2[$ $[x_2; x_3[$ $\dots [x_{n-1}; x_n]$ d'amplitude différentes alors pour chaque modalité on appelle « effectif corrigé » ou « hauteur corrigée » la hauteur qu'il faut donner dans l'histogramme à la colonne pour que sa surface soit proportionnelle à son effectif réel.*

Dans un histogramme normalisé on calcule les hauteurs corrigées par la formule :

$$h_i = \frac{f_i}{|x_{i+1} - x_i|} = \frac{n_i}{N|x_{i+1} - x_i|}$$

de telle sorte que la surface totale des colonnes soit égale à 1.

Preuve : chaque modalité $[x_i,x_{i+1}[$ à pour largeur $|x_{i+1} - x_i|$ et pour hauteur h_i sa surface est donc :

$$h_i \times |x_{i+1} - x_i| = f_i = \frac{n_i}{N}$$

est bien proportionnel à l'effectif réel, de plus il est évident que la somme des aires est alors $\sum_i f_i = 1$. □

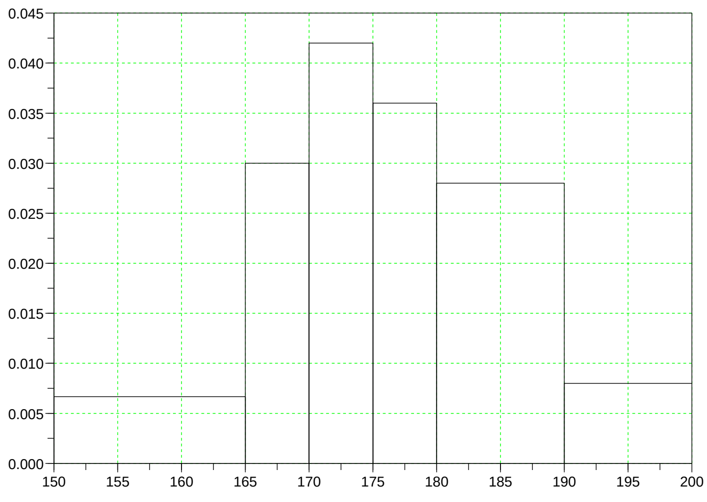
✎ **2.3** Reprenons le cas de la variable Y que l'on étudie avec les modalités :

[150,165[, [165,170[, [170,175[, [175,180[, [180,190[, [190,200[

Après avoir calculé les effectifs de chaque modalité, il faut cette fois calculer les hauteurs corrigées :

$Y_i =$	[150,165[	[165,170[	[170,175[	[175,180[	[180,190[	[190,200[
f_i	0.1	0.15	0.21	0.18	0.28	0.08
$\|x_{i+1} - x_i\|$	15	5	5	5	10	10
h_i	0.0066667	0.03	0.042	0.036	0.028	0.008

pour pouvoir dessiner l'histogramme



Un autre représentation permettant d'identifier la nature de la répartition des données statistiques est donnée par la fonction de répartition, qui est étroitement associée à la notion de cumul des fréquences.

Définition 2.4 (Fonction de répartition) *Soit X une série statistique alors la fonction de répartition de X*

$$\begin{array}{rccl} F: & \mathbb{R} & \longrightarrow & [0;1] \\ & x & \longmapsto & \displaystyle\sum_{\{i|x_i \leq x\}} f_i \end{array}$$

On retiendra que la fonction de répartition vérifie les conditions suivantes :

Proposition 2.5 *Soit F la fonction de répartition d'une série statistique X alors*

- *F est croissante*
- $\lim_{t\to-\infty} F(t) = 0$
- $\lim_{t\to+\infty} F(t) = 1$
- *si X est une variable discrète alors F est constante par morceau.*
- *si X est une variable continue alors F on la représentera par une fonction continue et affine par morceau.*

Preuve :

- si $s \leq t$ alors
$$F(s) = \sum_{\{i|x_i \leq s\}} f_i \leq \sum_{\{i|x_i \leq t\}} f_i = F(t)$$
- si $s < x_1$ (la première modalité) alors
$$F(s) = \sum_{\{i|x_i \leq s\}} f_i = 0 \Longrightarrow \lim_{s\to-\infty} F(s) = 0$$
- inversement si $t > x_n$ (dernière modalité) alors
$$F(t) = \sum_{\{i|x_i \leq t\}} f_i = f_1 + \cdots + f_n = 1 \Longrightarrow \lim_{t\to+\infty} F(t) = 1$$

Si X est une variable discrète, l'effectif d'une modalité $]-\infty,x]$ ne va varier que lorsque x prend pour valeur une des modalités x_i. Il est donc normal que le graphe de F soit constant par morceau. Par contre la représentation du graphe de F par une fonction affine par morceau dans le cas où X est une variable continue, est une approximation qui se justifiera d'un point de vue pratique (voir les exemples). □

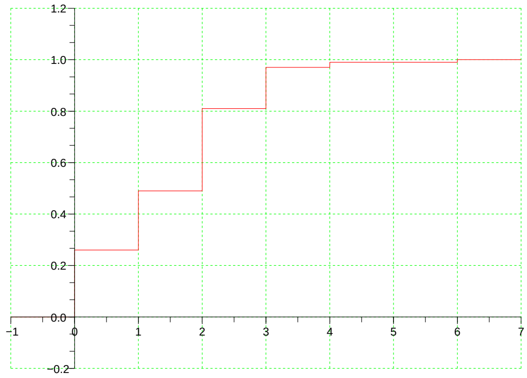
✎ **2.4** Pour construire la fonction de répartition de la variable X il faut donc faire le cumul des fréquences. Pour cela repartons du tableau statistique et ajoutons y une ligne pour calculer le cumul des fréquences :

$X_i =$	0	1	2	3	4	6
f_i	0.26	0.23	0.32	0.16	0.02	0.01
cumul	0.26	0.49	0.81	0.97	0.99	1

Cette dernière ligne donne la valeur de la fonction de répartition sur l'intervalle qui démarre à la modalité correspondante :

I	$]-\infty,0[$	$[0,1[$	$[1,2[$	$[2,3[$	$[3,4[$	$[4,6[$	$[6,+\infty[$
$F(t)$	0	0.26	0.49	0.81	0.97	0.99	1

ce qui donne :



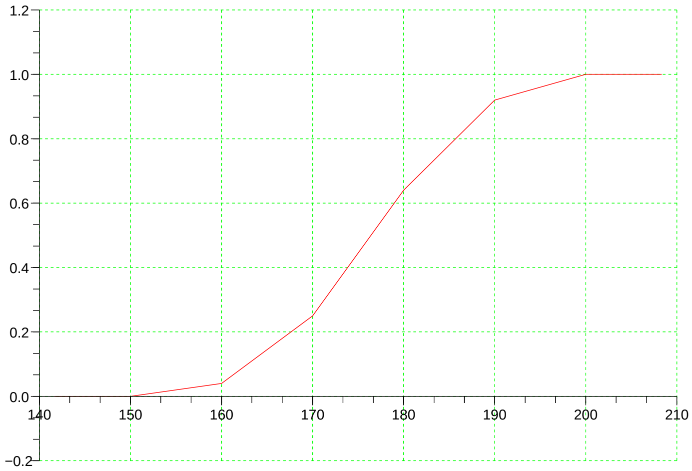
de même pour Y on repart du tableau statistique :

$Y_i =$	[150,160[	[160,170[	[170,180[	[180,190[	[190,200[
f_i	0.04	0.21	0.39	0.28	0.08
cumul	0.04	0.25	0.64	0.92	1

permet de calculer les valeurs de la fonction de répartition :

t	$]-\infty$,150[	160	170	180	190	200	[200, $+\infty$[
$F(t)$	0	0.04	0.25	0.64	0.92	1	1

par contre on doit représenter la fonction de répartition comme une fonction continue affine par morceau (*i.e.* son graphe est une suite de segments de droites) entre chaque borne des modalités:



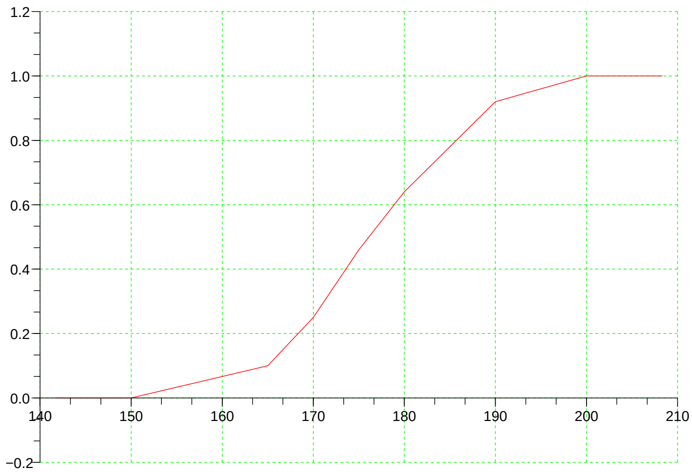
ce procédé permet d'obtenir des résultats numériquement cohérents quand on change les modalités lors de l'étude d'une série statistique. Reprenons le cas de Y avec des modalités de taille variable, ici on aura pas de problème d'effectifs corrigés, le tableau des fréquences devient :

$Y_i =$	[150,165[	[165,170[	[170,175[	[175,180[	[180,190[	[190,200[
f_i	0.1	0.15	0.21	0.18	0.28	0.08
cumul	0.1	0.25	0.46	0.64	0.92	1

ce qui donne cette fois pour la fonction de répartition :

t	$]-\infty,150[$	165	170	175	180	190	200	$[200,+\infty[$
$F(t)$	0	0.1	0.25	0.46	0.64	0.92	1	1

et pour le graphe :



2.2 Paramètres statistiques

L'analyse des données statistiques doit permettre d'isoler certaines valeurs remarque définissant la manière dont son réparties les données statistiques. Le plus connu est la moyenne empirique :

Définition 2.6 (Moyenne empirique)
Soit X une série statistique alors la moyenne de X, notée $\overline{X}$, est la valeur :

$$\overline{X} = \underbrace{\frac{1}{N}\sum_{k\in P} X(k)}_{\text{données brutes}} = \underbrace{\frac{1}{N}\sum_{i=1}^{n} n_i x_i}_{\text{effectifs par modalités}} = \underbrace{\sum_{i=1}^{n} f_i x_i}_{\text{fréquences par modalités}}$$

La moyenne permet de cerner facilement la valeur centrale de la série statistique.

Définition 2.7 (Moyenne approchée)
Soit X une variable continue regroupée suivant les modalités $[x_1;x_2[$ $[x_2;x_3[$ $\ldots[x_{n-1};x_n]$. Si on pose $xc_i = \frac{x_{i+1}+x_i}{2}$ le milieu de chaque modalité alors la moyenne approchée est définie par

$$\textit{moyenne approchée } = \sum_{i=1}^{n} f_i \times xc_i$$

✎ **2.5** Pour une variable discrète X on peut calculer la moyenne directement à partir du tableau statistique :

$X_i =$	0	1	2	3	4	6
n_i	26	23	32	16	2	1
f_i	0.26	0.23	0.32	0.16	0.02	0.01

en faisant

$$\overline{X} = 0.26 \times 0 + 0.23 \times 1 + 0.32 \times 2 + 0.16 \times 3 + 0.02 \times 4 + 0.01 \times 6 = 1.49$$

de même pour la variable Y on trouvera $\overline{Y} = 176.71$. Mais dans le cas d'une variable continue comme Y, si on a pas accès aux données brutes on peut quand même calculer une valeur approchée de la moyenne à partir du tableau des fréquences :

$Y_i =$	[150,160[	[160,170[	[170,180[	[180,190[	[190,200[
f_i	0.04	0.21	0.39	0.28	0.08

$$\begin{aligned} \text{moyenne approchée} &= 0.04 \times 155 + 0.21 \times 165 + 0.39 \times 175 + 0.28 \times 185 + 0.08 \times 195 \\ &= 176.5 \approx 176.71 = \overline{Y} \end{aligned}$$

la valeur est proche de la moyenne calculée avec les données brutes. Si on change les modalités on va trouver un résultat proche :

$Y_i =$	[150,165[	[165,170[	[170,175[	[175,180[	[180,190[	[190,200[
f_i	0.1	0.15	0.21	0.18	0.28	0.08

cette fois on trouve :

$$\begin{aligned} &\text{moyenne approchée} \\ &= 0.1 \times 157.5 + 0.15 \times 167.5 + 0.21 \times 172.5 + 0.18 \times 177.5 + 0.28 \times 185 + 0.08 \times 195 \\ &= 176.45 \approx 176.71 = \overline{Y} \end{aligned}$$

On retiendra aussi que si on manipule plusieurs séries statistiques sur la même population alors la moyenne est linéaire :

Proposition 2.8 (linéarité de la moyenne)
Soit X,Y deux séries statistiques (sur une même population) et $a,b \in \mathbb{R}$ alors

$$\overline{aX + bY} = a\overline{X} + b\overline{Y}$$

en particulier $\overline{aX + b} = a\overline{X} + b$.

La moyenne n'est pas suffisante pour comprendre la manière dont sont réparties les valeurs d'une série statistique X. On a besoin d'avoir une indication sur la dispersion des données. Le meilleur de ces indicateurs est l'écart-type qui est définit comme suit.

Définition 2.9 (variance et écart-type) *Soit X une série statistique de moyenne $\mu = \overline{X}$ alors on appelle variance de X la valeur*

$$\mathbb{V}\text{ar}(X) = \overline{(X - \mu)^2}$$

et l'écart-type de X est $\sigma_X = \sqrt{\mathbb{V}\text{ar}(X)}$.

On remarquera qu'on a toujours $\mathbb{V}\text{ar}(X) \geq 0$ puisque c'est la moyenne de nombres positifs et l'écart-type est donc toujours bien défini! L'écart-type σ donne une estimation de l'intervalle $[\overline{X} - \sigma; \overline{X} + \sigma]$ centré sur la moyenne où sont concentrées les 2/3 des données. il permet donc de comprendre à quel point les données sont concentrées (ou pas) autour de la moyenne. Pour une variable continue si on a pas accès aux données brutes on pourra aussi calculer une variance et un écart-type approché (comme pour la moyenne approché).

Dans la pratique on calculera la variance d'une série via le théorème suivant.

Théorème 2.10 (Koening)
Soit X une variable statistique alors

$$\mathbb{V}\text{ar}(X) = \overline{X^2} - \overline{X}^2$$

Preuve : On a que $(X - \mu)^2 = X^2 - 2\mu X + \mu^2$ donc en utilisant la linéarité de la moyenne on obtient

$$\begin{aligned}
\mathbb{V}\text{ar}(X) &= \overline{(X-\mu)^2} \\
&= \overline{X^2 - 2\mu X + \mu^2} \\
&= \overline{X^2} - 2\mu\overline{X} + \mu^2 \\
&= \overline{X^2} - 2\overline{X} \times \overline{X} + \overline{X}^2 \qquad \text{car } \mu = \overline{X} \\
&= \overline{X^2} - \overline{X}^2
\end{aligned}$$

□

2.6 Calculons la variance de X à partir du tableau statistique :

$X_i =$	0	1	2	3	4	6
n_i	26	23	32	16	2	1
f_i	0.26	0.23	0.32	0.16	0.02	0.01

on a déjà obtenu que $\overline{X} = 1.49$ de même on obtient

$$\begin{aligned}
\overline{X^2} &= 0.26 \times 0 + 0.23 \times 1 + 0.32 \times 4 + 0.16 \times 9 + 0.02 \times 16 + 0.01 \times 36 \\
&= 3.63
\end{aligned}$$

donc en utilisant le théorème de Koening on en tire pour la variance $\mathbb{V}\text{ar}(X) = 3.63 - 1.49^2 = 3.63 - 2.2201 = 1.4099$. On aurait trouvé le même résultat en calculant :

$$\begin{aligned}
\mathbb{V}\text{ar}(X) &= 0.26 \times (0 - 1.49) + 0.23 \times (1 - 1.49) + 0.32 \times (2 - 1.49) \\
&\quad + 0.16 \times (3 - 1.49) + 0.02 \times (4 - 1.49) + 0.01 \times (6 - 1.49) \\
&= 1.4099
\end{aligned}$$

dans tous les cas on trouve l'écart-type $\sigma_X = \sqrt{1.4099} = 1.1873921$.

Pour les cas où on a pas directement accès aux données brutes (en particulier pour les variables statistiques continues) il existe d'autres indicateurs de centralité et de dispersion qui peuvent être calculés à partir de l'histogramme ou du graphe de la fonction de répartition .

Définition 2.11 (Mode et médiane)
Soit X une série statistique alors

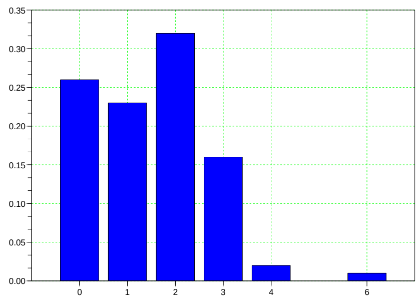
- *Le mode est la modalité la plus représentée (effectif maximum)*
- *la médiane m est la valeur qui permet de couper la population en 2 part égales: $\{i|X(i) \geq m\}$ et $\{i|X(i) \leq m\}$*
  - *si l'effectif total $N = 2p + 1$ est impair alors la médiane est la $p + 1^{ième}$ valeur*
  - *si l'effectif total $N = 2p$ est pair alors la médiane est la moyenne des $p^{ième}$ et $p + 1^{ième}$ valeurs*

On peut aussi caractériser la médiane à partir de la fonction de répartition, ce qui permet d'en calculer une valeur approchée si on a pas accès aux donnée brutes.

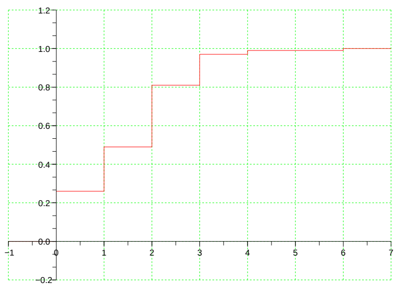
Proposition 2.12 (médiane et fonction de répartition) *Soit X une série statistique alors et F la fonction de répartition associée comme étant la solution de $F(x) = 0.5$ que l'on pourra déterminer graphiquement pour une variable statistique continue. Pour une variable statistique discrète la médiane il y a 2 cas :*

- *la première valeur telle que $F(x) > 0.5$ (si $F(x) = 0.5$ n'a pas de solution)*
- *la moyenne des valeurs telle que $F(x) = 0.5$ (si $F(x) = 0.5$ a plusieurs solutions)*

✎ **2.7** On reprend la variable X (nombre d'absences injustifiées) la moyenne est $\overline{X} = 1.49$ le mode et la médiane se trouvent facilement sur l'histogramme et le graphe de la fonction de répartition



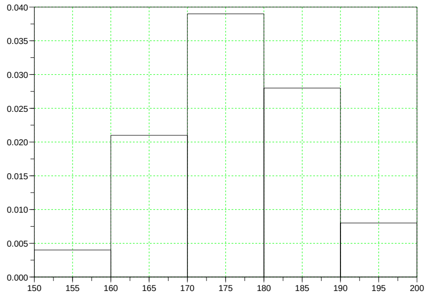
mode 2

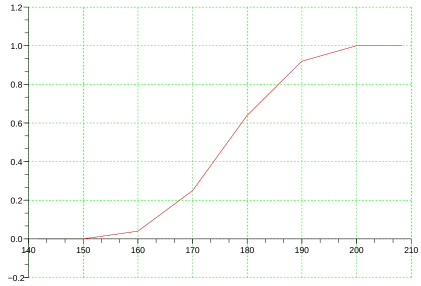


mediane 2

on peut remarquer qu'ici il n'y a pas de solution à l'équation $F(x) = 0.5$ et que la première valeur telle que $F(x) > 0.5$ est bien $x = 2$ d'après le tableau :

Pour la série continue Y on a pour moyenne brute $\overline{Y} = 176.71$.



