Probabilité et Statistiques en Seconde

Maximilien Dreveton

March 19, 2017

Contents

1 Statistique descriptive

1.1 Introduction; le langage statistique

(cf `http://tanopah.jo.free.fr/seconde/statisalpha.php`)

Le mot "statistique" vient du latin "status" qui signifie état. Une étude statistique consiste à observer et à étudier une particularité commune chez un groupe de personnes ou de choses. Les personnes ou les choses sont ce que l'on appelle des individus. Le groupe formé par ces individus est dénommé population. Cette particularité est elle appelée caractère.

Par exemple, on peut étudier les notes obtenues par les élèves d'une classe à un devoir. La population considérée est celle de la classe. Tous les individus de cette population présentent la particularité d'avoir une note. C'est le caractère.

L'étude statistique commence par un recueil de données. Par exemple dans le cadre des notes obtenues par une classe, on a que l'individu A a obtenu par exemple la note de 10.

Les différentes valeurs pouvant être prises par un caractère portent le nom de valeurs. Ainsi par exemple, 10 est une valeur que peut prendre la caractère note. 15 est une autre valeur possible de ce même caractère.

Mais savoir que tel individu présente telle modalité est peu parlant d'un point de vue global. Par contre savoir combien d'individus présentent cette modalité est nettement plus intéressant. Par valeur, on compte est le nombre d'individus qui la présentent.

1.2 Effectifs et fréquences

1.2.1 Vocabulaire. Notion d'effectif et de fréquence

Définition 1. *Une **série statistique** quantitative brute est une liste de valeurs $v_1, \dots, v_n$. Le nombre de valeur est appelé l'**effectif**.*

*On parle aussi de **population**, dans laquelle on étudie un caractère particulier. Le **caractère** peut prendre différente **valeurs** selon les individus de la population.*

Définition 2. *Dans une série statistique.*

*L'**effectif total** est le nombre d'individus dans la population.*

*L'**effectif d'une valeur** est le nombre d'individus de la population dont le caractère prend cette valeur. (le nombre de données correspondant à cette valeur)*

*L'**effectif cumulé croissant** d'une valeur est le nombr d'individus dont le caractère a une valeur inférieure ou égale à celle-ci.*

Exercice 1 : Dans une classe, les notes se répartissent tel quel : 4; 6; 8; 4; 8; 6; 7; 8; 4; 5. Donner l'effectif total, les effectifs de chaque valeur et les effectifs cumulés croissant

Proof. L'effectif total est de 10.

Note	4	5	6	7	8
Effectif	3	1	2	1	3
Effectif cumulé croissant	3	4	6	7	10

□

Définition 3. *La **fréquence** d'une valeur est le quotient de son effectif par l'effectif total.*

$$\textit{fréquence de la valeur} = \frac{\textit{effectif de cette valeur}}{\textit{effectif total}}$$

$$FCC = \textit{fréquence cumulée croissante de la valeur} = \frac{\textit{effectif cumulée croissante de cette valeur}}{\textit{effectif total}}$$

Exercice 2 : Reprendre l'exo précédent en calculant les fréquence d'apparition de chaque note.

Proof.

Note	4	5	6	7	8
Effectif	3	1	2	1	3
Fréquence	0,3	0,1	0,2	0,1	0,3

□

Exercice 3 : Montrer que la somme des fréquence de toutes les valeurs vaut 1.

Montrer que la fréquence cumulée croissante d'une valeur est égale à la somme de toutes les fréquences des valeurs qui lui sont inférieures ou égales.

Proof. La somme des effectifs de toutes les valeurs est l'effectif total (par définition), donc la somme des fréquences de toutes les valeurs vaut 1. □

1.2.2 Représentation graphique

Définition 4. *Nuage de points. Ensemble des points placé dans un repère où l'abscisse sont les valeurs et l'ordonnée les effectifs (ou fréquences) correspondantes.*

Courbe (polygone) des fréquences cumulées croissantes : FCC en fonction de valeur.

Histogramme : rectangles dont la base sont des intervalles et l'aire correspondante proportionnelles aux effectifs ou fréquence de ces intervalles. Idem pour le camembert.

