

Chapitre 14

Loi Binomiale

Dans ce chapitre nous allons étudier une loi de probabilité sur un ensemble discret : la loi binomiale. Ce sera l'occasion de faire un peu de dénombrement, de parler de répétitions d'expériences aléatoires indépendantes et de faire des estimations statistiques.

14.1 Introduction

Pour mieux comprendre l'aléa que nous allons étudier, il semble préférable de débiter par des exemples.

- Exemple 14.1.1.**
1. M. Tanguy, dans sa jeunesse, a du passer des oraux de mathématiques. En analyse, il y avait 67 leçons à apprendre et, le jour de l'épreuve, le candidat se retrouve à choisir parmi 2 sujets tirés au sort. Combien de leçons doit-il apprendre pour être certain de connaître au moins un des sujets tirés au sort avec une probabilité supérieure ou égale à 0.75 ?
 2. En lisant Crime et châtiment, Coline et Séraphin, ont trouvé 2 fautes de frappes dans une page contenant 250 mots. Sachant que le livre est composé de 1360 pages, comment estimer (avec une marge d'erreur de 5%) le nombre total de fautes de frappes ?
 3. Un fabricant de composants électroniques sait que 2% de ses produits possède un défaut (lesquels sont supposés indépendants). Un distributeur se fait livrer 350 de ces composants électronique, quelle est la probabilité que 25 d'entre eux soit défectueux ?

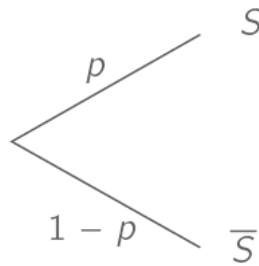
Pour répondre à ces questions, il va nous être nécessaire d'introduire de nouveaux objets mathématiques. Commençons par le plus simple d'entre eux.

14.2 Loi de Bernoulli

Dans cette section, nous considérons une expérience aléatoire très simple puisqu'elle ne comporte que deux issues.

Définition 14.2.1 (Epreuve de Bernoulli). Soit $p \in [0, 1]$, une épreuve de Bernoulli est une expérience aléatoire comportant **deux issues** : un **succès** S de probabilité p et un **échec** \bar{S} de probabilité $1 - p$.

Remarque. De manière schématique, cela peut se représenter par un arbre pondéré :



où S désigne l'issue appelée « succès ».

Voyons quelques exemples d'épreuves de Bernoulli.

- Exemple 14.2.1.** 1. Le Jeu de pile ou face (avec une pièce équilibrée) est une épreuve de Bernoulli car elle ne présente que deux issues. Si on désigne le succès S par le fait d'obtenir pile alors $\mathbb{P}(S) = p = \frac{1}{2}$.
2. Un sac contient 10 boules, indiscernables au touché, numérotées de 1 à 10. L'expérience aléatoire consiste à tirer une boule du sac. Le joueur gagne si le numéro est inférieur ou égale à 3. Il s'agit d'une épreuve de Bernoulli pour laquelle le succès est « obtenir un numéro inférieur à 3 » de probabilité $p = \frac{3}{10}$; l'échec est l'évènement « obtenir un numéro strictement supérieur à 3 » sa probabilité vaut $1 - 0,3 = 0,7$.
3. De nombreux exemples sont envisageables : jouer au loto (on gagne le gros lot ou non), conduire une voiture (avoir un accident ou non),...

Remarquons enfin que la notion de succès n'a pas forcément le même sens qu'en français (avec un aspect positif).

Exemple 14.2.2. Supposons qu'un industriel fabrique des ampoules. Pour chaque ampoule, il effectue l'épreuve de Bernoulli suivante : ou bien l'ampoule a un défaut ou bien elle n'en a pas. Il semble naturel que l'industriel cherche à identifier les ampoules défectueuses. C'est donc normal de définir le succès comme étant l'évènement « l'ampoule a un défaut ».

A une épreuve de Bernoulli, il est toujours possible d'associer une variable aléatoire.