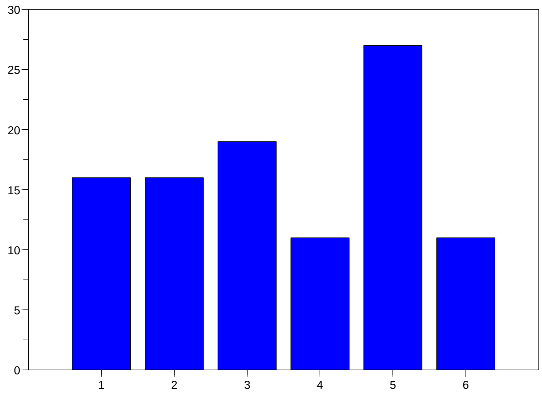


mode 175 $\in$ [170,180[mediane 177

alors que graphiquement on trouve que la médiane vaut ≈ 176.4.
La médiane est en général un bon indicateur de centralité mais sa détermination se fait graphiquement (dont de manière imprécise), au contraire le mode est plus facile à déterminer mais peut être très éloigné de la moyenne pour des séries fortement dispersées. Par exemple pour une série Z de résultats de $N = 100$ lancés d'un dé à 6 faces on obtient :

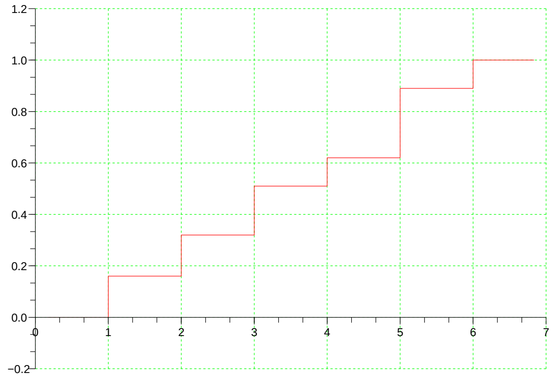
$Z_i =$	1	2	3	4	5	6
n_i	16	16	19	11	27	11
f_i	0.16	0.16	0.19	0.11	0.27	0.11

effectif total $N = 100$



le mode sera donc 5 assez éloigné de la moyenne empirique $\overline{Z} = 3.5$. Par contre le calcul de la médiane donne 3 d'après la fonction de répartition :

I	$]-\infty,1[$	$[1,2[$	$[2,3[$	$[3,4[$	$[4,5[$	$[5,6[$	$[6,+\infty[$
$F(t)$	0	0.16	0.32	0.51	0.62	0.89	1



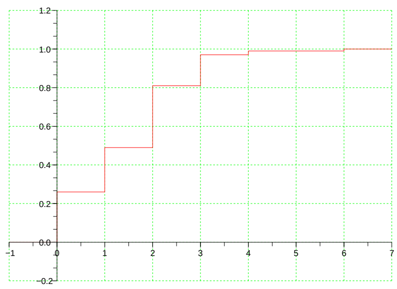
Définition 2.13 (quartiles)
Soit X une série statistique alors on définie les quartiles Q_1,Q_3 à partir de la fonction de répartition de la manière suivante :

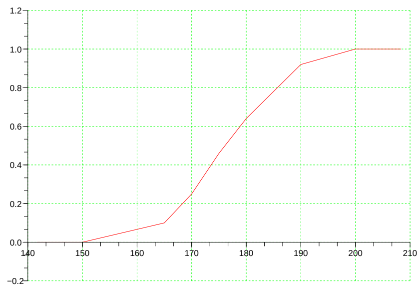
- *Q_i est la solution de $F(Q_1) = 0.25$ et $F(Q_3) = 0.75$ ou à défaut (pour une variable discrète):*
  - *$Q_i = x$ la première valeur telle que $F(x) > 0.25 \times i$ (si $F(x) = 0.25 \times i$ n'a pas de solution)*
  - *la moyenne des valeurs telle que $F(x) = 0.25 \times i$ (si $F(x) = 0.25 \times i$ a plusieurs solutions)*

l'inter-quartile est la valeur $Q_3 - Q_1$

La définition des quartiles est très similaire à celle de la médiane (d'ailleurs la médiane peut être vue comme le quartile Q_2). L'inter-quartile est un indicateur de dispersion, il donne la largeur d'un intervalle $[Q_1; Q_3]$ contenant plus de 50% des données. De ce fait en général l'inter-quartile est $\approx 1.3333\sigma_X$.

✎ **2.8** on peut reprendre le calcul des inter-quartile pour les séries X et Y:





- écart-type $\sigma_X = 1.1873921$
- inter-quartille $q_3 - q_1 = 2 - 0 = 2$

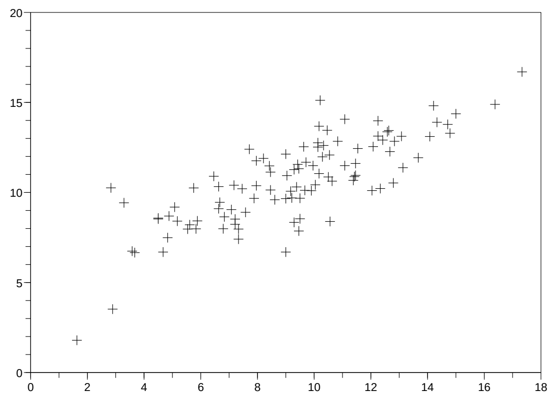
- écart-type $\sigma_Y = 9.1129523$
- inter-quartille $q_3 - q_1 = 183 - 170.5 = 12.5$

2.3 Corrélation

En statistique il est souvent important de rechercher s'il existe un lien entre deux variables X et Y, lien qui, dans l'idéal, pourrait s'exprimer par une équation comme $Y = aX + b$. Prenons un exemple simple : on considère la population des étudiants inscrit au S1 du DUT INFORMATIQUE de Lannion en 2008 ($N = 103$ étudiants) et les deux variables aléatoires associées à cette population

$M =$« note de maths au S1 »et $N =$« moyenne générale au S1 »

On voudrait quantifier le lien entre ces deux variables statistiques. Si on place les points de coordonnées (M_i, U_i) sur un graphe on obtient le "nuage" suivant :



ces points semblent grossièrement alignés sur une droite, cela signifie qu'on peut quantifier le lien entre M et U par une équation de la forme $U = aM + b$. Le but est donc de trouver les coefficients a et b de telle sorte que la droite $y = ax + b$ passe au plus près du maximum de points comme sur la figure FIG.6. C'est ce qu'on appelle faire une *régression linéaire*.

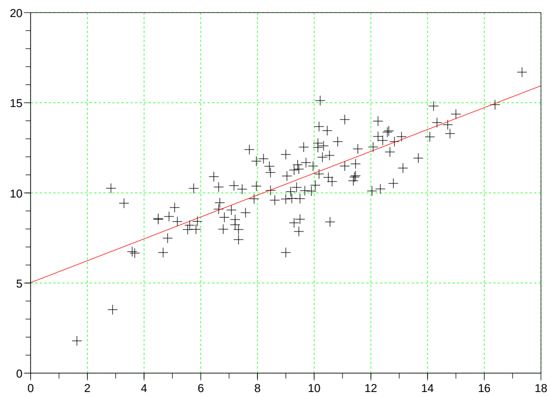


FIG. 6 – *droite de régression*

Cette méthode empirique n'est pas satisfaisante, pour obtenir une méthode plus consistante nous allons devoir définir une nouvelle quantité statistique : la *corrélation.*

Définition 2.14 (corrélation) *Soient X et Y deux séries statistiques d'effectif N alors la corrélation de X et Y est la valeur*

$$\sigma_{X,Y} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})$$

Cette définition est une généralisation de la définition de la variance , d'ailleurs on peut remarquer que $\sigma_{X,X} = \mathbb{V}\mathrm{ar}(X) = \sigma_X^2$. Et comme pour la variance on a une formule de Koening pour simplifier le calcul de la covariance :

Théorème 2.15 (Koening) *Soient X et Y deux séries statistiques de moyennes $\overline{X}$ et $\overline{Y}$ alors*

$$\sigma_{X,Y} = \overline{XY} - \overline{X}\,\overline{Y}$$

Preuve : La démonstration de cette formule est très similaire à c'autre formule de Koening exprimant la variance d'une série statistique en fonction de la moyenne des carrés :

$$\begin{aligned}
\sigma_{X,Y} &= \overline{(X-\overline{X})(Y-\overline{Y})} \\
&= \overline{XY - \overline{X}Y - X\overline{Y} + \overline{X}\times\overline{Y}} \\
&= \overline{XY} - \overline{X}\times\overline{Y} - \overline{X}\times\overline{Y} + \overline{X}\times\overline{Y} \\
&= \overline{XY} - \overline{X}\times\overline{Y}
\end{aligned}$$

□

On a aussi l'inégalité suivante qui va prendre tout son sens plus loin.

Proposition 2.16 (corrélation) *Soient X et Y deux séries statistiques d'écart-type σ_X et σ_Y alors*

$$|\sigma_{X,Y}| \leq \sigma_X\sigma_Y$$

Preuve : On pose $S = (X - \overline{X})$ et $T = (Y - \overline{Y})$ il est alors évident que

$$\overline{S^2} = \overline{(X-\overline{X})^2} = \mathbb{V}\mathrm{ar}(X),\ \ \overline{T^2} = \overline{(Y-\overline{Y})^2} = \mathbb{V}\mathrm{ar}(Y),\ \ \overline{ST} = \sigma_{X,Y}$$

Maintenant on décide de calculer :

$$\overline{(S+\lambda T)^2} = \overline{S^2 + 2\lambda ST + \lambda^2 T^2} = \overline{S^2} + 2\lambda\overline{ST} + \lambda^2\overline{T^2} = \sigma_X^2 - 2\lambda\sigma_{X,Y} + \lambda^2\sigma_Y^2$$

c'est donc un trinôme du second degré par rapport à la variable λ, mais ce polynôme est toujours positif car $\overline{(S+\lambda T)^2} \geq 0$, donc son discriminant est négatif ou nul :

$$\sigma_X^2 - 2\lambda\sigma_{X,Y} + \lambda^2\sigma_Y^2 \geq 0 \Longrightarrow \sigma_{X,Y}^2 - \sigma_X^2\sigma_Y^2 \leq 0 \Longrightarrow |\sigma_{X,Y}| \leq \sigma_X\sigma_Y$$

□

Maintenant pour calculer l'équation de la "meilleure" droite passant parmi les points du nuage, on va pouvoir utiliser une méthode découverte par Gauss au début du $XIX^{\text{ième}}$ siècle : la méthode des moindres carrés. Cette méthode consiste à minimiser l'erreur globale qu'on comment en écrivant que $Y = aX + b$:

Théorème 2.17 (méthode des moindres carrés) *Soient X et Y deux séries statistiques alors la méthode des moindres carrés consiste à chercher la "meilleure" droite, d'équation $y = ax + b$, passant par le nuage de points (X,Y) comme étant la droite qui minimise la somme des carré des écarts entre les points $(X(i),Y(i))$ et $(X(i),aX(i) + b)$ c'est à dire on cherche a et b qui minimisent la fonction*

$$\phi(a,b) = \sum_{i=1}^{N}(Y(i) - aX(i) - b)^2$$

En utilisant nos connaissances en analyse, on peut maintenant trouver le minimum de la fonction $\phi(a,b)$.

Théorème 2.18 (ajustement linéaire) *Soient X et Y deux séries statistiques alors la droite d'ajustement linéaire de Y en X a pour équation $y = ax + b$ avec :*

$$a = \frac{\sigma_{X,Y}}{\mathbb{V}\mathrm{ar}(X)} \quad \text{et} \quad b = \overline{Y} - a\overline{X}$$

Enfin on peut quantifier la qualité de l'approximation $Y = aX + b$ en utilisant le coefficient de corrélation.

Théorème 2.19 (coefficient de corrélation) *Soient X et Y deux séries statistiques, on suppose que la droite d'ajustement linéaire de Y en X à pour équation $y = ax + b$ et on définit sont coefficient de corrélation par :*

$$\rho = \frac{\sigma_{X,Y}^2}{\mathbb{V}\mathrm{ar}(X)\mathbb{V}\mathrm{ar}(Y)} \in [0,1]$$

alors le nuage de points (X,Y) est d'autant plus proche de la droite d'ajustement linéaire de Y en X que ρ est proche de 1. En particulier tous les points du nuage sont sur la droite $y = ax + b$ si et seulement si $\rho = 1$.

La corrélation sera dite forte (resp. faible) si $\rho \geq \frac{\sqrt{3}}{2} \approx 0.866$ (resp. $\rho < \frac{\sqrt{3}}{2} \approx 0.866$)

Preuve : le but est de trouver 2 équations qui permettent de trouver les a et b qui minimise la fonction $\phi(a,b)$. Pour cela on va minimiser ϕ d'abord par rapport à b et ensuite par rapport à a :

- recherche du minimum par rapport à b

$$\begin{aligned}
\phi(a,b) &= \sum_{i=1}^{N}(Y(i)-(aX(i)+b))^2 \\
&= \sum_{i=1}^{N}((Y(i)-aX(i))-b)^2 \\
&= \sum_{i=1}^{N}(Y(i)-aX(i))^2-2b(Y(i)-aX(i))+b^2 \\
&= N\left(b^2-2b\overline{Y-aX}+\overline{(Y-aX)^2}\right)
\end{aligned}$$

ce trinôme atteint son minimum quand $b=\overline{Y-aX}=\overline{Y}-a\overline{X}$, on peut remarquer que cette équation signifie que la "meilleur droite passe par le point "moyen" $(\overline{X},overlineY)$ puisque $\overline{Y}=a\overline{X}+b$.

- recherche du minimum par rapport à a

$$\begin{aligned}
\phi(a,b) &= \sum_{i=1}^{N}(Y(i)-(aX(i)+b))^2 \\
&= \sum_{i=1}^{N}((Y(i)-aX(i))-(\overline{Y}-a\overline{X}))^2 \\
&= \sum_{i=1}^{N}((Y(i)-\overline{Y})-a(X(i)-\overline{X}))^2 \\
&= \sum_{i=1}^{N}(Y(i)-\overline{Y})^2-2a(Y(i)-\overline{Y})(X(i)-\overline{X})+a^2(X(i)-\overline{X})^2 \\
&= N\left(a^2\overline{(X-\overline{X})^2}-2a\overline{(Y-\overline{Y})(X-\overline{X})}+\overline{(Y-\overline{Y})^2}\right) \\
&= N\left(a^2\mathbb{V}\mathrm{ar}(X)-2a\sigma_{X,Y}+\mathbb{V}\mathrm{ar}(Y)\right)
\end{aligned}$$

ce trinôme atteint son minimum quand $a=\frac{\sigma_{X,Y}}{\mathbb{V}\mathrm{ar}(X)}$

- Pour finir calculons la somme des écarts pour la "meilleure" droite:

$$\begin{aligned}
\phi(a,b) &= N\left(\frac{\sigma_{X,Y}^2}{\mathbb{V}\mathrm{ar}(X)^2}\mathbb{V}\mathrm{ar}(X)-2\frac{\sigma_{X,Y}^2}{\mathbb{V}\mathrm{ar}(X)}-\mathbb{V}\mathrm{ar}(Y)\right) \\
&= N\left(\mathbb{V}\mathrm{ar}(Y)-\frac{\sigma_{X,Y}^2}{\mathbb{V}\mathrm{ar}(X)}\right)=N\mathbb{V}\mathrm{ar}(Y)\left(1-\frac{\sigma_{X,Y}^2}{\mathbb{V}\mathrm{ar}(X)\mathbb{V}\mathrm{ar}(Y)}\right) \\
&= N\mathbb{V}\mathrm{ar}(Y)\left(1-\rho\right)
\end{aligned}$$

$\phi(a,b)$ étant toujours positive (par définition) on doit avoir que $\rho\leq 1$, ce qui est bien vérifié d'après la proposition 2.16. Maintenant on en déduit facilement que $\rho=1\iff\phi(a,b)=0$ ce qui signifie bien que tous les écarts sont nuls si $\rho=1$.

□

✎ **2.9** Reprenons l'exemple des notes du premier semestre :
$M =$« note de maths au S1 »et $N =$« moyenne générale au S1 »
et appliquons la méthode des moindres carrés. On calcule d'abord les différents paramètres :

- $\overline{M} = 9.2500971$ $\mathbb{V}\mathrm{ar}(M) = 9.8811854$ et $\sigma_M = 3.1434353$
- $\overline{N} = 10.636117$ $\mathbb{V}\mathrm{ar}(N) = 5.7622801$ et $\sigma_N = 2.400475$
- $\overline{MN} = 104.37805$

On en déduit l'ajustement de N par rapport à M

- $a = \frac{\sigma_{M,N}}{\mathbb{V}\mathrm{ar}(N)} = \frac{5.9929373}{3.1434353^2}$
- $b = \overline{N} - a\overline{M} = 10.636117 - 0.6064998 \times 9.2500971 = 5.0259342$
- $\rho = \frac{\sigma^2_{M,N}}{\mathbb{V}\mathrm{ar}(M)\mathbb{V}\mathrm{ar}(N)} = \frac{5.9929373^2}{9.8811854 \times 5.7622801} = 0.6307773$ (corrélation faible)

ce qui donne :

$$N = 0.6064998M + 5.0259342$$

qui correspond bien à ce qu'on obtient graphiquement (voir droite rouge sur les figures FIG.6 et FIG.7).

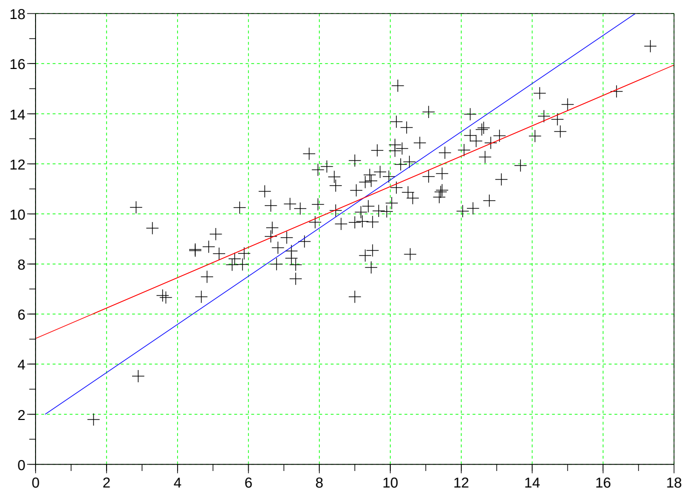


FIG. 7 – *comparaison des ajustements de N par rapport à M (en rouge) et de M par rapport à N (en bleu) avec une corrélation faible* $\rho = 0.6307773$

On aurait pu de la même manière chercher à faire l'ajustement de M par rapport à N. Il suffit pour cela d'échanger les rôles de N et M. On peut remarquer que ce changement ne modifie pas la valeur de la corrélation (car $\sigma_{X,Y} = \sigma_{Y,X}$) ni du coefficient de corrélation $\rho = 0.6307773$ (corrélation faible), on peut donc reprendre ces valeurs pour calculer les nouveaux coefficients de la droite d'ajustement de M par rapport à N :

- $a = \frac{\sigma_{N,M}}{\mathbb{V}\mathrm{ar}(N)} = \frac{5.9929373}{2.400475^2}$

- $b = \overline{M} - a\overline{N} = 9.2500971 - 1.0400288 \times 10.636117 = -1.8117705$

on trouverait donc :

$$M = 1.0400288N + -1.8117705$$

on remarquera que cette relation n'est pas équivalente à celle trouvée en ajustant Y par rapport à X, on peut s'en convaincre facilement sur la figure FIG.7 ! Ceci est lié au fait que la corrélation soit faible. Les deux méthodes donnent la même équation si et seulement si $\rho = 1$.

3 Probabilités

L'étude scientifique des probabilités est relativement récente dans l'histoire des mathématiques. Elle a démarrée au $XVII^{\text{ième}}$ de l'étude de l'aspect aléatoire, *i.e.* en partie imprévisible, de certains phénomènes, en particulier les jeux de hasard. Ceux-ci ont conduit les mathématiciens à développer une théorie qui a ensuite eu des implications dans des domaines aussi variés que la météorologie, la finance ou la chimie.

3.1 Épreuves, événements et probabilité

Dans les faits, la théorie des probabilités s'occupe d'ensembles et de la mesure de ces ensembles. Lorsque l'on considère une expérience aléatoire il faut donc faire le lien entre cette expérience et un ensemble contenant tous les résultats possibles sur lequel on effectuera des mesures. Le fait de créer ces objets mathématiques et de les relier à la réalité c'est faire une *modélisation.*

Définition 3.1 (modélisation d'une expérience aléatoire)
À toute expérience aléatoire on associe un espace probabilisable $(\Omega,\mathscr{T})$ *constitué de :*

- *un ensemble* Ω *contenant autant d'éléments que de résultat de l'expérience et appelé* l'univers *de l'expérience*
- *un autre ensemble* $\mathscr{T}$ *dont les éléments sont appelés des* événements

Si Ω *est un ensemble fini ou dénombrable on peut toujours prendre* $\mathscr{T} = \mathscr{P}(\Omega)$ *l'ensemble des parties de* Ω.

On représente en général la probabilité d'un événement par un nombre réel compris entre 0 et 1 de telle sorte que plus ce nombre est grand, plus le risque (ou la chance, selon le point de vue) que l'événement se produise est grand. Cette notion quelque peu « empirique » a pourtant une définition mathématique plus rigoureuse.