1.2.3 Caractéristiques d'une série statistique

Définition 5. *La moyenne est le quotient de la somme des valeurs de l'échantillon par le nombre total de valeurs (effectif total). Souvent notée $\bar{x}$.*

Proposition 1.1. *Soit $x_1, \ldots, x_p$ les valeurs prises par un échantillon, avec des effectifs respectifs de $n_1, \ldots, n_p$. Alors :*

$$\bar{x} = \frac{\sum_{i=1}^{p} n_i x_i}{\sum i = 1^p n_i} \tag{1}$$

Proof. Dans la moyenne, on regroupe les valeurs égales. □

Exercice 4 : Calcul de moyenne sur l'exemple d'avant.

Définition 6. *Médiane et quartiles.*

La médiane Me d'une série statistique est une valeur telle que la moitié des valeurs de la série lui soit inférieure ou égale et la moitié lui soit supérieur ou égale.

Le premier quartile d'une série statistique, noté Q_1 est la plus petite valeur de cette série telle qu'au moins 25% des valeurs de la série lui soient inférieures ou égales.

Le troisième quartile Q_3 d'une série statistique est la plus petite valeur de cette série telle qu'au moins 75% des valeurs de la série lui soient inférieures ou égales.

Faire un schéma.
Mentionner les déciles.

Exemple Calcul de la médiane en pratique.

- On range les valeurs de la série par ordre croissant.
- Si l'effectif total N est impair, alors la médiane est la valeur de rang $\frac{N+1}{2}$.
- Si l'effectif est pair, on choisit pour médiane la moyenne des valeurs de rangs $\frac{N}{2}$ et $\frac{N+1}{2}$.

Exercice 5 : Calcul de la médiane sur le premier exemple.

Exercice 6 : On trace les fréquence cumulées croissante en fonction des valeurs. Montrer que cette courbe est croissante, et que la médiane est l'antécédent de 0,5.

1.2.4 (*) Éclatement d'une série statistique autour de sa moyenne

Définition 7. *La variance, souvent notée V_X ou Var(X), est la moyenne des carrés des écarts à la moyenne.*

L'écart type, souvent notée σ_X, est la racine carré de la variance.

Proposition 1.2.

$$Var(X) = \frac{n_i(x_i - \bar{X})^2}{\sum_{i=1}^{p} n_i} \tag{2}$$

Proof. Découle de la définition et de la proposition 1.1. □

Proposition 1.3.

$$Var(X) = \frac{1}{n}\sum_{i=1}^{p}(n_i x_i^2) - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{p}(n_i x_i^2 - \bar{X}^2) \tag{3}$$

Proof. Bon exercice d'entrainement de manipulation d'expressions. □

Proposition 1.4. *Soit X et Y deux séries statistiques ayant même moyenne m; et pour variance respective Var(X) et Var(Y).*

Alors, si $Var(X) > Var(Y)$, les valeurs de X sont plus éclatées que celles de Y (Y est plus resseré autour de sa moyenne).

2 Probabilité

"The theory of probability as a mathematical discipline can and should be developed from axioms in exactly the same way as geometry and algebra." Andrei Kolmogorov

2.1 Introduction; vocabulaire

Le mot Hasard est un mot d'origine arabe : *az-zahr*, le dé. Il est apparu en français pour signifier tout d'abord un jeu de dés, puis plus généralement un événement non prévisible, et par extension le mode d'apparition de ce type d'événement.

Expérience aléatoire On peut voir une expérience aléatoire comme une boîte de laquelle sorte des choses (ce ne sont pas forcément des nombres). On ne peut pas prédire, à l'avance, de manière sûre quelle chose va sortir de la boîte. Mais :

- Le résultat de l'expérience est appelé **issue** (ou éventualité) de l'expérience aléatoire
- Les valeurs des issues de l'expérience sont connues à l'avance (l'ensemble des issues possible forme ce que l'on appelle l'**univers**)
- On affecte un poids à chaque valeur de l'ensemble de définition. Ce poids correspond à la probabilité que la variable aléatoire prenne cette valeur : c'est la probabilité élémentaire (ou atomique). La somme des poids fait 1.
- En observant un grand nombre de fois les variables aléatoires qui sortent de la boîte, on peut en déduire le domaine et les poids (loi des grand nombre).