Définition 14.2.2. *Considérons une épreuve de Bernoulli de paramètre p (la probabilité du succès), la variable aléatoire X qui prend la valeur 1 en cas de succès et la valeur 0 en cas d'échec est appelée **variable aléatoire de Bernoulli**; ceci sera noté $X \sim \mathcal{B}(p)$. Sa loi de probabilité est donnée par le tableau ci-dessous :*

Valeur x_i	0	1
$\mathbb{P}(X = x_i)$	$1 - p$	p

Voyons cela sur un exemple.

Exemple 14.2.3. Soit l'épreuve de Bernoulli qui consiste à lancer un dé tétraédrique et dont le succès est S : « obtenir la face 3 ». Puisqu'il s'agit d'une situation d'équiprobabilité, nous avons

$$p = \mathbb{P}(S) = \mathbb{P}(\text{obtenir la face 3}) = \frac{1}{4}.$$

Soit X la variable aléatoire qui vaut 1 en cas de succès, 0 sinon. Nous avons donc

$$\mathbb{P}(X = 1) = \mathbb{P}(\text{obtenir la face 3}) = \frac{1}{4} \quad \text{et} \quad \mathbb{P}(X = 0) = \mathbb{P}(\overline{S}) = \frac{3}{4}.$$

X est une variable aléatoire de Bernoulli de paramètre $p = \frac{1}{4}$.

Exercices à traiter : 39/40 page 383.

Lorsque X suit une loi de Bernoulli, il est assez simple de déterminer son espérance, sa variance et son écart-type.

Proposition 86. *Soit X une variable aléatoire de Bernoulli de paramètre $p \in [0; 1]$ alors*

- $\mathbb{E}[X] = p.$
- $\text{Var}(X) = p(1 - p).$
- $\sigma(X) = \sqrt{p(1 - p)}.$

Remarque. Nous rappelons le fait suivant, si X est une variable aléatoire discrète, que x_1, \dots, x_k sont les k valeurs prises par X avec probabilités $p_i = \mathbb{P}(X = x_i)$ pour tout $i \in \{1, \dots, k\}$ alors

$$\mathbb{E}[X] = \sum_{i=1}^k x_i p_i \quad \text{et} \quad \text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \quad \text{avec} \quad \mathbb{E}[X^2] = \sum_{i=1}^k x_i^2 p_i.$$

Démonstration. Grâce à la remarque, la démonstration est élémentaire puisque $X = 0$ ou $X = 1$ avec probabilités respectives $1 - p$ et p . □

Bien entendu, cette loi de probabilité est encore trop simple pour répondre aux questions de l'introduction. Comme nous allons le voir, elle vont servir de briques élémentaires permettant de construire la loi binomiale.

14.3 Schéma de Bernoulli

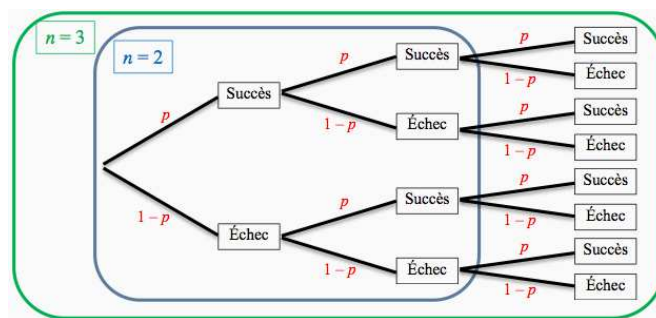
Reprenons l'exemple de l'industriel qui cherche à identifier les composants électroniques présentant un défaut. Après avoir effectué son test pour le premier composant, il paraît naturel qu'il recommence à nouveau pour les autres afin de passer son stock en revu. Ce procédé porte le nom de schéma de Bernoulli. Implicitement, nous supposons que le fait qu'un composant soit défectueux n'a aucun impact sur les autres : il s'agit de la notion d'indépendance.

Remarque. Rappelons, qu'en probabilité, deux évènements A et B sont dit indépendants lorsque $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$ (cf. cours sur les probabilités conditionnelles de 1ère).

Exercice à traiter : 33 page 382.