Définition 3.2
une probabilité $\mathbb{P}$ *sur un espace probabilisable* $(\Omega,\mathscr{T})$ *est une application qui vérifie*

- $\mathbb{P} : \mathscr{T} \longrightarrow [0,1]$
- $\mathbb{P}(\Omega) = 1$
- *soit* $I \subset \mathbb{N}$ *alors pour toute suite d'événements* $\{A_n | n \in I\} \subset \mathscr{T}$ *disjoints deux à deux* $(\forall i,j \in I,\ \ A_i \cap A_j = \emptyset$ *si* $i \neq j)$ *on a*

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}(A_0 \cup A_1 \cup \cdots) = \mathbb{P}(A_0) + \mathbb{P}(A_1) + \cdots = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

$(\Omega,\mathscr{T},\mathbb{P})$ *est appelé un espace probabilisé.*

Le dernier axiome est l'axiome d'additivité de $\mathbb{P}$. Dans le cas des ensembles finis on peut le remplacer par la propriété :

$$\forall A,B \in \mathscr{T}, \quad \text{tels que} \quad A \cap B = \emptyset, \quad \text{on a} \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

c'est d'ailleurs sous cette forme qu'on l'utilise le plus souvent

Proposition 3.3 *Soit* $(\Omega,\mathscr{T})$ *un espace probabilisable, avec* Ω *fini ou dénombrable, alors une probabilité* $\mathbb{P}$ *est entièrement déterminée par la donnée de* $\mathbb{P}(\{\omega\}),\forall\omega \in \Omega$

✎ **3.1** reprenons l'exemple 1.2 du lancer d'un dé équilibré. Cette expérience comporte 6 résultats et peut donc être modélisée par l'univers

$$\Omega = \{1;2;3;4;5;6\}$$

pour définir une probabilité sur cet ensemble il faut donner la valeur de $\mathbb{P}(\{k\})$ pour chaque $k = 1,2,\ldots,6$. L'hypothèse du dé équilibré doit nous amener à choisir

$$\forall k = 1,\ldots,6,\ \ \mathbb{P}(\{k\}) = \frac{1}{6}$$

à partir de là on peut calculer la probabilité de n'importe quel événement :

$$A = \{2;4;6\} \Longrightarrow \mathbb{P}(A) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

On retiendra les formules suivantes qui sont vraies pour toute fonction de probabilité.

Théorème 3.4 *Soit* $(\Omega,\mathscr{T},\mathbb{P})$ *un espace probabilisé alors on a :*
- $\mathbb{P}(\emptyset) = 0$
- $\forall A \in \mathscr{T},\ \ \mathbb{P}\left(\overline{A}\right) = 1 - \mathbb{P}(A)$
- $\forall A,B \in \mathscr{T},\ \ A \subset B \Longrightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
- $\forall A,B \in \mathscr{T},\ \ \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ *(Loi de la mesure)*

Preuve :

- pour tout événement A on a $A = A \cup \emptyset$ qui est une réunion d'ensembles disjoints (car $A \cap \emptyset = \emptyset$) donc

$$\mathbb{P}(A) = \mathbb{P}(A) + \mathbb{P}(\emptyset) \Longrightarrow \mathbb{P}(\emptyset) = 0$$

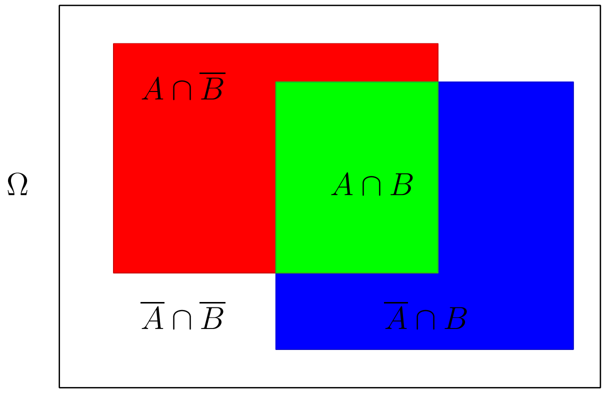
- pour tout événement A on a $\Omega = A \cup \overline{A}$ qui est une réunion d'ensembles disjoints (car $A \cap \overline{A} = \emptyset$) donc

$$\mathbb{P}(\Omega) = \mathbb{P}(A \cup \overline{A}) = \mathbb{P}(A) + \mathbb{P}(\overline{A}) = 1 \Longrightarrow \mathbb{P}\left(\overline{A}\right) = 1 - \mathbb{P}(A)$$

- pour deux événements $A \subset B$ on a $B = A \cup (\overline{A} \cap B)$ qui est une réunion d'ensembles disjoints (car $A \cap (B \cap \overline{A}) = \emptyset$) donc

$$\mathbb{P}(B) = \mathbb{P}(A \cup (\overline{A} \cap B)) = \mathbb{P}(A) + \mathbb{P}(\overline{A} \cap B) \geq \mathbb{P}(A) \ \ \textit{car} \ \ \mathbb{P}(\overline{A} \cap B) \geq 0$$

- quand les choses deviennent plus complexes il faut décomposer les ensembles présents en monômes canoniques, pour manipuler seulement des réunions d'ensembles disjoints, en utilisant un diagramme de Venn :



on en déduit facilement que :

$$A \cup B = (A \cap \overline{B}) \cup (A \cap B) \cup (\overline{A} \cap B) \Longrightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A \cap \overline{B}) + \mathbb{P}(A \cap B) + \mathbb{P}(\overline{A} \cap B)$$

si on compare à ce qu'on a pour A :

$$A = (A \cap \overline{B}) \cup (A \cap B) \Longrightarrow \mathbb{P}(A) = \mathbb{P}(A \cap \overline{B}) + \mathbb{P}(A \cap B)$$

et pour B :

$$B = (A \cap B) \cup (\overline{A} \cap B) \Longrightarrow \mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(\overline{A} \cap B)$$

On comprend facilement comment combiner ces résultats pour obtenir la formule recherchée :

$$\mathbb{P}(A) + \mathbb{P}(B) = \underbrace{\mathbb{P}(A \cap \overline{B}) + \mathbb{P}(\overline{A} \cap B) + \mathbb{P}(A \cap B)}_{\mathbb{P}(A \cup B)} + \mathbb{P}(A \cap B) = \mathbb{P}(A \cup B) + \mathbb{P}(A \cap B)$$

$\square$

3.2 Équiprobabilité

L'hypothèse la plus courante lors de la modélisation d'une expérience aléatoire est l'hypothèse d'*équiprobabilité*. Sous cette hypothèse on admet que tous les résultats envisageables ont la même probabilité d'apparaître. C'est le seul cas où l'on sache calculer exactement la probabilité de chaque événement.

Définition 3.5 (Équiprobabilité)
Soit $(\Omega, \mathscr{T}, \mathbb{P})$ *un espace probabilisé tel que* Ω *est un ensemble fini :*

$$\Omega = \{\omega_1; \omega_2; \ldots; \omega_{n-1}; \omega_n\} \Longrightarrow \mathrm{Card}(\Omega) = n < \infty$$

alors on dit que $\mathbb{P}$ *est l'équiprobabilité sur* Ω *si et seulement si :*

$$\forall k = 1, 2, \ldots, n, \quad \mathbb{P}(\{\omega_k\}) = \frac{1}{n}$$

On bon exemple d'équiprobabilité est celui du lancé d'un dé équilibré (exemple 3.1) où l'univers possède 6 éléments

$$\Omega = \{1; 2; \ldots ; 6\}$$

de probabilité $1/6^{\text{ième}}$ à chaque fois. Pour l'équiprobabilité le calcul des probabilités se ramène donc à des calculs de dénombrement.

Proposition 3.6
Soit $(\Omega, \mathscr{T}, \mathbb{P})$ *un espace probabilisé où* $\mathbb{P}$ *est l'équiprobabilité sur* Ω *alors :*

$$\forall A \subset \mathscr{T}, \quad \mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}$$

Preuve : Comme Ω est un ensemble fini (avec $\text{Card}(\Omega) = n$) c'est aussi le cas pour n'importe quel événement A que l'on peut donc écrire en extension :

$$A = \{\omega_1; \omega_2; \ldots ; \omega_k\} \Longrightarrow \text{Card}(A) = k$$

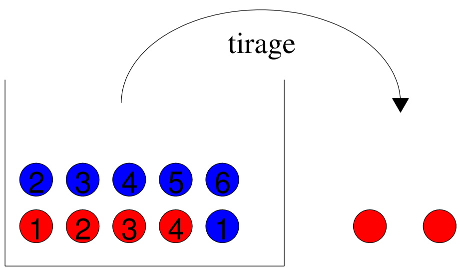
maintenant on peut calculer $\mathbb{P}(A)$ en utilisant l'additivité de $\mathbb{P}$ et l'hypothèse d'équiprobabilité :

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(\{\omega_1; \omega_2; \ldots ; \omega_k\}) \\
&= \mathbb{P}(\{\omega_1\}) + \mathbb{P}(\{\omega_2\}) + \cdots + \mathbb{P}(\{\omega_k\}) \\
&= \frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n} = \frac{k}{n} = \frac{\text{Card}(A)}{\text{Card}(\Omega)}
\end{aligned}$$

$\square$

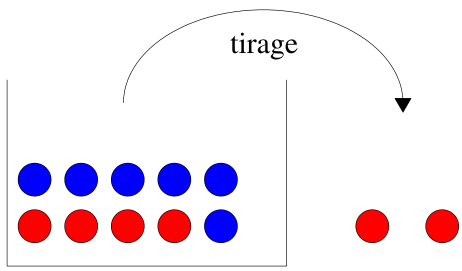
✎ **3.2** On considère l'expérience :
« tirage de deux boules simultanément parmi 4 boules rouges et 6 bleus »
et on veut calculer la probabilité de l'événement A :« obtenir 2 boules rouges »
on a alors 2 manières de modéliser l'expérience :

Considérer les boules discernables



on a 10 boules
$F = \{R1; \ldots R4; B1; \ldots B6\}$
et $\Omega = \{E \subset F | \text{Card}(E) = 2\}$
$\Longrightarrow \text{Card}(\Omega) = C_{10}^2 = 45$

Considérer les boules indiscernables



on a que 2 types de boules
$F = \{R; B\}$
et $\Omega = \{RR; RB; BB\}$
$\Longrightarrow \text{Card}(\Omega) = 3$

Dans la première représentation on peut clairement supposer l'équiprobabilité des tirages de Ω et donc :

$$\mathbb{P}(\{E\}) = \frac{1}{45} \quad \forall E \in \Omega$$

ce qui permet de calculer la probabilité de l'événement A :

$$A = \{B \subset \{R1;\ldots;R4\} | \text{Card}(B) = 2\} \Longrightarrow \mathbb{P}(A) = \frac{C_4^2}{45} = \frac{2}{15}$$

la présence ou l'absence de numéro sur les boules ne change en rien la probabilité d'obtenir 2 boules rouges, la probabilité de l'événement A doit donc être la même dans les 2 représentations :

$$\mathbb{P}(\{BB\}) = \mathbb{P}(A) = \frac{2}{15} \neq \frac{1}{3}$$

en particulier on en déduit qu'on ne peut pas avoir équiprobabilité dans la seconde représentation!

On retiendra de cet exemple la règle suivante :

Il faut toujours modéliser l'expérience en considérant que tous les objets sont discernables de telle sorte qu'on puisse utiliser l'équiprobabilité pour les calculs.

3.3 Probabilités conditionnelles

L'*indépendance* est une notion probabiliste qualifiant de manière intuitive des événements aléatoires n'ayant aucune influence l'un sur l'autre. Cette notion qui semble pourtant intuitive nécessite une définition mathématique très précise.

Définition 3.7 (indépendance) *Soient A et B deux événements alors*

$$A \text{ et } B \text{ indépendants} \Longleftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$$

Contrairement à ce qu'on pourrait penser il n'est évident de détecter une dépendance entre deux événement sans passer par le calcul.

3.3 reprenons encore une fois l'expérience 1.2 du lancer d'un dé équilibré. Cette expérience peut être modélisée par l'univers $\Omega = \{1;2;3;4;5;6\}$ muni de l'équiprobabilité $\forall k = 1,\ldots,6,\ \ \mathbb{P}(\{k\}) = \frac{1}{6}$. On considère les événements

$A =$« les résultats pair » $B =$« les résultats supérieur ou égal à trois »

pour savoir si ces événements sont indépendants il faut identifier $A \cap B$

$$A = \{2;4;6\} \ \ et \ \ B = \{3;4;5;6\} \Longrightarrow A \cap B = \{4;6\}$$

et calculer les probabilités

$$\mathbb{P}(A) = \frac{1}{2} \quad \mathbb{P}(B) = \frac{2}{3} \quad \mathbb{P}(A \cap B) = \frac{1}{3}$$

donc $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$ et A et B sont indépendants. C'est assez logique puisqu'il y a autant de résultats pairs et de résultats impairs dans B donc la réalisation de A ne donne aucune information sur la réalisation de B. Par contre si on considère les événements

$A =$« les résultats pairs » $C =$« les résultats strictement supérieur à trois » les mêmes calculs

$$A = \{2; 4; 6\} \quad et \quad C = \{4; 5; 6\} \Longrightarrow A \cap C = \{4; 6\}$$

conduisent à un résultat différent

$$\mathbb{P}(A) = \frac{1}{2} \quad \mathbb{P}(C) = \frac{1}{2} \quad \mathbb{P}(A \cap C) = \frac{1}{3}$$

donc $\mathbb{P}(A \cap C) \neq \mathbb{P}(A) \times \mathbb{P}(C)$ et A et C sont dépendants. Ce résultat est normal puisqu'il y a plus de résultats pairs que impairs dans C, donc savoir que A est réalisé indique qu'on a "plus" de chances de voir C réalisé.

Il ne faut pas confondre indépendance et incompatibilité qui sont deux choses très différentes :

- A et B indépendants $\Longleftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$
- A et B incompatibles $\Longleftrightarrow A \cap B = \emptyset \Longrightarrow \mathbb{P}(A \cap B) = 0$

d'ailleurs si A et B sont incompatibles et de probabilité non nulle il ne *sont pas* indépendants puisque $\mathbb{P}(A) \times \mathbb{P}(B) \neq 0 = \mathbb{P}(A \cap B)$

Quand on considère plus de 2 événements, l'indépendance 2 à 2 ne garantie pas l'indépendance d'un des événements par rapport à l'ensemble des autres événements. C'est pourquoi on doit modifier la définition d'indépendance dans ce cas.

Définition 3.8 (événements totalement indépendants)
On dit que n événements A_1,A_2,$\ldots A_n$ sont totalement indépendants (ou indépendant dans leur ensemble) si on a l'indépendance des événements pris 2 à 2, 3 à 3,...,n à n, ce qui donne :

- $\forall i \neq j, \ \mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \times \mathbb{P}(A_j)$
- $\forall \{i; j; k\} \subset \{1; \ldots; n\}, \ \mathbb{P}(A_i \cap A_j \cap A_k) = \mathbb{P}(A_i) \times \mathbb{P}(A_j) \times \mathbb{P}(A_k)$
- $\ldots$
- $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \times \mathbb{P}(A_2) \times \cdots \times \mathbb{P}(A_n)$

Pour trois événements A,B,C il faudra donc vérifier que :

- $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$
- $\mathbb{P}(B \cap C) = \mathbb{P}(B) \times \mathbb{P}(C)$
- $\mathbb{P}(C \cap A) = \mathbb{P}(C) \times \mathbb{P}(A)$
- $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \times \mathbb{P}(B) \times \mathbb{P}(C)$ (indépendance 3 à 3)

pour $n = 4$ il y aurait 4 indépendances 3 à 3 à vérifier et il faudrait ajouter l'indépendance des événements 4 à 4, et ainsi de suite... Donc plus n augmente plus il y a de conditions à vérifier, on peut même montrer que pour n événements il y aura $2^n - n - 1$ conditions à vérifier!

Quand 2 événements ne sont pas indépendants on introduit la notion de probabilités conditionnelles d'un des événements par rapport à l'autre.

Définition 3.9 (probabilité conditionnelle) *Soient A et B deux événements alors on appelle probabilité conditionnelle de A sachant B*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \text{« probabilité de } A \ \underline{\text{sachant}} \ B \text{ »}$$

Lorsque deux événements sont indépendants la probabilité conditionnelle de l'un par rapport à l'autre se ramène à la probabilité normale (puisqu'il n'y a pas d'apport d'information).

Théorème 3.10 *Soient A et B deux événements non vides alors*

$$A \text{ et } B \text{ indépendants} \Longleftrightarrow \mathbb{P}(A|B) = \mathbb{P}(A)$$

Preuve : Il s'agit d'une équivalence on doit donc montrer les deux équivalences :
$\Longrightarrow$ si A et B sont indépendants alors on a $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$ et

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A) \times \mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$$

$\Longleftarrow$ inversement si

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A) \Longrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$$

donc A et B sont indépendants.

□

On va maintenant voir comment certaines formules des probabilités conditionnelles vont nous aider à calculer d'autres probabilités.

Théorème 3.11 (des probabilités composées)
Soit A_1,A_2, … A_{n-1},A_n des événements alors

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_n|A_{n-1} \cap \cdots \cap A_2 \cap A_1) \times \cdots \times \mathbb{P}(A_2|A_1) \times \mathbb{P}(A_1)$$

Preuve :

- On note $\mathscr{P}_n$ la proposition $\forall A_1,A_2,\ldots A_{n-1},A_n \in \mathscr{T}$:

  $$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_n|A_{n-1} \cap \cdots \cap A_2 \cap A_1) \times \cdots \times \mathbb{P}(A_2|A_1) \times \mathbb{P}(A_1)$$

- **rang initial:** c'est $\mathscr{P}_2$ ici qui découle directement de la définition 3.9

  $$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2|A_1) \times \mathbb{P}(A_1) \Longleftrightarrow \mathbb{P}(A_2|A_1) = \frac{\mathbb{P}(A_2 \cap A_1)}{\mathbb{P}(A_1)}$$

- **passage** $\mathscr{P}_n \Longrightarrow \mathscr{P}_{n+1}$ se fait de la même manière en considérant les événements A_{n+1} et $B = A_n \cap A_{n-1} \cap \cdots \cap A_2 \cap A_1$ on peut écrire :

  $$\mathbb{P}(A_{n+1} \cap B) = \mathbb{P}(A_{n+1}|B) \times \mathbb{P}(B)$$

  il ne reste plus qu'à remplacer B par son expression en fonction des A_i et utiliser l'hypothèse de récurrence :

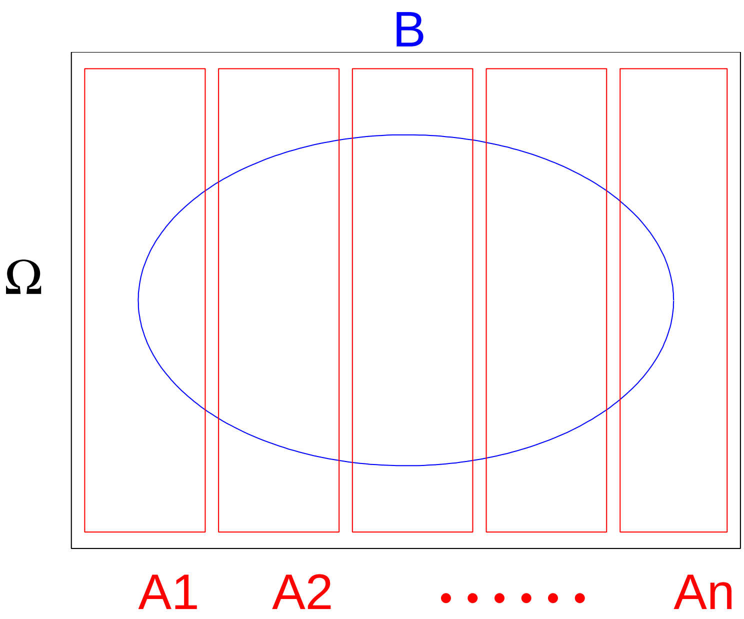
$$\begin{aligned}
\mathbb{P}(A_{n+1} \cap B) &= \mathbb{P}(A_{n+1}|B) \times \mathbb{P}(B) \\
&= \mathbb{P}(A_{n+1}|A_n \cap \cdots \cap A_2 \cap A_1) \times \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) \\
&= \mathbb{P}(A_{n+1}|A_n \cap \cdots \cap A_2 \cap A_1) \times \mathbb{P}(A_n|A_{n-1} \cap \cdots \cap A_2 \cap A_1) \\
&\quad \times \cdots \times \mathbb{P}(A_3|A_2 \cap A_1) \times \mathbb{P}(A_2|A_1) \times \mathbb{P}(A_1)
\end{aligned}$$

□

Théorème 3.12 (des probabilités totales) *Soit $A_1, A_2, \dots A_{n-1}, A_n$ des événements formant une partition de Ω et B un autre événement alors*

$$\mathbb{P}(B) = \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2) + \cdots + \mathbb{P}(B|A_n)\mathbb{P}(A_n) = \sum_{k=1}^{n} \mathbb{P}(B|A_k) \times \mathbb{P}(A_k)$$

Preuve : Le plus simple est de partir du diagramme de Venn :



comme les ensembles A_i forment une partition de Ω les ensembles $B \cap A_i$ forment une partition de B

$$B = B \cap (A_1 \cup A_2 \cup \cdots \cup A_n) = (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n)$$

en particulier c'est une réunion d'ensembles disjoints ce qui permet d'écrire :

$$\mathbb{P}(B) = \mathbb{P}(B \cap A_1) + \mathbb{P}(B \cap A_2) + \cdots + \mathbb{P}(B \cap A_n)$$

ensuite, d'après le théorème 3.11, on a

$$\forall k = 1, \dots n, \;\; \mathbb{P}(B \cap A_k) = \mathbb{P}(B|A_k) \times \mathbb{P}(A_k)$$

donc

$$\mathbb{P}(B) = \mathbb{P}(B|A_1) \times \mathbb{P}(A_1) + \mathbb{P}(B|A_2) \times \mathbb{P}(A_2) + \cdots + \mathbb{P}(B|A_n) \times \mathbb{P}(A_n)$$

□

Pour des raisons de causalité on connaît en général la probablités des « effets »en fonctions des « causes »mais pour les algorithmes d'auto-apprentissage il faut essayer de calculer les probabilités des « causes »en fonctions des « effets ». C'est ce que permet de faire la formule de Bayes.