Prenons l'exemple du lancer de dés à 6 faces. Je lance la pièce. L'ensemble $\Omega = \{1; 2; 3; 4; 5; 6\}$ constitue l'ensemble des résultats que je peux observer : c'est l'univers des possibilités.
Je lance la pièce une fois. Le résultat que j'obtiens est la variable aléatoire de l'expérience "lancer de dés". Je ne peux la prédire à l'avance. Par contre, si je lance la pièce un grand nombre de fois, je dois avoir le même nombre de 1 que de 2, 3, 4, 5 et 6 (si le dés est non pipé). Les résultats successif que j'obtient en lançant le dés plein de fois ne sont pas complètement le fruit du hasard : ils suivent une répartition bien précise, donnée par les probabilité de chaque événement élémentaire ("obtenir 1", "obtenir 2", ..., "obtenir 6"). Les probabilités élémentaires sont de $p_j = \frac{1}{6}$.

Exemple Le tableau suivant donne une longue liste d'expériene aléatoire, avec l'espace des possibles associé.

Expérience aléatoire	Univers des possible Ω
Lancer de dés	$\{1; \ldots ; 6\}$
Pile ou Face	{P;F}
Mutation du génome	{A;T;C;G} (les 4 bases de l'ADN)
Élection présidentielle	{ tous les candidats }
Nombre de personnes qui mangent à la cantine	$\mathbb{N}$ (ou $\{0, \ldots, N\}$ avec N nombres d'élèves au lycée)
Espérance de vie d'une ampoule	$[0; +\infty[$
Temps d'attente à l'arrêt de bus (le bus passant toutes les 10 min)	$[0; 10[$
Vitesse d'une molécule de gaz pendant un intervalle de temps $[t_1, t_2]$	Ensemble des fonctions (continues) de $[t_1; t_2]$ dans $\mathbb{R}$ (ou plutôt $[0; c]$, c vitesse de la lumière)
Marche aléatoire en dimension N	$\mathbb{Z}^N$
Position d'une particule	$\mathbb{R}^3$
Ce qui va se passer dans la salle aujourd'hui	aucune idée

Ces exemples sont avant tout là pour illustrer la grande variété d'univers Ω possibles. Il s'agit donc d'élaborer une théorie qui s'affranchit le plus possible de ces univers et créer une théorie qui englobe tous ces cas.

2.2 Définitions

Définition 8. *Un phénomène est dit aléatoire si, reproduit maintes fois dans des conditions identiques, il se déroule chaque fois différemment de telle sorte que le résultat de l'expérience change d'une fois sur l'autre de manière imprévisible.*

Définition 9. *On appelle **expérience aléatoire** une expérience E qui, reproduite dans des conditions identiques, peut conduire à plusieurs résultats possibles, et dont on ne peut prévoir le résultat par avance. L'espace de tous les résultats possibles, appelé univers des possibles (ou espace d'états associé à l'expérience), sera noté Ω. Un résultat possible de l'expérience est noté classiquement ω, et est une **issue** (ou éventualité) de l'expérience.*

Exemple Lancer de dés, Pile ou Face, loterie, jeux de cartes, durée de vie d'une ampoule (ou d'un être humain), sexe à la naissance, temps d'attente à l'arrêt de bus, etc... sont des expériences aléatoire.

Proposition 2.1. *Soit ω une issue d'une expérience aléatoire E, dont l'univers est Ω. Alors $\omega \in \Omega$*

2.3 Evènements

Définition 10. *Un événement est une partie (dit aussi sous-ensemble) de l'univers Ω.*

Si une issue appartient à l'événement A, on dit qu'elle réalise A, ou que A est réalisé par cette issue.

Un événement ne contenant qu'une seule issue est un événement élémentaire.