Définition 14.3.1 (Schéma de Bernoulli). *Considérons une épreuve de Bernoulli de paramètre de succès $\mathbb{P}(S) = p \in [0; 1]$. Un schéma de Bernoulli est la répétition de cette épreuve de Bernoulli de manière **indépendante** (autrement dit, il est supposé que le résultat d'une épreuve n'affecte pas le résultat des autres).*

Remarque. Schématiquement, un schéma de Bernoulli consiste à agrandir l'arbre d'une épreuve de Bernoulli autant de fois que de le nombre de répétitions (ci-dessous, nous avons 2 et 3 répétitions) :



Rappelons que nous utilisons les notations suivantes S : « succès » et \bar{S} : « échec ».

Voyons quelques exemples de schéma de Bernoulli.

- Exemple 14.3.1.**
1. Reprenons l'exemple de l'industriel et supposons que la probabilité qu'une ampoule soit défectueuse soit égale à 0,003 (autrement dit $\mathbb{P}(S) = 0,003$). Si l'industriel teste son stock de 4000 ampoules, il s'agit d'un schéma de Bernoulli avec 4000 répétitions.
 2. Un sac contient 10 boules, indiscernables au touché, numérotées de 1 à 10. Le joueur tire successivement et avec remise 4 boules du sac (on remet la boule obtenue dans le sac après chaque tirage). Le joueur gagne si le numéro est inférieur ou égale à 3.

Ici, l'épreuve de Bernoulli consiste à trier une boule du sac ; le succès correspond à obtenir « un numéro inférieur ou égale à 3 » avec $p = 0,3$. Cette épreuve est répétée quatre fois de façon indépendantes (puisque les tirages sont avec remise). Il s'agit d'un schéma de Bernoulli.

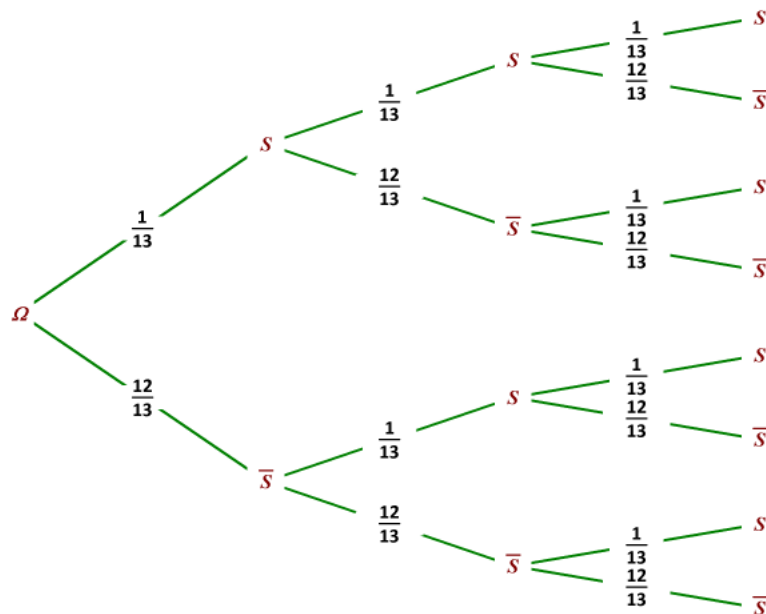
Maintenant que les définitions ont été donné et illustré, il est important de maîtriser les propriétés suivantes :

- Un **schéma de Bernoulli est représenté par un arbre pondéré** dont la **taille** dépend du **nombre de répétitions** (lesquelles doivent être identiques et indépendantes).
- Sur chaque branche de l'arbre est écrite la probabilité associée.
- La **probabilité d'un chemin est le produit des probabilités figurant sur les branches**¹.

Voyons un exercice que le lecteur pourra consulter pour se rappeler comment utiliser des arbres pondérés.

Exemple 14.3.2. Supposons que nous ayons à disposition un sac contenant 12 boules noires et 1 boule blanche ; ces boules sont supposées indiscernables au touché. Nous tirons une boule au hasard dans le sac, notons sa couleur puis replaçons la boule dans le sac. Nous répétons ce tirage trois fois de suite et appelons succès le fait d'obtenir une boule blanche.