Théorème 3.13 (des probabilités à posteriori-Formule de Bayes)
Soient A et B deux événements alors $\mathbb{P}(B|A) = \mathbb{P}(A|B)\dfrac{\mathbb{P}(B)}{\mathbb{P}(A)}$

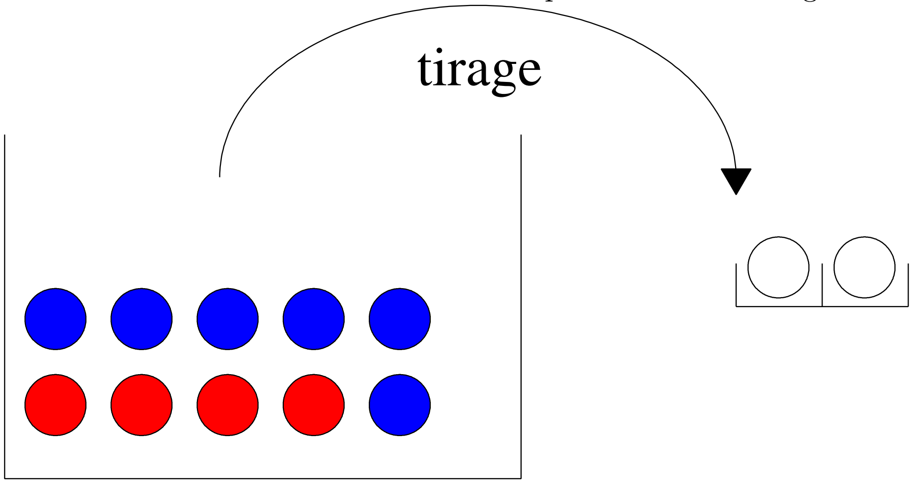
Preuve : Il s'agit d'une application directe de la définition 3.9

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B) \times \mathbb{P}(B)}{\mathbb{P}(A)}$$

□

Dans la pratique les probabilités conditionnelles apparaissent naturellement lors qu'il y a un enchaînement d'événements. Dans ce genre de situation il est souvent pratique d'utiliser un arbre pour représenter les différentes alternatives possibles.

✎ **3.4** Considérons l'expérience suivante :
« tirer 2 boules successivement et sans remises parmi 4 boules rouges et 6 bleus »



et on considère les événements

- R_i :« obtenir une boule rouge au $i^{\text{ième}}$ tirage »
- B_i :« obtenir une boule bleu au $i^{\text{ième}}$ tirage »

En supposant l'équiprobabilité à chaque tirage on calcule directement et sans difficultés les probabilités suivantes

- au départ il y a 6 bleus et 4 rouges
- $\mathbb{P}(R_1) = \frac{4}{10} = \frac{2}{5}$
- $\mathbb{P}(B_1) = \frac{6}{10} = \frac{3}{5}$

- après R_1 il reste 6 bleus et 3 rouges
- $\mathbb{P}(R_2|R_1) = \frac{6}{9} = \frac{2}{3}$
- $\mathbb{P}(B_2|R_1) = \frac{3}{9} = \frac{1}{3}$

- après B_1 il reste 5 bleus et 4 rouges
- $\mathbb{P}(R_2|B_1) = \frac{5}{9}$
- $\mathbb{P}(B_2|B_1) = \frac{4}{9}$

on peut alors placer ces probabilités sur l'arbre FIG.8. On apprendra à visualiser les différents calculs sur un arbre :

- la formule des probabilités enchaînées correspondent au produit des probabilités le long d'une branche de l'arbre

$$\mathbb{P}(\mathbb{R}_1 \cap \mathbb{R}_2) = \mathbb{P}(\mathbb{R}_2|\mathbb{R}_1)\mathbb{P}(R_1) = \frac{1}{3} \times \frac{2}{5} = \frac{2}{15}$$

- la formule des probabilités totales correspond à la somme des probabilités de plusieurs branches

$$\begin{aligned}\mathbb{P}(R_2) &= \mathbb{P}(R_1 \cap R_2) + \mathbb{P}(B_1 \cap R_2)\\ &= \mathbb{P}(R_2|R_1)\mathbb{P}(R_1) + \mathbb{P}(R_2|B_1)\mathbb{P}(B_1)\\ &= \frac{2}{15} + \frac{4}{9} \times \frac{3}{5} = \frac{6}{15} = \frac{2}{5}\end{aligned}$$

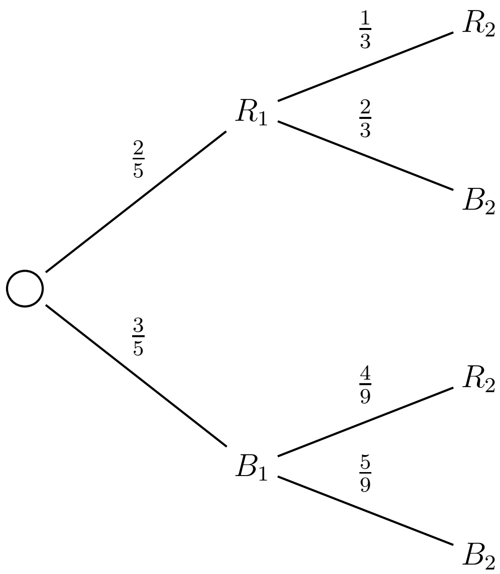


FIG. 8 – *arbre des probabilités pour 2 tirages successifs sans remises*

- la formule des probabilités a posteriori revient à exprimer la probabilité des cause (1^{er} niveau de l'arbre) en fonction des effets ($2^{\text{ième}}$ niveau)

$$\mathbb{P}(R_1|R_2) = \mathbb{P}(R_2|R_1)\frac{\mathbb{P}(R_1)}{\mathbb{P}(R_2)} = \frac{1}{3} \times \frac{2/5}{2/5} = \frac{1}{3}$$

4 Variables aléatoires discrètes

Dès que le résultat d'une expérience aléatoire peut être quantifié il est pratique d'utiliser une variable aléatoire. Historiquement cette notion est née de considération sur les jeux de hasard (dues en particulier à Blaise Pascal et Pierre de Fermat) où l'on avait besoin d'étudier le gain (= somme empochée - mise) envisageable selon la stratégie adoptée.

4.1 Loi, fonction de répartition et graphiques

De manière informelle, une variable aléatoire réelle est une variable numérique qui dépend d'une expérience soumise au hasard (gain lors d'un jeu de loterie par exemple). À chaque résultat de l'expérience on veut associer un nombre, mathématiquement il est donc naturel de définir une variable aléatoire comme une relation entre l'univers des possibles et un ensemble de nombres.

Définition 4.1 (Variable aléatoire) *Soit* $(\Omega,\mathscr{T},\mathbb{P})$ *un espace probabilisé, on appelle variable aléatoire une application de* Ω *dans* $\mathbb{R}$ *:*

$$X : \begin{array}{ccc} \Omega & \longrightarrow & \mathbb{R} \\ \omega & \longmapsto & X(\omega) \end{array}$$

Cette application crée un nouvel univers $X(\Omega)$ de réels sur lequel on peut construire une probabilité issue de $\mathbb{P}$. Dans la pratique on oublie l'univers Ω pour ne s'intéresser qu' à l'univers $X(\Omega)$.

Définition 4.2 (univers image) *Soit* X *une v.a. sur l'espace probabilisé* $(\Omega,\mathscr{T},\mathbb{P})$ *alors l'univers image de* X *est :*

$$X(\Omega) = \{X(\omega)|\omega \in \Omega\} \subset \mathbb{R}$$

Si $X(\Omega)$ *est fini ou dénombrable on dit que* X *est une variable aléatoire discrète. L'événement correspondant à « la variable* X *vaut* k *»sera notée :*

$$\forall k \in X(\Omega),\ \ [X = k] = X^{-1}(\{k\}) = \{\omega \in \Omega|X(\omega) = k\}$$

La nouvelle probabilité ainsi construite sur $X(\Omega)$ s'appelle loi de probabilité de X.

Définition 4.3 (loi d'une v.a.) *Soit* X *une v.a. sur l'espace probabilisé* $(\Omega,\mathscr{T},\mathbb{P})$ *la loi de* X *est la donnée des probabilités d'apparition de chaque résultat de* X *:*

$$\mathbb{P}(X = k),\ \ \forall k \in X(\Omega)$$

Une autre manière de définir la loi d'une variable aléatoire est de considérer la fonction de répartition.

Définition 4.4 (fonction de répartition) *Soit* X *une v.a. sur l'espace probabilisé* $(\Omega,\mathscr{T},\mathbb{P})$ *la fonction de répartition est définie par :*

$$F : \begin{array}{ccc} \mathbb{R} & \longrightarrow & [0,1] \\ t & \longmapsto & \mathbb{P}(X \leq t) \end{array}$$

Toutes les fonctions ne sont pas des fonctions de répartition, car une fonction de répartitions vérifie certaines propriétés particulières, listées dans la proposition suivante.

Proposition 4.5 *Soit F la fonction de répartition d'une v.a. X alors*

- *F est croissante*
- *F est continue à droite*
- $\lim_{t\to-\infty} F(t) = 0$
- $\lim_{t\to+\infty} F(t) = 1$
- $\mathbb{P}(a < X \leq b) = F(b) - F(a)$

si X est une variable discrète alors F est constante par morceau.

Preuve :

- si $s \leq t$ alors $[X \leq s] \subset [X \leq t]$ et d'après le théorème 3.4 la fonction F est bien croissante:

$$F(s) = \mathbb{P}([X \leq s]) \leq \mathbb{P}([X \leq t]) = F(t)$$

- pour montrer que F est continue à droite il faut considérer les événements $[X \leq s]$ pour $s \geq t$, on a alors

$$\lim_{s\to t^+} [X \leq s] = [X \leq t] \Longrightarrow \lim_{s\to t^+} F(s) = \lim_{s\to t^+} \mathbb{P}([X \leq s]) = \mathbb{P}([X \leq t]) = F(t)$$

  par contre ça ne marche pas si on prend la limite à gauche (par valeur inférieures) car

$$\lim_{s\to t^-} [X \leq s] = [X < t] \neq [X \leq t]$$

- $\lim_{t\to-\infty} F(t) = \lim_{t\to-\infty} \mathbb{P}(X \leq t) = \mathbb{P}(\emptyset) = 0$
- $\lim_{t\to+\infty} F(t) = \lim_{t\to+\infty} \mathbb{P}(X \leq t) = \mathbb{P}([X \in \mathbb{R}]) = 1$
- si on considère la réunion d'ensembles disjoints :

$$[X \leq b] = [X \leq a] \cup [a < X \leq b]$$

  alors

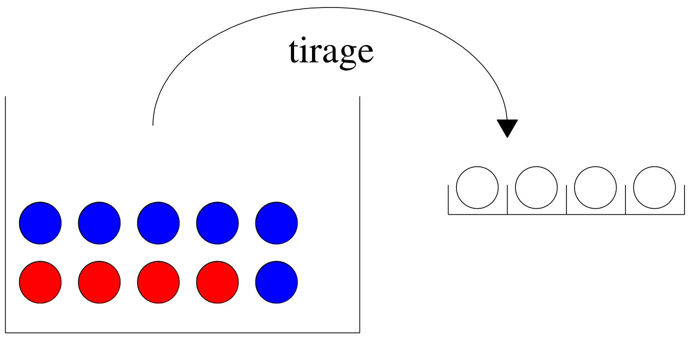
$$\begin{aligned} \mathbb{P}([X \leq b]) &= \mathbb{P}([X \leq a]) + \mathbb{P}([a < X \leq b]) \\ \Longrightarrow \mathbb{P}(a < X \leq b) &= \mathbb{P}([X \leq b]) - \mathbb{P}([X \leq a]) \\ &= F(b) - F(a) \end{aligned}$$

□

✎ 4.1 étude d'une variable aléatoire

On considère l'expérience « tirer 4 boules parmi 4 blanches et 6 bleus »
étudier la v.a. définie par $X =$« nombre de boules rouges obtenues »

Il s'agit d'étudier un ensemble de tirages simultanés (sans remise)



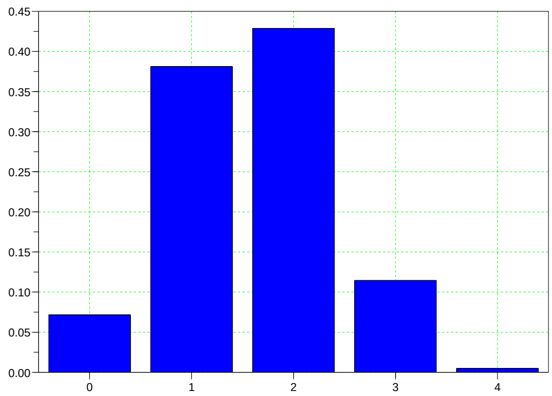
si on considère les boules discernables on a donc l'équiprobabilité sur un univers Ω tel que $\text{Card}(\Omega) = C_{10}^4 = 210$. Le nombre de boules rouges obtenues peut varier de 0 à 4, on calcule ensuite facilement le les probabilités pour chaque cas (résultats arrondis à 10^{-3} près ici):

k	0	1	2	3	4
$\mathbb{P}(X = k)$	0.071	0.381	0.429	0.114	0.005

Pour éviter une erreur dans la suite on prendra soin de vérifier que la somme des probabilités sur cette ligne vaut bien 1 :

$$\sum_{k=0}^{4} \mathbb{P}(X = k) = 0.071 + 0.381 + 0.429 + 0.114 + 0.005 = 1$$

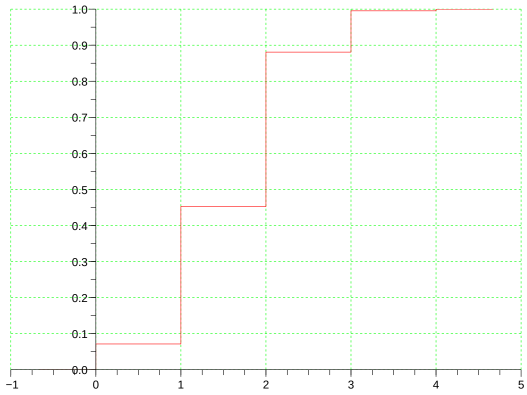
et on pourra représenter graphiquement cette loi par un diagramme en bâton :



on voit que le cas le plus probable est $X = 2$ et que les cas $X = 4$ et $X = 0$ sont très peu probables. La représentation de la fonction de répartition est plus difficile, il faut d'abord faire le tableau du cumul des probabilités qui apparaissent dans la loi de X :

$t \in I$	$]-\infty,0[$	$[0,1[$	$[1,2[$	$[2,3[$	$[3,4[$	$[4,+\infty[$
$F(t)$	0	0.071	0.452	0.881	0.995	1

chacune de ces probabilités correspond à la valeur de F sur un intervalle (ouvert à droite), on peut ensuite tracer le graphe de cette fonction.



4.2 Moyenne, variance et écart-type

Comme pour les variables statistiques on va définir deux indicateurs permettant de se représenter rapidement la loi d'une v.a.. Le premier de ces indicateurs est ce qu'on appelle la moyenne.

Définition 4.6 (Espérance) *Soit X une v.a. alors la moyenne de X, notée $\mathbb{E}(X)$ est la valeur :*

$$\mathbb{E}(X) = \sum_{k \in X(\Omega)} k\mathbb{P}(X = k)$$

plus généralement, pour fonction $f : \mathbb{R} \longrightarrow \mathbb{R}$ on pourra écrire

$$\mathbb{E}(f(X)) = \sum_{k \in X(\Omega)} f(k)\mathbb{P}(X = k)$$

Un cas particulier est celui des fonctions constantes, par exemple pour $f(x) = 1$:

$$\mathbb{E}(f(X)) = \mathbb{E}(1) = \sum_{k \in X(\Omega)} f(k)\mathbb{P}(X = k) = \sum_{k \in X(\Omega)} \mathbb{P}(X = k) = 1$$

Pour des combinaisons de plusieurs variables on a la formule suivante.

Proposition 4.7 (linéarité de $\mathbb{E}$)
Soit X,Y des v.a. et a,$b \in \mathbb{R}$ alors

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

en particulier $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.

Preuve : La démonstration de la deuxième formule est plus simple et permet peut être de mieux comprendre la première démonstration. Il suffit d'écrire l'espérance comme une somme et d'utiliser la linéarité de la somme :

$$\begin{aligned}\mathbb{E}(aX+b) &= \sum_{k\in X(\Omega)}(ak+b)\mathbb{P}([X=k])\\ &= a\left(\sum_{k\in X(\Omega)}k\mathbb{P}([X=k])\right)+b\left(\sum_{k\in X(\Omega)}\mathbb{P}([X=k])\right)\\ &= a\mathbb{E}(X)+b\times 1 = a\mathbb{E}(X)+b\end{aligned}$$

Pour la première formule les choses sont plus complexes car il y a une double sommation (une par rapport à chaque variable) et il faut comprendre que :

$$\begin{aligned}\sum_{q\in Y(\Omega)}\mathbb{P}([X=k]\cap[Y=q]) &= \mathbb{P}([X=k]\cap[Y=q_1])+\cdots+\mathbb{P}([X=k]\cap[Y=q_n])\\ &= \mathbb{P}([X=k])\end{aligned}$$

car si $Y(\Omega)=\{q_1;\ldots;q_n\}$ alors

$$\begin{aligned}[X=k] &= [X=k]\cap\Omega=[X=k]\cap([Y=q_1]\cup\cdots\cup[Y=q_n])\\ &= ([X=k]\cap([Y=q_1])\cup\cdots\cup([X=k]\cap[Y=q_n])\end{aligned}$$

qui est une décomposition en réunion d'ensembles disjoints de l'ensemble $[X=k]$. De même on a

$$\begin{aligned}\sum_{k\in X(\Omega)}\mathbb{P}([X=k]\cap[Y=q]) &= \mathbb{P}([X=k_1]\cap[Y=q])+\cdots+\mathbb{P}([X=k_n]\cap[Y=q])\\ &= \mathbb{P}([Y=q])\end{aligned}$$

ceci permet de simplifier la somme définissant $\mathbb{E}(aX+bY)$:

$$\begin{aligned}\mathbb{E}(aX+bY) &= \sum_{k\in X(\Omega)}\sum_{q\in Y(\Omega)}(ak+bq)\mathbb{P}([X=k]\wedge[Y=q])\\ &= \sum_{k\in X(\Omega)}ak\left(\sum_{q\in Y(\Omega)}\mathbb{P}([X=k]\wedge[Y=q])\right)\\ &\quad+\sum_{q\in Y(\Omega)}ak\left(\sum_{k\in X(\Omega)}\mathbb{P}([X=k]\wedge[Y=q])\right)\\ &= a\left(\sum_{k\in X(\Omega)}k\mathbb{P}([X=k])\right)+b\left(\sum_{q\in Y(\Omega)}k\mathbb{P}([X=k])\right)\\ &= a\mathbb{E}(X)+b\mathbb{E}(Y)\end{aligned}$$

□

attention pour deux variables aléatoires X et Y quelconques en général on a

$$\mathbb{E}(XY)\neq\mathbb{E}(X)\mathbb{E}(Y)$$

En effet, prenons le cas d'une variable X associée à un jeu de pile ou face

$$X(\Omega) = \{0;1\} \quad et \quad \mathbb{E}(X) = 1/2$$

et prenons $Y = 1 - X$ alors $\mathbb{E}(Y) = 1 - 1/2 = 1/2 = \mathbb{E}(X)$ mais

$$\mathbb{E}(XY) = 0 \times (1-0)\mathbb{P}(X=0) + 1 \times (1-1)\mathbb{P}(X=1) = 0 \neq \mathbb{E}(X)\mathbb{E}(Y)$$

ceci est du au fait que les deux variables X et Y ne sont pas indépendantes.