Exemple Dans le lancé de dés à 6 faces, $\Omega = \{1; \ldots; 6\}$. $A = \{1; 2; 4\}$ est un événement possible ("avoir un 1, un 2 ou un 4"). Calculer $p(\Omega)$ et p(A).

Cas particulier

- Ω est un événement certain : il est toujours réalisé, car il contient toutes les issues.
- $\emptyset$ est un événement impossible : il n'est jamais réalisé.
- Un événement est formé d'une seule issue est un événement élémentaire

Définition 11.
- *L'intersection de A et B, noté $A \cap B$ est l'événement des issues qui appartiennent à la fois à A et à B.*
- *La réunion de A et B, notée $A \cup B$, est l'événement constitué des issues qui appartiennent au moins à l'un des deux événements A ou B*
- *L'événement contraire de l'événement A, noté $\bar{A}$ (ou A^c) est l'événement constitué de toutes les issues qui n'appartiennent pas à A.*

Donner des exemples et faire les diagrammes.

2.4 (**) Probabilités d'événements

Dans la suite, on travaille uniquement dans le cas où l'univers Ω est fini.

Définition 12. *Soit Ω un ensemble fini et $\mathcal{F}$ l'ensemble des parties de Ω.*

Soit $\mathbb{P} : \mathcal{F} \mapsto [0; 1]$ tel que :

$$(a) \quad \mathbb{P}(\Omega) = 1 \tag{4}$$

$$(b) \quad \forall A, B \in \mathcal{F}, A \cap B = \emptyset \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \tag{5}$$

On dit que $\mathbb{P}$ est une mesure (ou loi) de probabilité sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$

Proposition 2.2. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, avec Ω fini. Notons $\omega_1, \ldots, \omega_n)$ les éléments de Ω.*

Alors la probabilité P sur Ω est entièrement caractérisée par ses probabilités atomiques.

Proof. On va montrer que pour toute partie A de Ω, on peut exprimer P(A) en fonction des probabilités élémentaires $p(\omega_i)$, $i = 1, \ldots, n$.

En effet, A est composé de ω_i, pour certains i : c'est donc l'union disjointe des $\{\omega_i\}$, donc avec 5

$$P(A) = \sum_{\omega \in A} p(\omega)$$

□

Remarque culturelle La proposition 2.4 devient fausse lorsque Ω est un ensemble avec un nombre infini d'éléments (et non dénombrable).

2.5 Récapitulatif - Reformulation

On considère une expérience aléatoire dont l'univers des possibles Ω est fini :

$$\Omega = \{e_1; e_2; \dots; e_n\} \tag{6}$$

Définir une loi de probabilité sur cet univers Ω, c'est associer à chaque événement élémentaire $\{e_i\}$ un réel positif ou nul p_i. Les réels p_i doivent de plus vérifier :

$$p_1 + \cdots + p_n = 1 \tag{7}$$

Le réel p_i est appelé la probabilité de l'événement élémentaire $\{e_i\}$.

Définition 13. *Soit A un sous ensemble de Ω. Alors la probabilité de l'événement A (noté $p(A)$) est la somme des probabilités des événements élémentaires qui composent A.*

Exercice 7 : Soit un dé à 6 faces pipé, tel que :

$$p(1) = \frac{1}{6} \quad p(2) = \frac{1}{12} \quad p(3) = \frac{1}{3} \quad p(4) = \frac{1}{4} \quad p(5) = \frac{1}{12} \quad p(6) = \frac{1}{12}$$

Vérifiez que l'on a bien une loi de probabilité. Quelle est la probabilité d'obtenir un nombre impair. Un nombre supérieur ou égal à 2 ? Quel est l'événement contraire de l'événement *Obtenir un nombre supérieur ou égal à 2.* Quelle est sa probabilité ?

2.6 Situation d'équiprobabilité

Définition 14. *On parle d'équiprobabilité lorsque toutes les probabilités élémentaires sont égales.*

Exemple Lancer de dés non pipé.
Contre-exemple : somme obtenue après le lancer de 2 dés.