1. Il s'agit bien d'une épreuve de Bernoulli puisque l'expérience aléatoire ne comporte que deux issues (on obtient une boule noire ou une boule blanche). Comme les boules sont indiscernables au touché, nous savons que $\mathbb{P}(S) = \frac{1}{13}$.
2. Ensuite, il convient de représenter la situation à l'aide d'un arbre.



3. Grâce à cet arbre, nous pouvons déterminer la probabilité d'obtenir 3 boules noires. Il est facile de réaliser que cet évènement correspond au chemin le plus bas dans l'arbre (i.e. obtenir $\bar{S}\bar{S}\bar{S}$). Nous avons donc

$$\mathbb{P}(\text{obtenir trois boules noires}) = \frac{12}{13} \times \frac{12}{13} \times \frac{12}{13} \approx 0,786$$

Nous avons multiplié entre elles les probabilités apparaissant sur le chemin correspondant à l'évènement « obtenir trois boules noires ».

4. Imaginons que nous souhaitons calculer la probabilité d'obtenir deux boules blanches. Nous constatons que plusieurs chemins correspondent à cet évènement : le chemin $SS\bar{S}$, le chemin $S\bar{S}S$ et le chemin $\bar{S}SS$. Pour chacun d'entre eux, nous obtenons les probabilités suivantes :

$$\mathbb{P}(SS\bar{S}) = \frac{1}{13} \times \frac{1}{13} \times \frac{12}{13} \approx 0,005 \quad ; \quad \mathbb{P}(S\bar{S}S) = \frac{1}{13} \times \frac{12}{13} \times \frac{1}{13} \approx 0,005$$

1. Et ceci est une conséquence de l'hypothèse d'indépendance

et

$$\mathbb{P}(\bar{S}SS) = \frac{12}{13} \times \frac{1}{13} \times \frac{1}{13} \approx 0,005$$

Enfin, pour déterminer la probabilité de l'évènement « obtenir deux boules blanches » il suffit d'additionner ces trois résultats :

$$\mathbb{P}(\text{obtenir deux boules blanches}) = \mathbb{P}(SS\bar{S}) + \mathbb{P}(S\bar{S}S) + \mathbb{P}(\bar{S}SS) \approx 3 \times 0,005 = 0,015.$$

5. Enfin, il est possible qu'on nous demande quelle est la probabilité de l'évènement : « obtenir au moins une boule noire ». En réfléchissant, nous voyons qu'il y a de nombreuses possibilités (obtenir 1 boule noire, obtenir 2 boules noires ou obtenir 3 boules noires). Nous pourrions procéder comme dans le point précédent mais cela serait un peu laborieux. Mieux vaut utiliser un argument qui utilise une formule vue en seconde (pour tout évènement A , nous avons l'égalité $\mathbb{P}(A) = 1 - \mathbb{P}(\bar{A})$) :

$$\begin{aligned} \mathbb{P}(\text{obtenir au moins une boule noire}) &= 1 - \mathbb{P}(\overline{\text{obtenir au moins une boule noire}}) \\ &= 1 - \mathbb{P}(\text{ne pas obtenir de boule noire}) \\ &\approx 1 - \frac{1}{13} \times \frac{1}{13} \times \frac{1}{13} = 0,999. \end{aligned}$$

Exercice à traiter : 1 page 369 (modélisation et utilisation d'un arbre) et 3 page 371 (schéma de Bernoulli) ; 50 page 384 (à la maison).

14.4 Loi binomiale

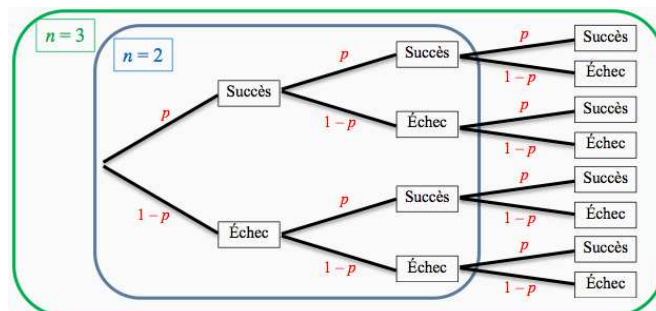
La loi binomiale est la loi naturellement associée à un schéma de Bernoulli et cette modélisation va nous permettre de répondre à certaines questions de l'introduction.