Comme pour les variables statistiques on à aussi un indicateur de dispersion.

Définition 4.8 (variance et écart-type) *Soit X une v.a. de moyenne $\mu = \mathbb{E}(X)$ alors la variance de X, notée $\mathbb{V}\mathrm{ar}(X)$, est la quantité :*

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}\left((X-\mu)^2\right) = \sum_{k \in X(\Omega)} (k-\mu)^2 \mathbb{P}(X=k)$$

et on note $\sigma_X = \sqrt{\mathbb{V}\mathrm{ar}(X)}$ l'écart-type de X.

L'écart-type est bien définit car comme $\forall x \in \mathbb{R}, \quad f(x) = (x-\mu)^2 \geq 0$ la variance est forcément positive :

$$\mathbb{V}\mathrm{ar}(X) = \sum_{k \in X(\Omega)} (k-\mu)^2 \mathbb{P}(X=k) \geq 0$$

L'écart-type permet de comprendre comment les valeurs prises par X sont réparties autour de l'espérance $\mathbb{E}(X)$. En général on ne calcule pas la variance d'une v.a. en utilisant la définition précédente mais en utilisant la formule de Koening.

Théorème 4.9 (Koening)
Soit X une v.a. alors

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2$$

Preuve : On a que $(X-\mu)^2 = X^2 - 2\mu X + \mu^2$ donc en utilisant la linéarité de l'espérance on obtient

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(X) &= \mathbb{E}\left((X-\mu)^2\right) \\
&= \mathbb{E}\left(X^2 - 2\mu X + \mu^2\right) \\
&= \mathbb{E}\left(X^2\right) - 2\mu\mathbb{E}(X) + \mu^2 \\
&= \mathbb{E}\left(X^2\right) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 \quad \text{car } \mu = \mathbb{E}(X) \\
&= \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2
\end{aligned}$$

□

Comme pour l'espérance on a une formule pour calculer l'espérance de la v.a. $aX + b$ à partir de celle de X.

Proposition 4.10 (non-linéarité de $\mathbb{V}\mathrm{ar}$) *Soit X une v.a. et $a,b \in \mathbb{R}$ alors $\mathbb{V}\mathrm{ar}(aX+b) = a^2\mathbb{V}\mathrm{ar}(X)$.*

Preuve : comme $\mathbb{E}(X+b) = \mathbb{E}(X) + b$

$$\begin{aligned}\mathbb{V}\mathrm{ar}(X+b) &= \mathbb{E}\left((X+b-\mathbb{E}(X+b))^2\right) \\ &= \mathbb{E}\left((X+b-(\mathbb{E}(X)+b))^2\right) \\ &= \mathbb{E}\left((X-\mathbb{E}(X))^2\right) \\ &= \mathbb{V}\mathrm{ar}(X)\end{aligned}$$

de même comme $\mathbb{E}(aX) = a\mathbb{E}(X)$

$$\begin{aligned}\mathbb{V}\mathrm{ar}(aX) &= \mathbb{E}\left((aX-\mathbb{E}(aX))^2\right) \\ &= \mathbb{E}\left((aX-a\mathbb{E}(X))^2\right) \\ &= \mathbb{E}\left(a^2(X-\mathbb{E}(X))^2\right) \\ &= a^2\mathbb{E}\left((X-\mathbb{E}(X))^2\right) \\ &= a^2\mathbb{V}\mathrm{ar}(X)\end{aligned}$$

enfin en combinant les 2 résultats on a bien

$$\mathbb{V}\mathrm{ar}(aX+b) = \mathbb{V}\mathrm{ar}(aX) = a^2\mathbb{V}\mathrm{ar}(X)$$

□

attention pour deux variables aléatoires X et Y quelconques en général on a

$$\mathbb{V}\mathrm{ar}(X+Y) \neq \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y)$$

4.2 calculs avec des variables aléatoires

On considère l'expérience « tirer 4 boules parmi 4 rouges et 6 bleus »
calculer la moyenne et l'écart-type des variables définies par :

- $X =$« nombre de boules rouges obtenues »
- $Y =$« nombre de boules bleus obtenues »

pour X on peut repartir de la loi de cette variable aléatoire

k	0	1	2	3	4
$\mathbb{P}(X=k)$	0.071	0.381	0.429	0.114	0.005

pour calculer la moyenne

$\mathbb{E}(X) = 0\times0.0714286+1\times0.3809524+2\times0.4285714+3\times0.1142857+4\times0.0047619 = 1.6$

puis la moyenne des carrés

$\mathbb{E}(X^2) = 0^2\times0.0714286+1^2\times0.3809524+2^2\times0.4285714+3^2\times0.1142857+4^2\times0.0047619 = 3.2$

ce qui permet de calculer la variance d'après la formule de Koening

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 3.2 - 1.6^2 = 0.64$$

d'où l'on déduit l'écart-type $\sigma_X = \sqrt{0.64} = 0.8$.

On pourrait recommencer les mêmes calculs pour Y mais il est beaucoup plus efficace de remarquer que $Y = 4 - X$ (car on tire 4 boules en tout) dans ce cas on obtient directement :

$$\mathbb{E}(Y) = \mathbb{E}(4-X) = 4 - \mathbb{E}(X) = 4 - 1.6 = 2.4$$

et

$$\mathbb{V}\mathrm{ar}(Y) = \mathbb{V}\mathrm{ar}(4-X) = (-1)^2\mathbb{V}\mathrm{ar}(X) = 0.64 \Longrightarrow \sigma_Y = \sqrt{0.64} = 0.8$$

4.3 Lois usuelles

Il existe un certain nombre de situation que l'on retrouve très fréquemment en théorie des probabilités. On va donc définir des variables aléatoires spécifiques pour ces types d'expériences, et quand l'une des expériences à modéliser se ramènera à l'une des situations listées ci-après on pourra utiliser directement les résultats de cette section pour calcules les probabilités, moyennes, variances dont on aura besoin.

La première de ces lois est la loi uniforme.

Définition 4.11 (loi uniforme)
On dit qu'une v.a. X suit une loi uniforme de paramètres a et b si c'est une v.a. d'univers image $\{a; a+1; \ldots ; b\}$ de cardinal $n = 1 + b - a$ vérifiant :

$$\forall k = a, a+1, \ldots, b, \ \ \mathbb{P}(X = k) = \frac{1}{n}$$

L'exemple le plus simple de loi uniforme est le lancer d'un dé *équilibré*.

Proposition 4.12 *On considère l'expérience :*
« lancer un dé (équilibrée) et on note X le résultat de la face supérieure »
alors X suit une loi uniforme de paramètres $a = 1$ et $b = 6$.

on a les formules suivantes pour l'espérance et la variance.

Théorème 4.13 (moment d'une loi uniforme)
Soit X une v.a. qui suit une loi uniforme de paramètres $a,b \in \mathbb{Z}$ (avec $n = 1 + b - a$) alors

$$\mathbb{E}(X) = \frac{a+b}{2} \quad \text{et} \quad \mathbb{V}\text{ar}(X) = \frac{n^2-1}{12}$$

Preuve : Pour la moyenne les calculs reposent sur la formule de la série arithmétique $\sum_{k=0}^{n} k = \frac{n(n+1)}{2}$:

$$\mathbb{E}(X) = \sum_{k=a}^{b} k \times \frac{1}{n} = \sum_{q=0}^{n-1} \frac{a+q}{n} = a + \frac{n(n-1)}{2n} = a + \frac{b-a}{2} = \frac{a+b}{2}$$

de même pour la moyenne des carrés en utilisant la formule $\sum_{q=0}^{n} q^2 = \frac{n(n+1)(2n+1)}{6}$:

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k=a}^{b} k^2 \times \frac{1}{n} = \sum_{q=0}^{n-1} \frac{(a+q)^2}{n} = \sum_{q=0}^{n-1} \frac{a^2 + 2aq + q^2}{n} \\
&= a^2 + 2a\frac{n(n-1)}{2n} + \frac{n(n-1)(2n-1)}{6n} = a^2 + a(b-a) + \frac{(n-1)(2n-1)}{6} \\
&= ab + \frac{4n^2 - 6n + 2}{12} = ab + \frac{3n^2 - 6n + 3}{12} + \frac{n^2-1}{12} \\
&= ab + \frac{(n-1)^2}{4} + \frac{n^2-1}{12} = ab + \frac{(b-a)^2}{4} + \frac{n^2-1}{12} \\
&= \frac{(a+b)^2}{4} + \frac{n^2-1}{12} = \left(\frac{a+b}{2}\right)^2 + \frac{n^2-1}{12}
\end{aligned}$$

donc

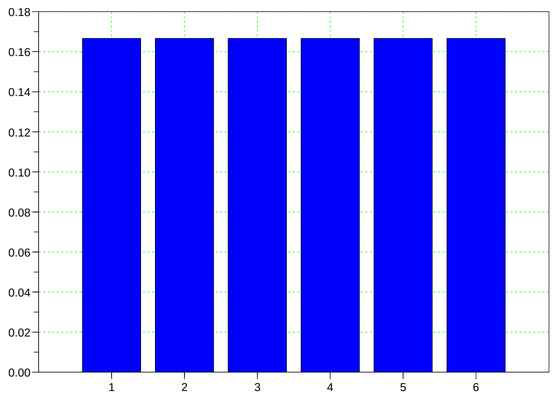
$$\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{n^2-1}{12}$$

□

✎ **4.3** on considère l'expérience « lancer un dé équilibré »étudier la v.a. définie par $X =$« numéro de la face supérieure »
On a $X \sim \mathscr{U}(a=1, b=6)$. On en déduit directement la loi de X :

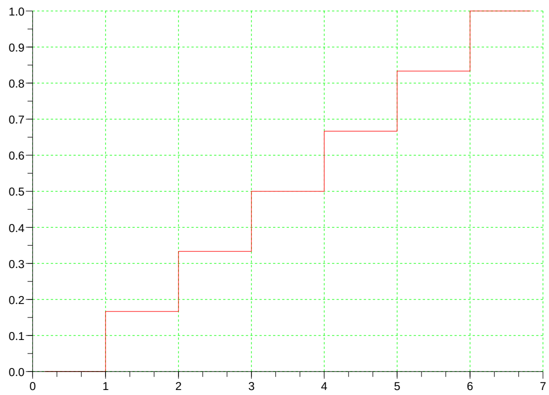
k	1	2	3	4	5	6
$\mathbb{P}(X=k)$	0.167	0.167	0.167	0.167	0.167	0.167

et son diagramme en bâton



ce qui donne pour la fonction de répartition :

$t \in I$	$]-\infty,1[$	$[1,2[$	$[2,3[$	$[3,4[$	$[4,5[$	$[5,6[$	$[6,+\infty[$
$F(t)$	0	0.167	0.333	0.5	0.667	0.833	1



On peut vérifier que les formules donnant les moments d'une loi binomiale sont correctes, pour la moyenne

$$\begin{aligned}
\mathbb{E}(X) &= 1 \times 0.1666667 + 2 \times 0.1666667 + 3 \times 0.1666667 \\
&\quad +4 \times 0.1666667 + 5 \times 0.1666667 + 6 \times 0.1666667 \\
&= 3.5 = \frac{1+6}{2}
\end{aligned}$$

et pour la variance

$$\begin{aligned}
\mathbb{E}(X^2) &= 1^2 \times 0.1666667 + 2^2 \times 0.1666667 + 3^2 \times 0.1666667 \\
&\quad +4^2 \times 0.1666667 + 5^2 \times 0.1666667 + 6^2 \times 0.1666667 \\
&= 15.166667 \\
\mathbb{V}\mathrm{ar}(X) &= 15.166667 - 3.5^2 = 2.9166667 = \frac{6^2-1}{12}
\end{aligned}$$

La première de ces lois est la loi de Bernoulli, elle correspond au type d'expérience le plus simple conduisant à un choix *« succès/échec »*.

Définition 4.14 (loi de Bernoulli)
On dit qu'une v.a. X suit une loi de Bernoulli de paramètre $p \in [0,1]$ une variable aléatoire qui ne prend que deux valeurs $X(\Omega) = \{0;1\}$, déterminée par :

$$\mathbb{P}(X=1) = p \quad \textit{et} \quad \mathbb{P}(X=0) = 1-p$$

L'exemple le plus simple de loi de Bernoulli est le jeu de *« pile ou face »*.

Proposition 4.15 *On considère l'expérience :*
« On effectue le lancer d'une pièce (équilibrée) et on note le résultat : pour pile $X = 1$ (succès) ou pour face $X = 0$ (échec). »
alors X suit une loi de Bernoulli de paramètre $p = \frac{1}{2}$.

les formules pour l'espérance et la variance nous serons utiles pour l'étude de la loi binomiale.

Théorème 4.16 (moments d'une loi de Bernoulli)
Soit X une v.a. qui suit une loi de Bernoulli de paramètre $p \in [0,1]$ alors

$$\mathbb{E}(X) = p \quad \textit{et} \quad \mathbb{V}\mathrm{ar}(X) = p(1-p)$$

Preuve :

$$\mathbb{E}(X) = 0 \times \mathbb{P}(X=0) + 1 \times \mathbb{P}(X=0) = 0 + p = p$$

ensuite

$$\mathbb{E}(X^2) = 0^2 \times \mathbb{P}(X=0) + 1^2 \times \mathbb{P}(X=0) = 0 + p = p$$

enfin

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = p - p^2 = p(1-p)$$

□

Les lois suivantes consistent en des répétitions d'expériences de Bernoulli.

Définition 4.17 (loi binomiale) *On dit qu'une v.a. X suit une loi binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0,1]$, noté $X \sim \mathscr{B}(n,p)$, si :*

$$\forall k = 0,1,\ldots,n, \ \ \mathbb{P}(X=k) = C_n^k p^k (1-p)^k$$

en particulier $X(\Omega) = \{0;1;2;\ldots;n\}$.

On peut vérifier que la formule $\mathbb{P}(X=k) = C_n^k p^k (1-p)^k$ définit bien la loi d'une v.a. puisque d'après la formule du binôme de Newton (*cf.*) on a bien

$$\mathbb{P}(\Omega) = \sum_{k=1}^{n} \mathbb{P}(X=k) = \sum_{k=1}^{n} C_n^k p^k (1-p)^k = (p + (1-p))^n = 1^n = 1$$

une loi binomiale compte le nombre de succès lors des répétition d'une expérience de Bernoulli.

Proposition 4.18 *Soit l'expérience :*
« On effectue n répétitions indépendantes d'une loi de Bernoulli (de paramètre p) et on compte le nombre de succès. »
alors $X \sim \mathscr{B}(n,p)$.

Preuve : on peut représenter l'univers de cette expérience par des n-uplet

$$\Omega = \{(x_1,\ldots,x_n) | x_i = 0 \quad ou \quad 1\} = \{0;1\}^n$$

un n-uplet correspondant à k succès va comporter k « 1 »et $n-k$ « 0 »et aura donc une probabilité $p^k(1-p)^{n-k}$ si les différents tirages sont indépendants. Il reste ensuite à « choisir la place des k succès parmi les n essais »la probabilité de $[X=k]$ sera donc C_n^k fois la probabilité d'un des cas. Au passage remarquons qu'une loi de Bernoulli est une loi binomiale de paramètre $n=1$. □

On a deux formule qui permettent un calcul rapide de l'espérance et de la variance.

Théorème 4.19 (moments d'une loi binomiale) *Soit* $X \sim \mathscr{B}(n,p)$ *alors*

$$\mathbb{E}(X) = np, \qquad \mathbb{V}\text{ar}(X) = np(1-p)$$

Preuve : soient $(X_i)_{i=0,\ldots,n}$ la loi de Bernoulli (de paramètre p) correspondant au $i^{\text{ième}}$ tirage alors on a

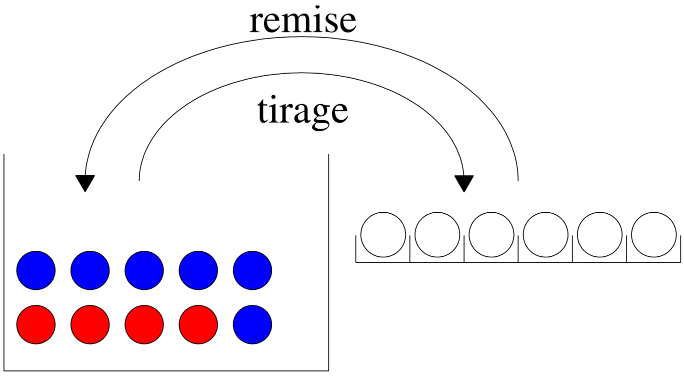
$$X = X_1 + \cdots + X_n \Longrightarrow \mathbb{E}(X) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = np$$

car les variables sont indépendantes. De même puisque les variables sont indépendantes on a

$$\mathbb{V}\text{ar}(X) = \mathbb{V}\text{ar}(X_1) + \cdots + \mathbb{V}\text{ar}(X_n) = np(1-p)$$

□

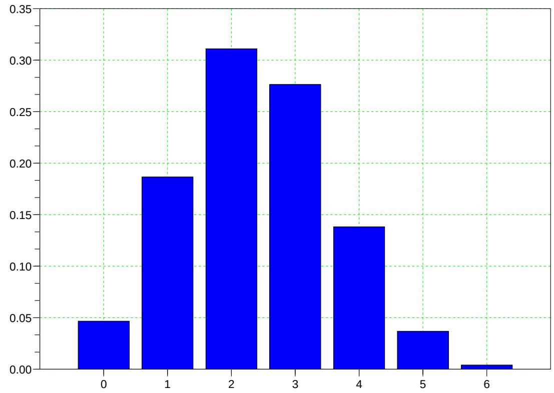
✎ **4.4** on considère l'expérience « tirer successivement et avec remise 6 boules parmi 4 rouges et 6 bleus »étudier la v.a. définie par « nombre de boules rouges obtenues »



tirer une boule est une expérience de Bernoulli de probabilité de succès (« obtenir une boule rouge ») $p = 0.4$. Cette expérience est répétée $n = 6$ fois de suite de manière indépendante (car il y a remise de la boule tirée à chaque fois) donc $X \sim \mathscr{B}(n=6, p=0.4)$. On en déduit directement la loi de X :

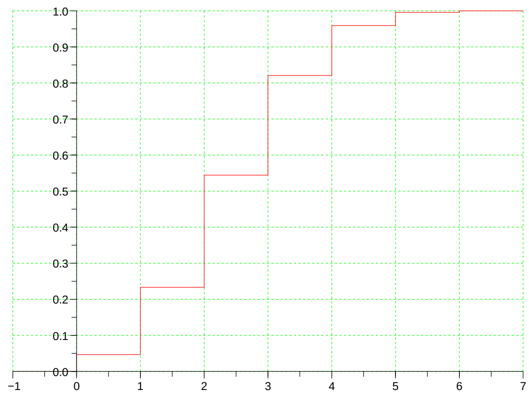
k	0	1	2	3	4	5	6
$\mathbb{P}(X=k)$	0.047	0.187	0.311	0.276	0.138	0.037	0.004

et son diagramme en bâton



ce qui donne pour la fonction de répartition :

$t \in I$	$]-\infty,0[$	$[0,1[$	$[1,2[$	$[2,3[$	$[3,4[$	$[4,5[$	$[5,6[$	$[6,+\infty[$
$F(t)$	0	0.047	0.233	0.544	0.821	0.959	0.996	1



On peut vérifier que les formules donnant les moments d'une loi binomiale sont correctes, pour la moyenne

$$\begin{aligned}\mathbb{E}(X) &= 0\times 0.046656 + 1\times 0.186624 + 2\times 0.31104 + 3\times 0.27648\\ &\quad +4\times 0.13824 + 5\times 0.036864 + 6\times 0.004096\\ &= 2.4 = 6\times 0.4\end{aligned}$$

et pour la variance

$$\begin{aligned}\mathbb{E}(X^2) &= 0^2\times 0.046656 + 1^2\times 0.186624 + 2^2\times 0.31104 + 3^2\times 0.27648\\ &\quad +4^2\times 0.13824 + 5^2\times 0.036864 + 6^2\times 0.004096\\ &= 7.2\\ \mathbb{V}\mathrm{ar}(X) &= 7.2 - 2.4^2 = 1.44 = 6\times 0.4\times 0.6\end{aligned}$$

Définition 4.20 (loi Hypergéométrique)
On dit qu'une v.a. X suit une loi hypergéométrique de paramètres $n,N \in \mathbb{N}^*$ avec $n \leq N$, et $p \in [0,1]$, noté $X \sim \mathscr{H}(n,p,N)$, si :

$$\forall k = 0,1,\ldots,n, \quad \mathbb{P}(X = k) = \frac{C_{pN}^{k} \times C_{(1-p)N}^{n-k}}{C_N^n}$$

en particulier $X(\Omega) = \{0; 1; 2; \ldots ; n\}$.