Proposition 2.3. *Dans une situation d'équiprobabilité ayant n issues possibles (ie ayant n événements élémentaires), les probabilités élémentaires valent toutes $\frac{1}{n}$.*

Exemple Lancer de dé à 6 faces; 6 événements élémentaires

Proof. Soient $e_1, \dots, e_n$ les probabilités élémentaires sont égales, (car équiprobabilité) notons les p. Leur somme vaut 1, donc np=1, d'où le résultat. □

Proposition 2.4. *Dans une situation d'équiprobabilité ayant n issues possibles, la probabilité d'un événement A est égale à :*

$$p(A) = \frac{\textit{nombres d'événements élémentaires constituant } A}{\textit{nombres total d'événement élémentaires}}$$

Proof. Soit une situation d'équiprobabilité et A un événement constitué de m issues. Alors

$$p(A) = \frac{1}{n} + \cdots + \frac{1}{n} = \frac{m}{n}$$

□

2.7 Lien probabilité - fréquence

Motivation : trouver expérimentalement la probabilité d'une expérience aléatoire. On va donc réaliser l'expérience aléatoire un grand nombre de fois, et espérer pouvoir en déduire les probabilité des événements constituant l'univers.

Exemple Je vous donne un dé. Déterminer une méthode pour savoir si celui-ci est truqué ou non.

Exemple Je demande aux élèves de lancers 50 fois une pièce de monnaie (non truquée). Une élève paresseuse refuse de faire l'exercice et sort une suite "au hasard" de P et F. Ai-je un moyen de prouver qu'elle n'a pas fait ce qui était demandé ?

Théorème 2.5. *Loi des grands nombres*
Si on effectue n fois une expérience aléatoire, alors la fréquence de réalisation d'un événement se rapproche de la probabilité de cet événement lorsque n devient très grand.

Proof. Largement hors programme. □

2.8 Calculs de probabilités

Définition 15. *Deux événements A et B sont dit incompatibles si* $A \cap B = \emptyset$.

Proposition 2.6. *Soient A et B deux événements incompatibles. Alors* $p(A \cup B) = p(A) + p(B)$

Exemple Lancer de carte dans un jeu à 32 cartes. A=*Tirer un coeur* et B = *tirer un pique.* On a bien A et B incompatibles, et $p(A \cup B) = p(A) + p(B)$

Proposition 2.7. *Soient A et B deux événements. Alors :*

$$P(A) + P(B) = P(A \cup B) + P(A \cap B) \tag{8}$$

Proof. On remarque que si $A \cap B = \emptyset$, alors on retrouve bien l'axiome 5 dans la définition d'une loi de probabilité.

Notons A_1 l'événement dont les issues sont celles de A qui n'appartiennent pas à B. Les événements A_1 et B sont incompatibles et on a $A_1 \cup B = A \cup B$.

Donc :

$$P(A \cup B) = P(A_1 \cup B) = P(A_1) + P(B)$$

Les événements A_1 et $A \cap B$ sont incompatibles, et $A_1 \cup (A \cap B) = A$.

Donc $P(A) = P(A_1) + P(A \cap B)$, d'où $P(A_1) = P(A) - P(A \cap B)$.

Finalement, on a :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

□

Définition 16. *Soit A u, événement, on note* $\bar{A}$ *sont événement contraire : c'est l'événement constitué de toutes les issues qui n'appartiennent pas à A (mais qui appartiennent à* Ω*).*

Proposition 2.8. *Soit A un événement. On a :*

$$P(A) + P(\bar{A}) = 1$$

Proof. A et $\bar{A}$ sont incompatibles, et leur réunion vaut l'univers Ω tout entier (donc $p(A \cup \bar{A}) = 1$. On applique ensuite la proposition 2.6 □

2.9 Exemples ultra-classiques

Exercice 8 : Somme de deux dés.

Exercice 9 : Un square est équippé de 3 bancs à deux places. Deux personnes arrivent successivement et s'installent au hasard. Quelle est la probabilité que les deux personnes s'assoient à côté ?

Proof. On propose deux manières. On suppose l'équiprobabilité dans les choix des deux personnes.