Définition 14.4.1 (Loi binomiale). Soit X la variable aléatoire qui **compte le nombre de succès** dans un schéma de Bernoulli comportant n répétitions **identiques et indépendantes** et dont la probabilité de succès vaut $p \in [0; 1]$. Nous dirons que cette variable aléatoire suit une loi binomiale de paramètre n et p (nous noterons cela par $X \sim \mathcal{B}(n; p)$). De plus, sa loi est donnée par

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{pour tout } k \in \{0; \dots; n\}.$$

Remarque. 1. Nous avons donc une formule pour calculer directement la probabilité d'obtenir k succès sur n répétitions : étant donné n et p , il suffit ensuite de remplacer k par la valeur adéquate afin d'obtenir la probabilité de succès associée.

2. Prenons un instant pour expliquer ce que représente la quantité $\binom{n}{k}$ (lire « k parmi n ») ; la valeur $\binom{n}{k}$ est aussi désignée sous l'appellation **coefficient binomial**. Elle correspond au nombre de façon de choisir k éléments dans une liste comportant n éléments. Ainsi, en considérant un schéma de Bernoulli avec n répétitions, cela correspond **au nombre de chemins contenant exactement k succès parmi les n répétitions**. Nous avons déjà fait ce genre de calculs, à la main, en utilisant un arbre. Par exemple, ci-dessous, nous avons $n = 3$ et il y a précisément 3 chemins comportant exactement 1 succès : autrement dit $\binom{3}{1} = 3$



Bien entendu, lorsque n est plus grand ($n = 10$ ou $n = 20$ par exemple) il ne semble pas raisonnable de dessiner l'arbre associé. Si $n = 10$, pour déterminer le nombre de chemins comportant exactement 3 succès, il suffit de calculer 3 parmi 10 (i.e. $\binom{10}{3}$) à l'aide de la calculatrice. Voici une vidéo expliquant la manipulation à faire :

<https://www.youtube.com/watch?v=ZiY1AWJu8tg&list=PLVUDmbpupCapZdNo8QaVDkWz3eqneIYK1&index=29>

Cette année, il sera important d'être capable de faire cette manipulation.

3. Pour certains, dans l'optique de l'enseignement supérieur, il pourrait être utile d'être capable de faire ce calcul à la main. Pour cela, il est nécessaire d'introduire définition suivante : pour tout $n \in \mathbb{N}$ et tout entier $k \leq n$,

$$\binom{k}{n} = \frac{n!}{k!(n-k)!} \quad (14.4.1)$$

où, pour tout entier $k \in \mathbb{N}$, $k! = k \times (k-1) \times (k-2) \dots \times 2 \times 1$ et, par convention, $0! = 1$; il s'agit de la notion de *factorielle*. Ainsi, par exemple, nous avons

$$\binom{10}{3} = \frac{10!}{3!(10-3)!} = \frac{10!}{3!7!}$$

or

$$\frac{10!}{3!7!} = \frac{10 \times 9 \times 8 \times 7 \times \dots \times 1}{3! \times 7 \times 6 \times \dots \times 1} = \frac{10 \times \dots \times 8}{3!} = \frac{10 \times 9 \times 8}{3 \times 2 \times 1} = 10 \times 3 \times 4 = 120$$

après avoir simplifier la fraction à plusieurs reprises. En français (i.e. en expliquant cette notion avec des phrases), $k!$ (lire « k factorielle ») correspond **au nombre de mots qu'il est possible de former à partir de k lettres**.