Contrairement à la loi binomiale, cette loi modélise les tirages simultanés.

Proposition 4.21 *Soit l'expérience :*
« On tire simultanément n boules dans une urne contenant pN boules gagnantes et $(1-p)N$ boules perdantes. On compte alors le nombre de boules gagnantes extraites et on appelle X la variable aléatoire donnant le nombre de boules gagnantes. »
alors $X \sim \mathscr{H}(n,p,N)$.

Preuve : pour des tirages simultanées dans un ensemble de N boules on a un univers de cardinal $\mathrm{Card}(\Omega) = C_N^n$ sur lequel on a équiprobabilité. Il faut donc compter le nombre de cas correspondants à « tirer k boules gagnantes parmi les pN boules gagnantes de l'urne »(et donc tirer $n-k$ boules perdantes parmi $(1-p)N$) il y en a $C_{pN}^{k} \times C_{(1-p)N}^{n-k}$ ce qui donne bien :

$$\mathbb{P}(X = k) = \frac{C_{pN}^{k} \times C_{(1-p)N}^{n-k}}{C_N^n}$$

□

il faut noter que $\mathbb{P}(X = k) = 0$ quand il n'y a pas assez de boules gagnantes ou au contraire trop de boules perdantes ce qui correspond à $k \notin [n-(1-p)N, pN]$. On retrouve bien ce résultat avec la formule :

$$\mathbb{P}(X = k) = \frac{C_{pN}^{k} \times C_{(1-p)N}^{n-k}}{C_N^n}$$

car quand $k > pN$ on a $C_{pN}^{k} = 0$ (pas assez de boule gagnantes) et quand $n-k > (1-p)N$ on a $C_{(1-p)N}^{n-k} = 0$ (trop de boules perdantes).

On a là aussi deux formules pour la moyenne et la variance.

Théorème 4.22 (moments d'une loi hypergéométrique)
Soit $X \sim \mathscr{H}(n,p,N)$ alors

$$\mathbb{E}(X) = np, \qquad \mathbb{V}\mathrm{ar}(X) = np(1-p)\frac{N-n}{N-1}$$

La démonstration de la formule pour la variance d'une loi hypergéométrique est assez compliquée, nous nous contenterons de la démonstration de la formule pour l'espérance.

Preuve : On considère un modèle d'urnes à tirage simultané, c'est-à-dire non ordonné et sans remise, avec 2 type de boules :

- $N_B = pN$ boules bleus « gagnantes »

- $N_R = (1-p)N$ boules rouges « perdantes »

. On peut bien sûr supposer que les boules sont numérotées, et définir les v.a. :

$$X_k = \text{« 1 si la boule bleu } n^\circ k \text{ a été tirée, 0 sinon »}$$

qui sont donc des v.a. de Bernoulli. Il est clair que $\forall k = 1, \ldots, N_B,\ \mathbb{E}(X_k) = \mathbb{E}(X_1)$ et que $X = X_1 + X_2 + \cdots + X_{N_B}$ donc

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{k=1}^{N_B} X_k\right) = \sum_{k=1}^{N_B} \mathbb{E}(X_k) = N_B \mathbb{E}(X_1)$$

Or

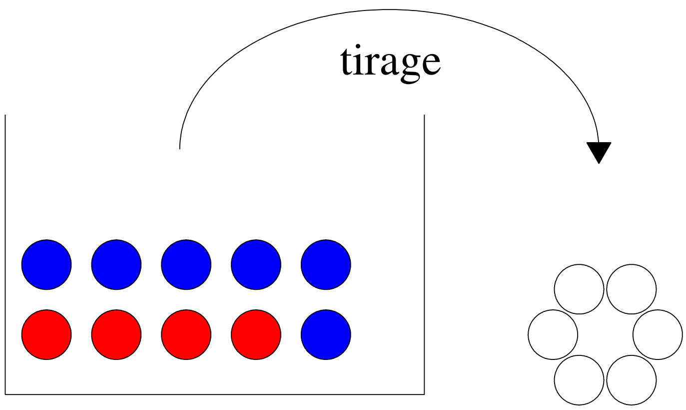
$$\mathbb{P}(X_1 = 0) = \frac{C_{N-1}^n}{C_N^n} = \frac{N-n}{N} \Longrightarrow \mathbb{P}(X_1 = 1) = 1 - \mathbb{P}(X_1 = 0) = \frac{n}{N}$$

et donc

$$\mathbb{E}(X_1) = 0 \times \mathbb{P}(X_1 = 0) + 1 \times \mathbb{P}(X_1 = 0) = \frac{n}{N} \Longrightarrow \mathbb{E}(X) = \frac{n}{N} \times pN = np$$

□

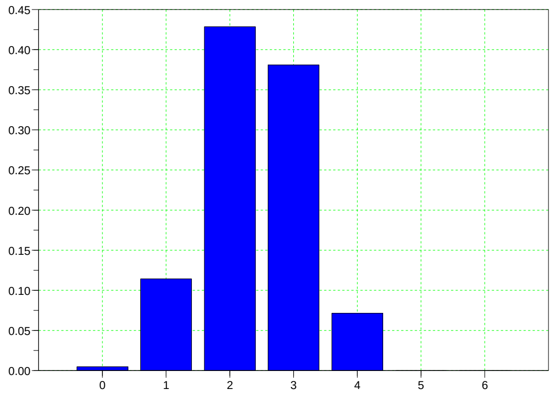
✎ **4.5** on considère l'expérience « tirer simultanément 6 boules parmi 4 rouges et 6 bleus »étudier la v.a. définie par « nombre de boules rouges obtenues »



la proportion de boules gagnantes (les rouges) est $p = 0.4$ on tire $n = 6$ boules parmi les $N = 10$ donc $X \sim \mathscr{H}(n = 6, p = 0.4, N = 10)$. On en déduit directement la loi de X :

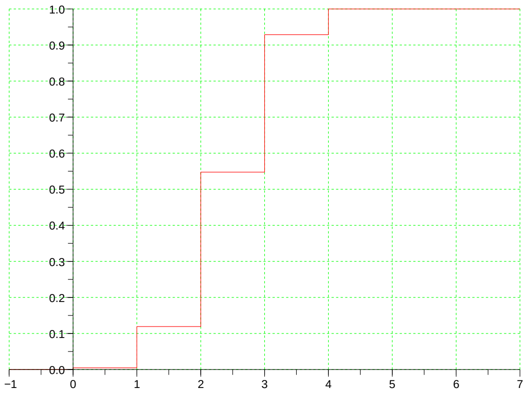
k	0	1	2	3	4	5	6
$\mathbb{P}(X = k)$	0.005	0.114	0.429	0.381	0.071	0	0

et son diagramme en bâton



ce qui donne pour la fonction de répartition :

$t \in I$	$]-\infty,0[$	$[0,1[$	$[1,2[$	$[2,3[$	$[3,4[$	$[4,5[$	$[5,6[$	$[6,+\infty[$
$F(t)$	0	0.005	0.119	0.548	0.929	1	1	1



On peut vérifier que les formules donnant les moments d'une loi binomiale sont correctes, pour la moyenne

$$\begin{aligned}
\mathbb{E}(X) &= 0 \times 0.0047619 + 1 \times 0.1142857 + 2 \times 0.4285714 + 3 \times 0.3809524 \\
&\quad +4 \times 0.0714286 + 5 \times 0 + 6 \times 0 \\
&= 2.4 = 6 \times 0.4
\end{aligned}$$

et pour la variance

$$\begin{aligned}
\mathbb{E}(X^2) &= 0^2 \times 0.0047619 + 1^2 \times 0.1142857 + 2^2 \times 0.4285714 + 3^2 \times 0.3809524 \\
&\quad +4^2 \times 0.0714286 + 5^2 \times 0 + 6^2 \times 0 \\
&= 6.4 \\
\mathbb{V}\mathrm{ar}(X) &= 6.4 - 2.4^2 = 0.64 = 6 \times 0.4 \times 0.6 \times \frac{10-6}{10-1}
\end{aligned}$$

Définition 4.23 (loi de Poisson) *On dit qu'une v.a. X suit une loi de Poisson de paramètre $\lambda > 0$, noté $X \sim \mathscr{P}(\lambda)$, si :*

$$\forall k = \in \mathbb{N}, \quad \mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

en particulier $X(\Omega) = \mathbb{N}$.

On peut vérifier que la formule $\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ définit bien la loi d'une v.a. puisque d'après la formule de la série exponentielle on a bien

$$\mathbb{P}(\Omega) = \sum_{k=1}^{n} \mathbb{P}(X = k) = \sum_{k=1}^{n} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{n}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1$$

Cette loi modélise le comptage des « événements rares »

Proposition 4.24 *Soit l'expérience :*
« Un serveur reçoit en moyenne n connexions par minutes. Soit X la variable aléatoire qui compte le nombre de connexions au serveur durant un temps T minutes. »
alors $X \sim \mathscr{P}(\lambda)$ avec $\lambda = nT$.

Passons aux formules pour la moyenne et la variance.

Théorème 4.25 (moments d'une loi de Poisson) *Soit $X \sim \mathscr{P}(\lambda)$ alors*

$$\mathbb{E}(X) = \lambda, \qquad \mathbb{V}\mathrm{ar}(X) = \lambda$$

Preuve : Pour l'espérence :

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=1}^{+\infty} k\,\mathbb{P}(X = k) = \sum_{k=1}^{+\infty} k\,e^{-\lambda}\frac{\lambda^k}{k!} \\
&= e^{-\lambda}\sum_{k=1}^{+\infty}\frac{\lambda^k}{(k-1)!} = \lambda\,e^{-\lambda}\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda\,e^{-\lambda}\,e^{\lambda} = \lambda
\end{aligned}$$

pour la variance on calcule d'abord :

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k=1}^{+\infty} k^2\,\mathbb{P}(X = k) = \sum_{k=1}^{+\infty} k^2\,e^{-\lambda}\frac{\lambda^k}{k!} \\
&= \lambda\,e^{-\lambda}\sum_{k=1}^{+\infty}\frac{k\lambda^{k-1}}{(k-1)!} \\
&= \lambda\,e^{-\lambda}\sum_{k=1}^{+\infty}\frac{(k-1+1)\lambda^{k-1}}{(k-1)!} \\
&= \lambda\,e^{-\lambda}\left(\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!} + \frac{\lambda^{k-1}}{(k-2)!}\right) \\
&= \lambda\,e^{-\lambda}\left(\left(\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!}\right) + \left(\sum_{k=0}^{+\infty}\lambda\frac{\lambda^k}{k!}\right)\right) \\
&= \lambda\,e^{-\lambda}\left(e^{\lambda} + \lambda e^{\lambda}\right) = \lambda + \lambda^2
\end{aligned}$$

et avec la formule de Koening on obtient

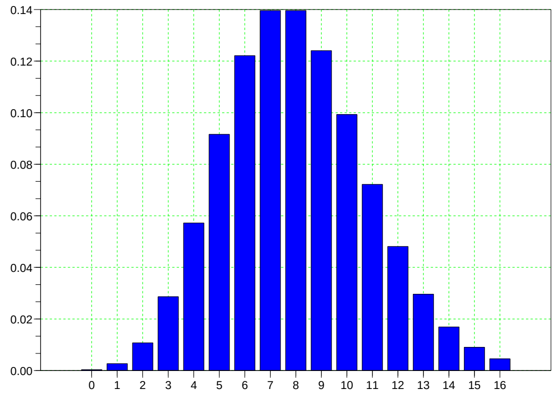
$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$$

$\square$

✎ **4.6** Soit l'expérience « un standard reçoit en moyenne 2 appels par quart d'heure »étudier la v.a. définie par $X =$« nombre d'appels reçut en une heure »
Par définition il s'agit d'une v.a. suivant une loi de Poisson de paramètre $\lambda = 2\times 4 = 8$ donc $X \sim \mathscr{P}(\lambda = 8)$. Bien entendu comme l'univers image de cette loi n'est pas fini on ne peut pas la représenter par un tableau. Cependant si on calcule les probabilités de X on peut voir qu'au-delà de $k = 15$ les probabilités deviennent très faibles (inférieures à 1%) :

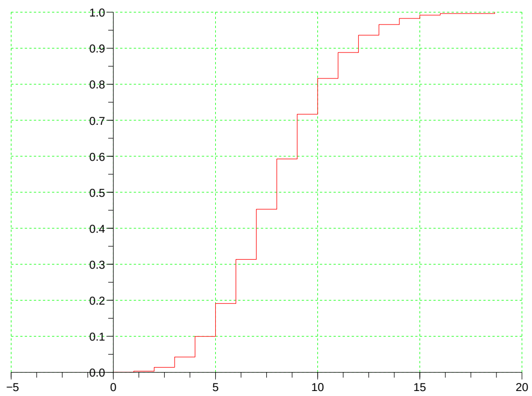
k	0	1	2	3	4	5	6	7
$\mathbb{P}(X=k)$	0	0.003	0.011	0.029	0.057	0.092	0.122	0.14
k	8	9	10	11	12	13	14	15
$\mathbb{P}(X=k)$	0.14	0.124	0.099	0.072	0.048	0.03	0.017	0.009

c'est aussi bien visible sur le diagramme en bâton



ce qui donne pour la fonction de répartition :

$t \in I$	$]-\infty,0[$	$[0,1[$	$[1,2[$	$[2,3[$	$[3,4[$	$[4,5[$	$[5,6[$	$[6,7[$	$[7,8[$
$F(t)$	0	0	0.003	0.014	0.042	0.1	0.191	0.313	0.453
$t \in I$	$[7,8[$	$[8,9[$	$[9,10[$	$[10,11[$	$[11,12[$	$[12,13[$	$[13,14[$	$[14,15[$	$[15,+\infty[$
$F(t)$	0.593	0.717	0.816	0.888	0.936	0.966	0.983	0.992	1



ici impossible de calculer moyenne ou variance sans passer par la formule. On peut quand même vérifier qu'on trouve un résultat proche les formules donnant les moments d'une loi de Poisson sont correctes, pour la moyenne

$$\mathbb{E}(X) = 8 \quad \approx \quad \sum_{k=0}^{15} k \times \mathbb{P}(X = k) \approx 7.9341519$$

et pour la variance

$$\begin{aligned} \mathbb{E}(X^2) &\approx 70.829705 \\ \mathbb{V}\mathrm{ar}(X) &= 8 \approx 70.829705 - 7.9341519^2 \approx 7.8789386 \end{aligned}$$

Définition 4.26 (loi géométrique) *On dit qu'une v.a. X suit une loi géométrique de paramètre $p \in]0,1]$, noté $X \sim \mathscr{G}(p)$, si :*

$$\forall k =\in \mathbb{N}^*, \quad \mathbb{P}(X = k) = p(1-p)^{k-1}$$

en particulier $X(\Omega) = \mathbb{N}^*$.

On peut vérifier que la formule $\mathbb{P}(X = k) = p(1-p)^{k-1}$ définit bien la loi d'une v.a. puisque d'après la formule de la série géométrique on a bien

$$\mathbb{P}(\Omega) = \sum_{k=1}^{\infty} \mathbb{P}(X = k) = \sum_{k=1}^{\infty} p(1-p)^{k-1} = p\sum_{k=0}^{\infty}(1-p)^k = p\frac{1}{1-(1-p)} = 1$$

Proposition 4.27 *On considère l'expérience : « répéter un jeu de pile ou face jusqu'à obtenir pile »et la variable aléatoire $X =$« nombre de lancer pour obtenir le premier pile »*
alors X suit une loi géométrique de paramètre $p = \frac{1}{2}$, $X \sim \mathscr{G}(p = 0.5)$.

Preuve : On peut représenter l'univers correspondants à k lancers de pile ou face par un ensemble de n-uplets :

$$\Omega = \{(x_1, \ldots, x_k) | x_i = 0 \quad ou \quad 1\}$$

où $\mathbb{P}(x_i = 0) = 1 - p$ et $\mathbb{P}(x_i = 1) = p$. Ensuite l'événement $[X = k] =$« obtenir le premier pile au $k^{\text{ième}}$ lancer »correspond au n-uplet $(0, \ldots, 0, 1)$ et à donc pour probabilité (en supposant les lancers indépendants) :

$$\mathbb{P}(X = k) = \underbrace{(1-p) \times \cdots \times (1-p)}_{k-1 \text{ lancers avec face}} \times p = p(1-p)^{k-1}$$

comme il faut au moins 1 lancer pour obtenir pile l'univers image est bien $\mathbb{N}^*$. □
Comme pour les autres lois on a une formule simple pour calculer l'espérance et la variance.

Théorème 4.28 (moments d'une loi géométrique) *Soit $X \sim \mathscr{G}(p)$ alors*

$$\mathbb{E}(X) = \frac{1}{p}, \qquad \mathbb{V}\mathrm{ar}(X) = \frac{1-p}{p^2}$$

Preuve : Les résultats sur l'espérance reposent sur les formules suivantes, valables pour tout $x \in [0,1[$,

$$\begin{aligned}
f(x) &= \sum_{k=1}^{+\infty} x^{k-1} = \sum_{k=0}^{+\infty} x^k = \frac{1}{1-x} \\
f'(x) &= \sum_{k=1}^{+\infty} kx^{k-1} = \sum_{k=0}^{+\infty} (k+1)x^k = \frac{1}{(1-x)^2} \\
f''(x) &= \sum_{k=1}^{+\infty} k(k+1)x^{k-1} = \sum_{k=1}^{+\infty} k^2x^{k-1} + \sum_{k=1}^{+\infty} kx^{k-1} = \frac{2}{(1-x)^3}
\end{aligned}$$

d'où l'on tire

$$\sum_{k=1}^{+\infty} k^2x^{k-1} = \frac{2}{(1-x)^3} - \frac{1}{(1-x)^2}$$

On peut alors calculer l'espérance de X en prenant $x = 1-p$:

$$\mathbb{E}(X) = \sum_{k=1}^{+\infty} kp(1-p)^{k-1} = \frac{p}{(1-(1-p))^2} = \frac{1}{p}$$

et pour la moyenne des carrés :

$$\mathbb{E}(X^2) = \left(\sum_{k=1}^{+\infty} k^2p(1-p)^{k-1}\right) = \frac{2p}{(1-(1-p))^3} - \frac{p}{(1-(1-p))^2} = \frac{2}{p^2} - \frac{1}{p}$$

et enfin pour la variance :

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2}$$

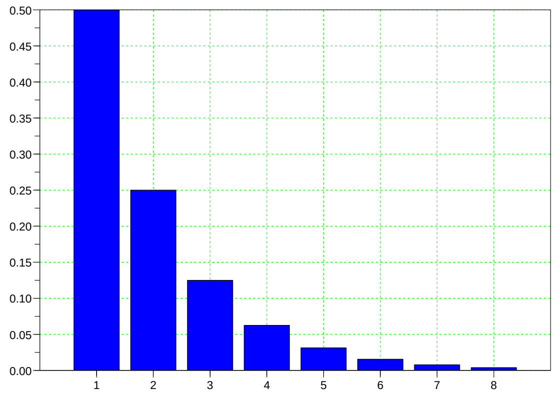
□

✎ **4.7**

soit l'expérience « répéter un jeu de pile ou face jusqu'à obtenir pile »
étudier la variable aléatoire $X =$« nombre de lancers pour obtenir le premier pile »
on a déjà dit que X suit une loi géométrique de paramètre $p = \frac{1}{2}$, $X \sim \mathscr{G}(p = 0.5)$. Bien entendu comme l'univers image de cette loi n'est pas fini on ne peut pas la représenter par un tableau. Cependant si on calcule les probabilités de X on peut voir qu'au-delà de $k = 7$ les probabilités devient très faibles (inférieures à 1%) :

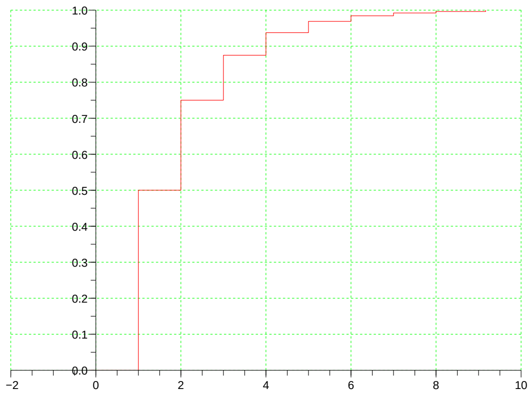
k	1	2	3	4	5	6	7	8
$\mathbb{P}(X = k)$	0.5	0.25	0.125	0.063	0.031	0.016	0.008	0.004

c'est aussi bien visible sur le diagramme en bâton



ce qui donne pour la fonction de répartition :

$t \in I$	$]-\infty{,}1[$	$[1{,}2[$	$[2{,}3[$	$[3{,}4[$	$[4{,}5[$	$[5{,}6[$	$[6{,}7[$	$[7{,}8[$	$[8,+\infty[$
$F(t)$	0	0.5	0.75	0.875	0.938	0.969	0.984	0.992	0.996



Là encore impossible de calculer moyenne ou variance sans passer par la formule. On peut quand même vérifier qu'on trouve un résultat proche les formules donnant les moments d'une loi géométrique sont correctes, pour la moyenne

$$\mathbb{E}(X) = 2 \quad \approx \quad \sum_{k=0}^{8} k \times \mathbb{P}(X=k) \approx 1.9609375$$

et pour la variance

$$\begin{aligned} \mathbb{E}(X^2) &\approx 5.6015625 \\ \mathbb{V}\mathrm{ar}(X) &= 2 \approx 5.6015625 - 1.9609375^2 \approx 1.7562866 \end{aligned}$$

pour ce dernier calcul on voit combien il est désatreux de ne pas prendre tous les termes de la somme!