Première réponse On numérote les places 1,2,3,4,5 et 6, avec les paires (1,2), (3,4) et (5,6) correspondant à des places sur les même bancs.

Il y a en tout 15 paires possibles, chacune avec une probabilité $\frac{1}{15}$.

Alors P(assis à côté)= P(1,2)+P(3,4)+P(5,6)=$\frac{3}{15} = \frac{1}{5}$

Deuxième réponse On numérote les trois bancs A,B et C. Les résultats de l'expérience sont codés par des couples du type (B,A) : première personne sur le banc B et deuxième sur le banc A.

Il y a en tout 9 couples possibles, chacun avec une probabilité $\frac{1}{9}$.

Donc P(assis à côté)=P(A,A)+P(B,B)+P(C,C)=$\frac{3}{9} = \frac{1}{3}$. □

Exercice 10 : On choisit une commune au hasard parmi toutes les communes de France.

Quelle est la probabilité que l'événement A : "cette commune soit dans la région Rhône-Alpes" ?

Quelle est la probabilité de l'événement B : "cette commune a plus de 3500 habitants".

Définir les événements $A \cap B$ et $A \cup B$ et calculer leurs probabilités respectives.

Exercice 11 : Dans une classe, quelle est la probabilité que deux élèves est la même date d'anniversaire ? Supposons qu'il y ait des jumeaux. Que vaut cette probabilité ?

Exercice 12 : Marche aléatoire.

2.10 Applications

Exemple Algorithme qui permet de trouver le maximum d'une fonction (sur un intervalle [a,b]).
Version déterministe : prendre un N entier et chercher le max des $f(a + k \times \frac{b-a}{N})$ où $k \in \{1, \ldots, N\}$.
Version probabiliste : On initialise $max = f(a)$. On choisit aléatoirement des $x \in [a; b]$. Si $f(x) > max$ alors on change max en $f(x)$. Sinon on garde max. Fair ça pour N valeurs aléatoires de x.

Exemple Monte-Carlo

Exemple Une fonction est-elle monotone?
Version déterministe & probabiliste.

Observation	Formulation ensembliste
Evènement élémentaire	$\{\omega\}, \omega \in \Omega$
Evènement	$A, A \subset \Omega$
$A \Rightarrow B$	$A \subset B$
A ou B	$A \cup B$
A et B	$A \cap B$
Absence de A	$\bar{A}$
Évènement impossible	$\emptyset$
Évènement certain	Ω
Évènement incompatibles	$A \cap B = \emptyset$

3 Échantillonnage

On étudie un caractère dans une population. La connaissance de la population entière n'est en général ni envisageable ni possible. Par contre, on peut connaître un échantillon de cette population (penser aux sondages : avec moins de 1 000 personnes, on en déduit les résultats d'une élection, et malgré les critiques cela s'avère souvent être correct). Par contre, pour des produits de beauté, on a souvent des études "80% d'utilisateurs satisfait", suivi d'une petite astérisque précisant que l'étude a été effectuée sur 50 personnes. La question est donc d'estimer l'erreur faite lorsque l'on interroge uniquement n personnes.

Définition 17. *On appelle* ***échantillon de taille n*** *une liste de n résultats obtenus par n répétition indépendantes d'une même expérience aléatoire.*

Exemple On demande à n personnes pour qui ils vont voter; on regarde la durée de vie de n ampoules; etc...

Évidemment, sauf gros cas particulier, les n résultats ne sont pas identiques: il y a des **fluctuations d'échantillonnage**.

Théorème 3.1. *On étudie un caractère ayant une probabilité p d'être réalisé.*
La probabilité que la fréquence du caractère observé appartienne à l'intervalle $[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}]$ est de 95%.

Proof. Largement hors programme. Le résultat écrit tel quel est d'ailleurs (légèrement) faux. Sera éventuellement démontré en terminale dans des cas simples. □

Ce résultat impose quelques remarque.

Définition 18. *L'intervalle $[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}]$ est appelé intervalle de fluctuation de la fréquence f au seuil 95% (ou de risque 5%).*