4. Mentionnons en passant que le triangle de Pascal, lié aux identités remarquables, est un moyen pratique pour déterminer les coefficients binomiaux $\binom{n}{k}$:

$$\begin{array}{rcccccc} n = 0 : & & & & & & 1 \\ n = 1 : & & & & 1 & & 1 \\ n = 2 : & & & 1 & 2 & 1 & \\ n = 3 : & & 1 & 3 & 3 & 1 & \\ n = 4 : & 1 & 4 & 6 & 4 & 1 & \end{array}$$

Pour chaque ligne, la valeur de k vaut 0 pour le terme le plus à gauche puis augmente de 1 jusqu'à atteindre la valeur n sur le terme le plus à droite. Ceci est notamment utilisé dans la formule du binôme 13 introduite dans un chapitre précédent.

Exercices à traiter : 5/6 page 373. (reconnaitre une loi binomiale et utilisation calculatrice); 7/8 page 373 (utilisation de la calculatrice pour la loi binomiale). Voici un tutoriel pouvant aider :

<https://www.youtube.com/watch?v=ZiY1AWJu8tg>

Voici un exemple permettant de dresser un bilan de tout ce qui précède.

Exemple 14.4.1. 1. Reprenons notre exemple d'une urne contenant 1 boules blanche et 12 boules noires. Le succès correspondait à obtenir une boule blanche (i.e. $\mathbb{P}(S) = \frac{1}{13}$), il y avait $n = 3$ répétitions indépendantes. Si X est la variable aléatoire qui compte le nombre de succès (après ces 3 répétitions) nous savons que

$$X \sim \mathcal{B}(3; \frac{1}{13}).$$

De plus, la probabilité d'obtenir deux succès (i.e. deux boules blanches sur les trois répétitions avec remise) vaut

$$\mathbb{P}(X = 2) = \binom{3}{2} \left(\frac{1}{13}\right)^2 \left(\frac{12}{13}\right)^{3-2} = 3 \times \left(\frac{1}{13}\right)^2 \left(\frac{12}{13}\right)^1 \approx 0,015.$$

2. Un QCM est composé de 3 questions dont chacune d'entre elles possède 4 propositions dont une seule est exacte. Un élève répond au hasard au QCM et nous supposons que ses réponses sont indépendantes les unes des autres, notons alors X la variable aléatoire qui compte le nombre de réponses correctes données par l'élève.
- Justifier que la situation peut être modélisée par une loi binomiale dont les paramètres seront à préciser.
 - Calculer $\mathbb{P}(X = 2)$ et interpréter le résultat obtenu. Arrondir au millième
 - Calculer la probabilité que l'élève ait au plus deux réponses correctes ? Arrondir au millième

Remarque. La calculatrice peut-être d'une aide précieuse pour déterminer ces probabilités :

<https://www.youtube.com/watch?v=7k4ZYdfWEY8&list=PLVUDmbpupCapZdNo8QaVDkWz3eqneIYK1&index=27>

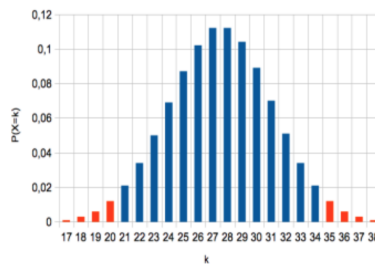
Exercices à traiter : 62,63 page 385.

Comme pour la loi de Bernoulli, il est possible de déterminer les paramètres probabilistes associés à une loi binomiale.

Proposition 87. Si $X \sim \mathcal{B}(n; p)$ alors

- $\mathbb{E}[X] = np$.
- $\text{Var}(X) = np(1 - p)$.
- $\sigma(X) = \sqrt{np(1 - p)}$.

Remarque. Graphiquement, la loi binomiale peut se représenter à l'aide d'un histogramme symétrique, centré (environ) en son espérance. Celui-ci à l'allure d'une « cloche »²



Plus l'écart-type est petit (lorsque p est proche de 0 ou de 1), plus la courbe semble « haute et étroite ».

Exercices à traiter : 9 page 375 (représentation graphique) et 64 page 385.

Avant d'aborder un dernier exemple, voyons de quelle manière utiliser sa calculatrice pour déterminer des « quantiles » associés à un seuil donné.