Dans la pratique il est souvent fastidieux d'utiliser les formules précédentes pour calculer les probabilités $\mathbb{P}(X=k)$ d'une loi donnée, c'est pourquoi on a pris l'habitude d'utiliser des tables donnant les valeurs numériques de ces probabilités suivant la loi (et ses paramètres).

$$\mathbb{P}(X = k) = p \times (1-p)^{k-1}$$

p \ k	1	2	3	4	5	6	7	8	9	10	11	12
0.05	0.05	0.0475	0.0451	0.0429	0.0407	0.0387	0.0368	0.0349	0.0332	0.0315	0.0299	0.0284
0.1	0.1	0.09	0.081	0.0729	0.0656	0.059	0.0531	0.0478	0.043	0.0387	0.0349	0.0314
0.15	0.15	0.1275	0.1084	0.0921	0.0783	0.0666	0.0566	0.0481	0.0409	0.0347	0.0295	0.0251
0.2	0.2	0.16	0.128	0.1024	0.0819	0.0655	0.0524	0.0419	0.0336	0.0268	0.0215	0.0172
0.25	0.25	0.1875	0.1406	0.1055	0.0791	0.0593	0.0445	0.0334	0.025	0.0188	0.0141	0.0106
0.3	0.3	0.21	0.147	0.1029	0.072	0.0504	0.0353	0.0247	0.0173	0.0121	0.0085	0.0059
0.35	0.35	0.2275	0.1479	0.0961	0.0625	0.0406	0.0264	0.0172	0.0112	0.0072	0.0047	0.0031
0.4	0.4	0.24	0.144	0.0864	0.0518	0.0311	0.0187	0.0112	0.0067	0.004	0.0024	0.0015
0.45	0.45	0.2475	0.1361	0.0749	0.0412	0.0226	0.0125	0.0069	0.0038	0.0021	0.0011	0.0006
0.5	0.5	0.25	0.125	0.0625	0.0313	0.0156	0.0078	0.0039	0.002	0.001	0.0005	0.0002
0.55	0.55	0.2475	0.1114	0.0501	0.0226	0.0101	0.0046	0.0021	0.0009	0.0004	0.0002	0.0001
0.6	0.6	0.24	0.096	0.0384	0.0154	0.0061	0.0025	0.001	0.0004	0.0002	0.0001	0
0.65	0.65	0.2275	0.0796	0.0279	0.0098	0.0034	0.0012	0.0004	0.0001	0.0001	0	0
0.7	0.7	0.21	0.063	0.0189	0.0057	0.0017	0.0005	0.0002	0	0	0	0
0.75	0.75	0.1875	0.0469	0.0117	0.0029	0.0007	0.0002	0	0	0	0	0
0.8	0.8	0.16	0.032	0.0064	0.0013	0.0003	0.0001	0	0	0	0	0
0.85	0.85	0.1275	0.0191	0.0029	0.0004	0.0001	0	0	0	0	0	0
0.9	0.9	0.09	0.009	0.0009	0.0001	0	0	0	0	0	0	0
0.95	0.95	0.0475	0.0024	0.0001	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0

FIG. 9 – *Table de la loi géométrique en fonction du paramètre p*

Par exemple, dans un jeu de pile ou face, la probabilité d'obtenir le premier pile au $4^{\text{ième}}$ lancer est $\mathbb{P}(X = 4) = 6{,}25\%$ et se trouve dans cette table à l'intersection de la ligne $p = 0.5$ et de la colonne $k = 4$.

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

λ \ k	0	1	2	3	4	5	6	7	8	9	10
0.1	0.9048	0.0905	0.0045	0.0002	0	0	0	0	0	0	0
0.2	0.8187	0.1637	0.0164	0.0011	0.0001	0	0	0	0	0	0
0.3	0.7408	0.2222	0.0333	0.0033	0.0003	0	0	0	0	0	0
0.4	0.6703	0.2681	0.0536	0.0072	0.0007	0.0001	0	0	0	0	0
0.5	0.6065	0.3033	0.0758	0.0126	0.0016	0.0002	0	0	0	0	0
0.6	0.5488	0.3293	0.0988	0.0198	0.003	0.0004	0	0	0	0	0
0.7	0.4966	0.3476	0.1217	0.0284	0.005	0.0007	0.0001	0	0	0	0
0.8	0.4493	0.3595	0.1438	0.0383	0.0077	0.0012	0.0002	0	0	0	0
0.9	0.4066	0.3659	0.1647	0.0494	0.0111	0.002	0.0003	0	0	0	0
1	0.3679	0.3679	0.1839	0.0613	0.0153	0.0031	0.0005	0.0001	0	0	0
1.1	0.3329	0.3662	0.2014	0.0738	0.0203	0.0045	0.0008	0.0001	0	0	0
1.2	0.3012	0.3614	0.2169	0.0867	0.026	0.0062	0.0012	0.0002	0	0	0
1.3	0.2725	0.3543	0.2303	0.0998	0.0324	0.0084	0.0018	0.0003	0.0001	0	0
1.4	0.2466	0.3452	0.2417	0.1128	0.0395	0.0111	0.0026	0.0005	0.0001	0	0
1.5	0.2231	0.3347	0.251	0.1255	0.0471	0.0141	0.0035	0.0008	0.0001	0	0
1.6	0.2019	0.323	0.2584	0.1378	0.0551	0.0176	0.0047	0.0011	0.0002	0	0
1.7	0.1827	0.3106	0.264	0.1496	0.0636	0.0216	0.0061	0.0015	0.0003	0.0001	0
1.8	0.1653	0.2975	0.2678	0.1607	0.0723	0.026	0.0078	0.002	0.0005	0.0001	0
1.9	0.1496	0.2842	0.27	0.171	0.0812	0.0309	0.0098	0.0027	0.0006	0.0001	0
2	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.012	0.0034	0.0009	0.0002	0
2	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.012	0.0034	0.0009	0.0002	0
2.2	0.1108	0.2438	0.2681	0.1966	0.1082	0.0476	0.0174	0.0055	0.0015	0.0004	0.0001
2.4	0.0907	0.2177	0.2613	0.209	0.1254	0.0602	0.0241	0.0083	0.0025	0.0007	0.0002
2.6	0.0743	0.1931	0.251	0.2176	0.1414	0.0735	0.0319	0.0118	0.0038	0.0011	0.0003
2.8	0.0608	0.1703	0.2384	0.2225	0.1557	0.0872	0.0407	0.0163	0.0057	0.0018	0.0005
3	0.0498	0.1494	0.224	0.224	0.168	0.1008	0.0504	0.0216	0.0081	0.0027	0.0008
3.2	0.0408	0.1304	0.2087	0.2226	0.1781	0.114	0.0608	0.0278	0.0111	0.004	0.0013
3.4	0.0334	0.1135	0.1929	0.2186	0.1858	0.1264	0.0716	0.0348	0.0148	0.0056	0.0019
3.6	0.0273	0.0984	0.1771	0.2125	0.1912	0.1377	0.0826	0.0425	0.0191	0.0076	0.0028
3.8	0.0224	0.085	0.1615	0.2046	0.1944	0.1477	0.0936	0.0508	0.0241	0.0102	0.0039
4	0.0183	0.0733	0.1465	0.1954	0.1954	0.1563	0.1042	0.0595	0.0298	0.0132	0.0053
5	0.0067	0.0337	0.0842	0.1404	0.1755	0.1755	0.1462	0.1044	0.0653	0.0363	0.0181
6	0.0025	0.0149	0.0446	0.0892	0.1339	0.1606	0.1606	0.1377	0.1033	0.0688	0.0413
7	0.0009	0.0064	0.0223	0.0521	0.0912	0.1277	0.149	0.149	0.1304	0.1014	0.071
8	0.0003	0.0027	0.0107	0.0286	0.0573	0.0916	0.1221	0.1396	0.1396	0.1241	0.0993
9	0.0001	0.0011	0.005	0.015	0.0337	0.0607	0.0911	0.1171	0.1318	0.1318	0.1186
10	0	0.0005	0.0023	0.0076	0.0189	0.0378	0.0631	0.0901	0.1126	0.1251	0.1251

FIG. 10 – *Table de la loi de Poisson en fonction du paramètre λ*

Par exemple, pour une loi de Poisson de paramètre $\lambda = 1$ la probabilité que $X = 2$ est $\mathbb{P}(X = 2) = 18{,}39\%$ et se trouve dans cette table à l'intersection de la ligne $\lambda = 1$ et de la colonne $k = 2$.

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

n	k \ p	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
2	0	0.9025	0.81	0.7225	0.64	0.5625	0.49	0.4225	0.36	0.3025	0.25
2	1	0.095	0.18	0.255	0.32	0.375	0.42	0.455	0.48	0.495	0.5
2	2	0.0025	0.01	0.0225	0.04	0.0625	0.09	0.1225	0.16	0.2025	0.25
3	0	0.8574	0.729	0.6141	0.512	0.4219	0.343	0.2746	0.216	0.1664	0.125
3	1	0.1354	0.243	0.3251	0.384	0.4219	0.441	0.4436	0.432	0.4084	0.375
3	2	0.0071	0.027	0.0574	0.096	0.1406	0.189	0.2389	0.288	0.3341	0.375
3	3	0.0001	0.001	0.0034	0.008	0.0156	0.027	0.0429	0.064	0.0911	0.125
4	0	0.8145	0.6561	0.522	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
4	1	0.1715	0.2916	0.3685	0.4096	0.4219	0.4116	0.3845	0.3456	0.2995	0.25
4	2	0.0135	0.0486	0.0975	0.1536	0.2109	0.2646	0.3105	0.3456	0.3675	0.375
4	3	0.0005	0.0036	0.0115	0.0256	0.0469	0.0756	0.1115	0.1536	0.2005	0.25
4	4	0	0.0001	0.0005	0.0016	0.0039	0.0081	0.015	0.0256	0.041	0.0625
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.116	0.0778	0.0503	0.0313
5	1	0.2036	0.3281	0.3915	0.4096	0.3955	0.3601	0.3124	0.2592	0.2059	0.1563
5	2	0.0214	0.0729	0.1382	0.2048	0.2637	0.3087	0.3364	0.3456	0.3369	0.3125
5	3	0.0011	0.0081	0.0244	0.0512	0.0879	0.1323	0.1811	0.2304	0.2757	0.3125
5	4	0	0.0005	0.0022	0.0064	0.0146	0.0284	0.0488	0.0768	0.1128	0.1563
5	5	0	0	0.0001	0.0003	0.001	0.0024	0.0053	0.0102	0.0185	0.0313
6	0	0.7351	0.5314	0.3771	0.2621	0.178	0.1176	0.0754	0.0467	0.0277	0.0156
6	1	0.2321	0.3543	0.3993	0.3932	0.356	0.3025	0.2437	0.1866	0.1359	0.0938
6	2	0.0305	0.0984	0.1762	0.2458	0.2966	0.3241	0.328	0.311	0.278	0.2344
6	3	0.0021	0.0146	0.0415	0.0819	0.1318	0.1852	0.2355	0.2765	0.3032	0.3125
6	4	0.0001	0.0012	0.0055	0.0154	0.033	0.0595	0.0951	0.1382	0.1861	0.2344
6	5	0	0.0001	0.0004	0.0015	0.0044	0.0102	0.0205	0.0369	0.0609	0.0938
6	6	0	0	0	0.0001	0.0002	0.0007	0.0018	0.0041	0.0083	0.0156
7	0	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.049	0.028	0.0152	0.0078
7	1	0.2573	0.372	0.396	0.367	0.3115	0.2471	0.1848	0.1306	0.0872	0.0547
7	2	0.0406	0.124	0.2097	0.2753	0.3115	0.3177	0.2985	0.2613	0.214	0.1641
7	3	0.0036	0.023	0.0617	0.1147	0.173	0.2269	0.2679	0.2903	0.2918	0.2734
7	4	0.0002	0.0026	0.0109	0.0287	0.0577	0.0972	0.1442	0.1935	0.2388	0.2734
7	5	0	0.0002	0.0012	0.0043	0.0115	0.025	0.0466	0.0774	0.1172	0.1641
7	6	0	0	0.0001	0.0004	0.0013	0.0036	0.0084	0.0172	0.032	0.0547
7	7	0	0	0	0	0.0001	0.0002	0.0006	0.0016	0.0037	0.0078
8	0	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039
8	1	0.2793	0.3826	0.3847	0.3355	0.267	0.1977	0.1373	0.0896	0.0548	0.0313
8	2	0.0515	0.1488	0.2376	0.2936	0.3115	0.2965	0.2587	0.209	0.1569	0.1094
8	3	0.0054	0.0331	0.0839	0.1468	0.2076	0.2541	0.2786	0.2787	0.2568	0.2188
8	4	0.0004	0.0046	0.0185	0.0459	0.0865	0.1361	0.1875	0.2322	0.2627	0.2734
8	5	0	0.0004	0.0026	0.0092	0.0231	0.0467	0.0808	0.1239	0.1719	0.2188
8	6	0	0	0.0002	0.0011	0.0038	0.01	0.0217	0.0413	0.0703	0.1094
8	7	0	0	0	0.0001	0.0004	0.0012	0.0033	0.0079	0.0164	0.0313
8	8	0	0	0	0	0	0.0001	0.0002	0.0007	0.0017	0.0039
9	0	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.002
9	1	0.2985	0.3874	0.3679	0.302	0.2253	0.1556	0.1004	0.0605	0.0339	0.0176
9	2	0.0629	0.1722	0.2597	0.302	0.3003	0.2668	0.2162	0.1612	0.111	0.0703
9	3	0.0077	0.0446	0.1069	0.1762	0.2336	0.2668	0.2716	0.2508	0.2119	0.1641
9	4	0.0006	0.0074	0.0283	0.0661	0.1168	0.1715	0.2194	0.2508	0.26	0.2461
9	5	0	0.0008	0.005	0.0165	0.0389	0.0735	0.1181	0.1672	0.2128	0.2461
9	6	0	0.0001	0.0006	0.0028	0.0087	0.021	0.0424	0.0743	0.116	0.1641
9	7	0	0	0	0.0003	0.0012	0.0039	0.0098	0.0212	0.0407	0.0703
9	8	0	0	0	0	0.0001	0.0004	0.0013	0.0035	0.0083	0.0176
9	9	0	0	0	0	0	0	0.0001	0.0003	0.0008	0.002
10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.006	0.0025	0.001
10	1	0.3151	0.3874	0.3474	0.2684	0.1877	0.1211	0.0725	0.0403	0.0207	0.0098
10	2	0.0746	0.1937	0.2759	0.302	0.2816	0.2335	0.1757	0.1209	0.0763	0.0439
10	3	0.0105	0.0574	0.1298	0.2013	0.2503	0.2668	0.2522	0.215	0.1665	0.1172
10	4	0.001	0.0112	0.0401	0.0881	0.146	0.2001	0.2377	0.2508	0.2384	0.2051
10	5	0.0001	0.0015	0.0085	0.0264	0.0584	0.1029	0.1536	0.2007	0.234	0.2461
10	6	0	0.0001	0.0012	0.0055	0.0162	0.0368	0.0689	0.1115	0.1596	0.2051
10	7	0	0	0.0001	0.0008	0.0031	0.009	0.0212	0.0425	0.0746	0.1172
10	8	0	0	0	0.0001	0.0004	0.0014	0.0043	0.0106	0.0229	0.0439
10	9	0	0	0	0	0	0.0001	0.0005	0.0016	0.0042	0.0098
10	10	0	0	0	0	0	0	0	0.0001	0.0003	0.001

FIG. 11 – *Table de la loi Binomiale en fonction des paramètres n et p*

Avec cette dernière table il faut faire un peu plus attention. Par exemple, dans un jeu de pile ou face, la probabilité d'obtenir 3 pile au cours de 5 est $\mathbb{P}(X=3)$ pour $X \sim \mathscr{B}(n=5, p=0.5)$. Pour la trouver dans la table il faut d'abord chercher le groupe de lignes correspondant à $n=5$ puis regarder la ligne $k=3$. la probablité cherchée se trouve alors à à l'intersection avec la colonne $p=0.5$ et vaut 31.25%.

On peut aussi calculer les probabilités d'une loi $X \sim \mathscr{B}(n,p)$ avec $p > 1/2$ en utilisant la table binomiale. Il suffit pour cela de compter le nombre d'échecs $Y = n - X$ qui suit aussi une loi binomiale $\mathscr{B}(n, 1-p)$ mais avec $1-p < 1/2$!

5 Théorème asymptotiques

La convergence de suites de variables aléatoires est un concept important qui va nous permettre de faire le lien entre la théorie abstraite des probabilités et celle plus concrète des statistiques. Elle peut être utile soit pour simplifier le calcul des probabilités de certaines situations soit pour estimer la validité de certaines hypothèses (comme l'indépendance, la valeur de la moyenne ...)

5.1 Approximation de lois

Nous allons commencer par deux théorèmes qui permettent de ramener le calcul des probabilités d'une loi compliqué à une plus simple.

Théorème 5.1 (Approximation d'une loi Hypergéométrique) *Soient* $X \sim \mathscr{H}(n,p,N)$ *et* $Z \sim \mathscr{B}(n,p)$, *si* $N \to \infty$ *alors on peut dire que* $X \approx Z$ *à partir du moment où* $n < \frac{N}{10}$ *ce qui signifie que :*

$$\forall k \in [0,n], \quad \mathbb{P}(X = k) \approx \mathbb{P}(Z = k)$$

Pour illustrer le théorème 5.1 on pourra constater sur les histogrammes de la figure FIG.12 que l'écart entre la loi binomiale (pour $n = 50$ et $p = 0.1$) et la loi de hypergéométrique (pour $n = 50$, $p = 0.1$ et $N = 1000$) est inférieur à 0.5%.

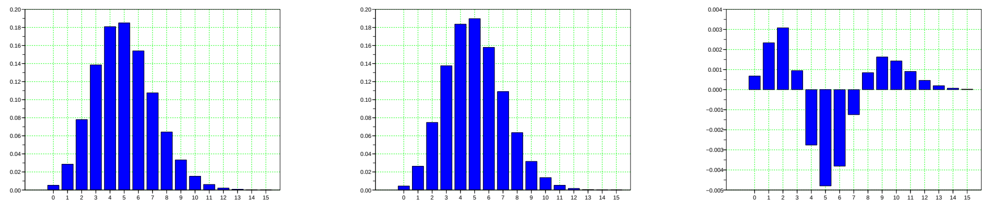


FIG. 12 – *histogramme des lois* $\mathscr{B}(n = 50,p = 0.1)$ $\mathscr{H}(n = 50,p = 0.1,N = 1000)$ *et de leur écart*

L'intérêt de cette Approximation est qu'il est beaucoup plus facile de calculer les probabilités d'une loi de Binomiale que celles d'une loi hypergéométrique, à cause des coefficients C_n^k qui nécessitent le calcul de $n!$ ou du triangle de Pascal correspondant. Tout ce passe comme si pour N grand la probabilité de tirer un membre d'une des populations ne varie pas d'un essai à l'autre. Cette approximation est utilisé dans la théorie des sondages pour estimer le résultat à partir d'une loi binomiale.

Théorème 5.2 (convergence d'une loi Binomiale vers une loi de Poisson) *Soient* $X \sim \mathscr{B}(n,p)$ *et* $Z \sim \mathscr{P}(\lambda)$, *si* $\lambda = np$ *et* $n \to \infty$ *alors on peut dire que* $X \approx Z$ *à partir du moment où* $n \geq 30$ *et* $np \leq 5$, *ce qui signifie que :*

$$\forall k \in [0,n], \quad \mathbb{P}(X = k) \approx \mathbb{P}(Z = k)$$

Pour illustrer le théorème 5.2 on pourra constater sur les histogrammes de la figure FIG.13 que l'écart entre la loi binomiale (pour $n = 50$ et $p = 0.1$) et la loi de Poisson (pour $\lambda = np = 5$) est inférieur à 1%.

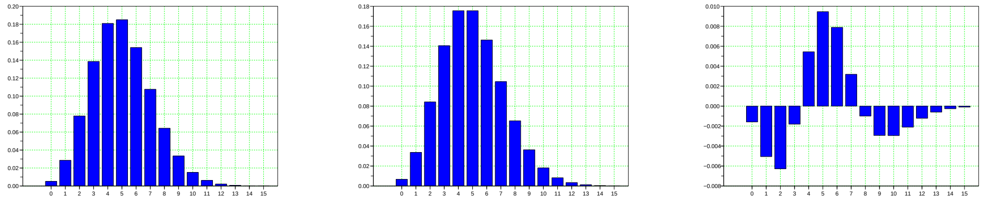


FIG. 13 – *histogramme des lois $\mathscr{B}(n = 50,p = 0.1)$ $\mathscr{P}(\lambda = 5)$ et de leur écart*

l'intérêt de cette approximation est que pour n grand il est beaucoup plus facile de calculer les probabilités d'une loi de Poisson que celles d'une loi Binomiale, à cause des coefficients C_n^k qui nécessitent le calcul de $n!$ ou du triangle de Pascal correspondant. Cependant pour que l'approximation reste valable il faut que np reste petit, de ce fait on dit que la loi de Poisson est bien adaptée pour estimer les « événements rares ».

5.2 La loi des grands nombres

le deuxième résultat de cette section permet de faire le lien entre probabilités et statistiques. La loi des grands nombres indique que lorsque l'on fait un tirage aléatoire dans une série de grande taille, plus on augmente la taille de l'échantillon, plus les caractéristiques statistiques de l'échantillon se rapprochent des caractéristiques statistiques de la population. Cette loi repose sur l'inégalité de Tchebyshev.

Proposition 5.3 (inégalité de Tchebyshev) *Soit X une v.a. positive admettant une espérance $\mathbb{E}(X)$ alors pour tout $t > 0$ on a :*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}$$

Preuve : La démonstration repose sur la définition de l'espérance

$$\mathbb{E}(X) = \sum_{k \in X(\Omega)} k\mathbb{P}(X = k)$$

et le fait que dans cette somme k est toujours positif :

$$\begin{aligned}
\mathbb{P}(X \geq t) &= \sum_{k\in X(\Omega),k\geq t} \mathbb{P}(X=k) \\
&= \sum_{k\in X(\Omega),k\geq t} \frac{k}{k}\mathbb{P}(X=k) \quad \text{car } k \geq 1 \text{ donc non nul} \\
&\leq \sum_{k\in X(\Omega),k\geq t} \frac{k}{t}\mathbb{P}(X=k) \quad \text{car } k \geq t \geq 1 \\
&\leq \frac{1}{t}\sum_{k\in X(\Omega),k\geq t} k\mathbb{P}(X=k) \quad \text{car } \frac{1}{t} \text{ ne dépend pas de } k \\
&\leq \frac{1}{t}\sum_{k\in X(\Omega)} k\mathbb{P}(X=k) \quad \text{on ajoute des termes positifs dans la somme} \\
&\leq \frac{\mathbb{E}(X)}{t}
\end{aligned}$$

□

L'inégalité de Bienaymé-Tchebyshev est une version légèrement modifiée de l'inégalité précédente faisant intervenir l'espérance et l'écart-type de la loi considérée.

Proposition 5.4 (inégalité de Bienaymé-Tchebyshev) *Soit X une v.a. admettant une espérance $\mathbb{E}(X) = \mu$ et une variance $\mathbb{V}\mathrm{ar}(X) = \sigma^2$ alors pour tout $t > 0$ on a :*

$$\mathbb{P}(|X-\mu| \geq \sigma t) \leq \frac{1}{t^2}$$

Preuve : On applique l'inégalité de Tchebyshev à la variable $Y = (X-\mu)^2$ qui est positive et dont l'espérance est :

$$\mathbb{E}(Y) = \mathbb{E}\left((X-\mu)^2\right) = \mathbb{V}\mathrm{ar}(X) = \sigma^2$$

ce qui donne pour tout $s = \sigma^2 t^2 > 0$

$$\mathbb{P}(Y \geq s) \leq \frac{\mathbb{E}(Y)}{s} \iff \mathbb{P}((X-\mu)^2 \geq \sigma^2 t^2) \leq \frac{1}{t^2} \iff \mathbb{P}(|X-\mu| \geq \sigma t) \leq \frac{1}{t^2}$$

car les événements suivants sont égaux :

$$\left[(X-\mu)^2 \geq \sigma^2 t^2\right] = [|X-\mu| \geq \sigma t]$$

□

L'inégalité de Bienaymé-Tchebyshev permet de démontrer la loi faible des grands nombres.

Théorème 5.5 (loi faible des grands nombres) *Soit X_i une suite de v.a. indépendantes de même loi, admettant une espérance $\mathbb{E}(X_i) = \mu$ et une variance $\mathbb{V}\mathrm{ar}(X_i) = \sigma^2$ alors on a :*

$$\mathbb{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \sigma t\right) \leq \frac{1}{nt^2}$$

Preuve : On applique l'inégalité de Bienamé-Tchebyshev à la variable $S_n = \frac{X_1+X_2+\cdots+X_n}{n}$ qui a pour espérance :

$$\begin{aligned}\mathbb{E}(S_n) &= \mathbb{E}\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)\\ &= \frac{\mathbb{E}(X_1)+\mathbb{E}(X_2)+\cdots+\mathbb{E}(X_n)}{n}\\ &= \frac{\mu+\mu+\cdots+\mu}{n} = \mu\end{aligned}$$

et pour variance

$$\begin{aligned}\mathbb{V}\mathrm{ar}(S_n) &= \mathbb{V}\mathrm{ar}\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)\\ &= \frac{\mathbb{V}\mathrm{ar}\,(X_1+X_2+\cdots+X_n)}{n^2}\\ &= \frac{\mathbb{V}\mathrm{ar}(X_1)+\mathbb{V}\mathrm{ar}(X_2)+\cdots+\mathbb{V}\mathrm{ar}(X_n)}{n^2}\\ &= \frac{\sigma^2}{n}\end{aligned}$$

ce qui donne pour tout $s = \sqrt{n}t > 0$

$$\mathbb{P}(|S_n-\mu| \geq \frac{\sigma}{\sqrt{n}}s) \leq \frac{1}{s^2} \Longleftrightarrow \mathbb{P}\left(|S_n-\mu| \geq \sigma t\right) \leq \frac{1}{nt^2}$$

□

On retiendra que c'est la loi des grand nombre qui permet de faire le lien entre les résultats issus des statistiques et ceux issus de la théorie des probabilités. En effet si on passe à la limite (quand $n \to \infty$) dans la loi faible des grands nombres on obtient :

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1+X_2+\cdots+X_n}{n} - \mu\right| \geq \varepsilon\right) = 0$$

ce qui indique que les moyennes empiriques de n répétitions indépendantes d'une même loi convergent vers l'espérance de cette loi $\mu = \mathbb{E}(X_i)$.

Une version plus précise de la loi des grands nombres permet de relier les fréquences observées lors de répétitions d'une v.a. X avec la probabilité théorie $\mathbb{P}(X = k)$.

Théorème 5.6 (intervalle de confiance) *Soit X une v.a. et X_i une série statistique de n valeurs, obtenues comme une suite réalisation indépendantes de X, alors les fréquences observées (f_k) de $X = k$ convergent vers $\mathbb{P}(X = k) = p_k$ avec une erreur proportionnelle à $\frac{1}{\sqrt{n}}$. Plus précisément :*

$$\forall \alpha \in [0,1], \quad \mathbb{P}\left(|f_k - p_k| \geq \frac{1}{2\sqrt{n\alpha}}\right) \leq \alpha$$

Preuve : On va appliquer la loi des grands nombres à la variable suivante :

$$Y = f(X) = \begin{cases} 1 & si \quad X = k \\ 0 & sinon \end{cases}$$

cette nouvelle variable Y une v.a. de Bernoulli de paramètre $p_k = \mathbb{P}(X = k)$. En particulier[2] sa moyenne est $\mu = p_k$ et sa variance $\sigma^2 = p_k(1 - p_k) \leq \frac{1}{4}$. Maintenant appliquons la loi faible des grands nombres à plusieurs réalisation indépendantes de Y. Si on notre $Y_i = f(X_i)$ alors $\frac{1}{n}\sum_i Y_i$ est la fréquence observée dans la série statistique X_i de l'événement $X = k$ et on doit avoir :

$$\mathbb{P}\left(\left|\frac{Y_1 + Y_2 + \cdots + Y_n}{n} - p_k\right| \geq \sigma t\right) \leq \frac{1}{nt^2}$$

quitte à perdre un peu en précision on peut remplacer σ par $1/2$ puisque $\sigma^2 = p_k(1 - p_k) \leq \frac{1}{4}$:

$$\mathbb{P}\left(\left|\frac{Y_1 + Y_2 + \cdots + Y_n}{n} - p_k\right| \geq \frac{t}{2}\right) \leq \frac{1}{nt^2}$$

en particulier pour un seuil de confiance α, fixé à l'avance, si on pose $t = \frac{1}{\sqrt{n\alpha}}$ on a

$$\mathbb{P}\left(\left|\frac{Y_1 + Y_2 + \cdots + Y_n}{n} - p_k\right| \geq \frac{1}{2\sqrt{n\alpha}}\right) \leq \alpha$$

en d'autre terme les fréquences observées de $X = k$ dans la série statistique X_i se trouve dans l'intervalle $[p_k - \frac{1}{2\sqrt{n\alpha}}; p_k + \frac{1}{2\sqrt{n\alpha}}]$ avec une probabilité $1 - \alpha$ et la taille de cet intervalle diminue proportionnellement à $\frac{1}{\sqrt{n}}$. □

✎ **5.1** À partir d'une variable aléatoire X de loi $\mathscr{U}(1,10)$ on génère plusieurs séries statistiques de tailles $n = 10^k$ pour $k = 2,3,4,5$ et on trace les histogrammes comme sur la figure FIG.15. Il est évident que lorsque n tend vers l'infini l'histogramme des fréquences observées se rapproche de plus en plus de l'histogramme correspondant aux probabilités de X qui sont bien les fréquences théoriques attendues lors de cette expérience. On peut même vérifier que l'écart maximal entre fréquence théoriques et fréquences observées est bien dans l'intervalle donné par le théorème 5.6 comme dans le tableau suivant :

n	écart observé	écart théorique
10^2	0.05	0.5
10^3	0.02	0.1581
10^4	0.0047	0.05
10^5	0.00121	0.0158

ici l'écart théorique a été calculé avec pour seuil de défiance $\alpha = 1\%$ ce qui donne un écart théorique de $\frac{1}{2\sqrt{n\alpha}} = \frac{5}{\sqrt{n}}$.

2. Remarquer que pour $x \in [0,1]$ la fonction $h(x) = x(1 - x)$ admet un maximum de $1/4$ en $x = 1/2$

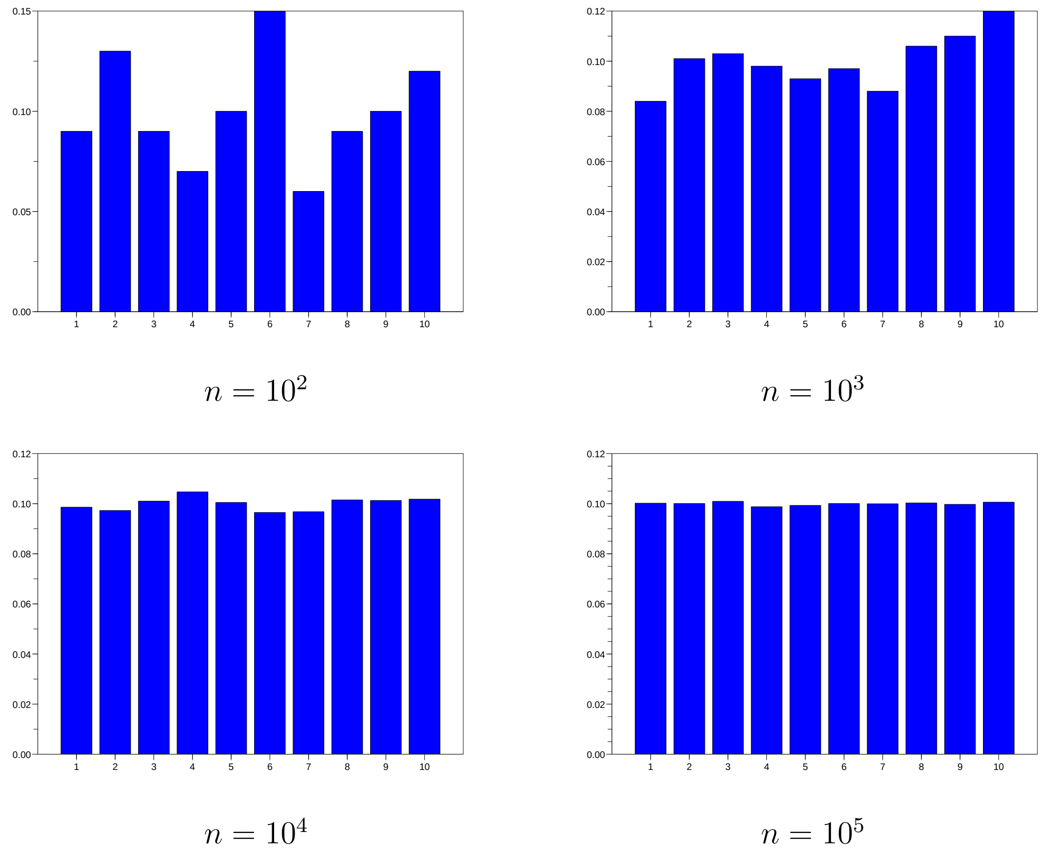


FIG. 14 – *Convergence des fréquences observées vers les fréquences théoriques pour une loi uniforme*

On peut faire le même genre d'études avec une autre loi que la loi uniforme. Par exemple si on considère la v.a. X comptant le nombre de pile obtenus lors de 10 lancers d'une pièce équilibrée, alors $X \sim \mathscr{B}(n = 10, p = 0.5)$ en répétant les essais on verra les fréquences observées converger vers les fréquences théoriques d'une loi binomiale avec la vitesse annoncée par le théorème 5.6.

n	écart observé	écart théorique
10^2	0.0561	0.5
10^3	0.0131	0.1581
10^4	0.0044	0.05
10^5	0.0029	0.0158

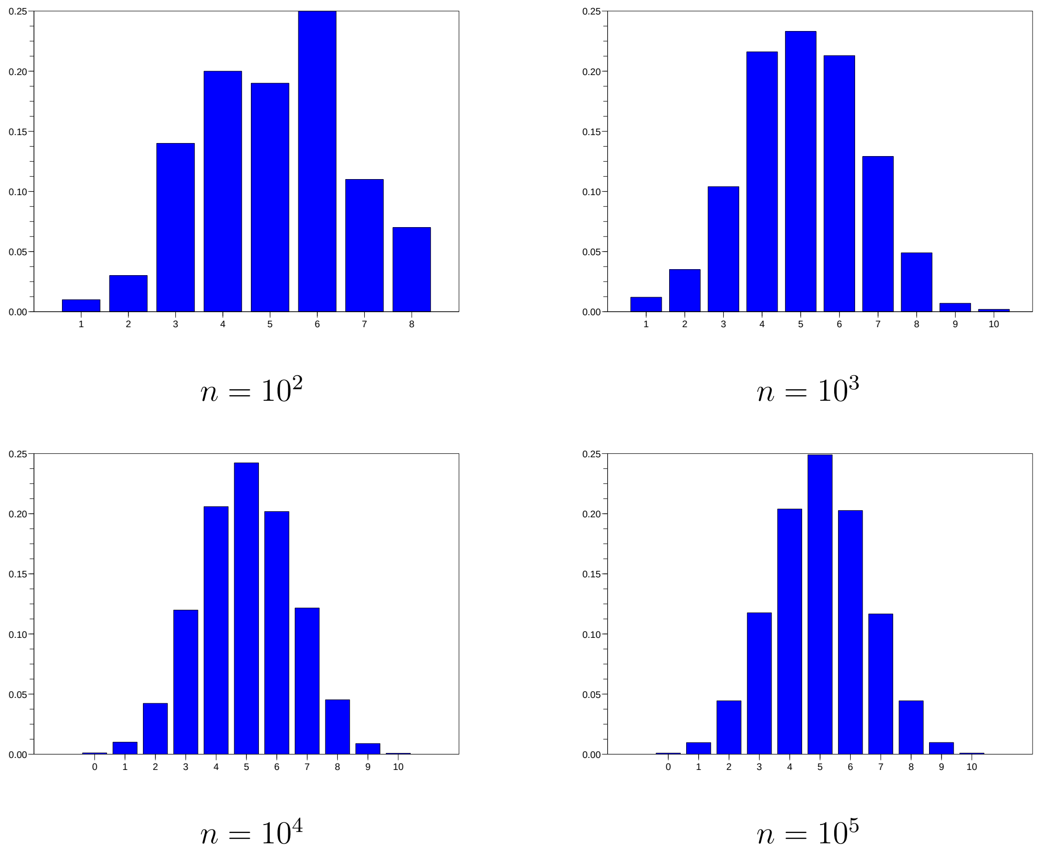


FIG. 15 – *Convergence des fréquences observées vers les fréquences théoriques pour une loi binomiale*

Cette loi possède de nombreuses applications pratiques dont le plus important est certainement la théorie des sondages

✎ 5.2 Théorie des sondages

Quand on veut évaluer l'opinion d'une population on interroge un nombre suffisamment important de personnes (en leur demandant s'il sont « pour »ou « contre »telle proposition) et on fait la moyenne des résultats obtenus. Mathématiquement demander son opinion à une personne correspond à une expérience de Bernoulli que l'on représente par une v.a. de Bernoulli, qui vaut 1 si la personne est « pour »et 0 si elle est « contre », de paramètre p qui est le pourcentage de personnes favorables. Faire le sondage revient donc à estimer la valeur de p en calculant la fréquence observée f de $X = 1$ lors de n réalisations indépendantes[3] de X. Pour estimer la précision du sondage en fonction du nombres n de personnes sondées on peut utiliser le théorème 5.6 :

$$\mathbb{P}\left(|f - p| \geq \frac{1}{2\sqrt{n\alpha}}\right) \leq \alpha$$

3. Pour cela on s'assure que l'on a sondé toutes les catégories socio-professionnelles de la population en utilisant la « méthode des quotas »par exemples

Si on fixe notre taux de confiance dans le sondage à 95% (donc le sondage est erroné dans moins de $\alpha = 5\%$ des cas) alors

- si on a interrogé $n = 1000$ personnes sa précision sera de plus ou moins :

$$\frac{1}{2\sqrt{n\alpha}} = 0.0707107\cdots \approx 7\%$$

- si on veut un résultat précis à 1% près $\frac{1}{2\sqrt{n\alpha}} = 1\%$ alors il faudra interroger

$$n = \frac{1}{4 \times 0.01^2 \times \alpha} = 50000\,\text{personnes}$$

La comparaison de ces chiffres avec ceux diffusés tous les jours dans les médias laissent rêveur sur la manière dont sont utilisées les probabilités dans le domaine de l'information ... mais dans d'autres domaine leur utilisation est beaucoup plus rationnelle. Par exemple, dans les assurances, cette loi permet de fixer le prix des polices d'assurances, en fonction des caractéristiques du client pour être sûr d'obtenir un bénéfice donné à partir de la connaissance des probabilités que les sinistres se réalisent ou pas (et vu les bénéfices des assurances ça doit marcher!).

Nous verrons l'an prochain qu'il est possible d'aller beaucoup plus loin dans l'étude des écarts entre fréquence empirique et fréquence théorique. Dans le cours de probabilité de deuxième année on démontrera le *théorème centrale limite* qui permet d'affirmer que la variable aléatoire correspondant aux écarts à la moyenne tend à suivre une loi particulièrement simple : *la loi normale* ...

Bibliographie

[1] http://www.math93.com/theoreme/probabilites.html
[2] http://fr.wikipedia.org/wiki/Probabilité

Index