Exercices à traiter : 15/16 page 378 et 17/18 page 379 et 18/19 page 379 (facultatifs : 93/94 et 97/98 page 388). Le lien suivant peut-être utile à ce sujet :

<https://www.youtube.com/watch?v=7k4ZYdfWEY8&list=PLVUDmbpupCapZdNo8QaVDkWz3eqneIYK1&index=27>

Définition 14.4.2 (Intervalles de fluctuations). Soit X une variable aléatoire et $\alpha \in]0; 1[$ un seuil donné. Nous dirons que

- I est un intervalle de fluctuation de seuil α si $\mathbb{P}(X \in I) \geq 1 - \alpha$.
- $I = [a, b]$ est un intervalle de fluctuation centré de seuil α si $\mathbb{P}(X < a) < \frac{\alpha}{2}$ et $\mathbb{P}(X > b) < \frac{\alpha}{2}$.

Remarque. Un intervalle de fluctuation centré vérifie aussi $\mathbb{P}(a \leq X \leq b) \geq 1 - \alpha$ (cf. graphique section 5).

Exemple 14.4.2. Nous faisons l'hypothèse que 55% des électeurs ont voté pour le candidat A . Nous interrogeons ensuite, au hasard, à la sortie des urnes 50 individus sur leur vote. Bien entendu, il est supposé que la réponse d'une personne est indépendante des autres réponses obtenues.

Soit X est la variable aléatoire qui compte le nombre de personnes ayant voté pour le candidat A .

2. Ce genre de courbe est habituellement désigné comme étant une « gaussienne » en l'honneur de Gauss et de sa loi normale.

1. Déterminer la loi de X et préciser la valeurs des paramètres associés.
2. Déterminer des réels k_1 et k_2 tels que : $\mathbb{P}(k_1 \leq X \leq k_2) \geq 0,95$.³
3. Donner une interprétation du résultat précédent.

La conclusion attendue est que pour un échantillon de 50 personnes, il y a au moins 95% de chance qu'il y ait entre 42% et 68% personnes ayant voté pour le candidat A . Nous constatons que l'échantillon n'est pas suffisamment grand pour prédire la victoire du candidat A (puisque nous savons déjà que 55% des électeurs ont voté pour lui).

Il est possible, si le temps le permet, de basculer rapidement dans le monde des statistiques et voir de quelle manière la loi binomiale peut-être employée. Ceci est un moyen d'explorer plus avant ces histoires d'intervalles de fluctuations et de seuil.

Exercice Bilan : 15 page 392 ; 17,19 page 379 ; autres exercices 73,76 page 386 et 101/103 page 390 ; sujets type bac .

<https://www.youtube.com/watch?v=8f-cfVFHixg&list=PLVUDmbpupCapZdNo8QaVDkWz3eqneIYK1&index=29>.

14.5 Statistiques (pour aller plus loin)

De manière générale, nous faisons face à la situation suivante : au sein d'une population, nous supposons que la proportion d'un certain caractère vaut $p \in [0; 1]$. Nous sommes souvent amené à faire des suppositions sur cette proportion. Voyons cela sur un exemple.

Exemple 14.5.1. L'étude de l'équilibre d'une pièce

$$i.e. \quad H_0 : p = \frac{1}{2} \quad \text{contre} \quad H_1 : p \neq \frac{1}{2} \quad (14.5.1)$$

rentre dans le cadre décrit au début de la section.

A partir d'un échantillon de taille n dans lequel le caractère en question est présent avec une fréquence valant f . Qu'est-il possible de décider ? Au vu des observations, devons-nous accepter l'hypothèse H_0 portant sur p ou la rejeter ? Dans tous les cas, quelle est l'erreur commise ?

Remarque. Implicitement, nous utilisons une loi binomiale $\mathcal{B}(n, p)$ dont l'épreuve de Bernoulli associée serait

$$S : \text{le caractère étudié apparait} \quad ; \quad \mathbb{P}(S) = p$$

et le nombre de répétitions indépendantes correspond à la taille de l'échantillon .

Afin d'obtenir une règle de décision, permettant de trancher lors d'un test, nous aurons besoin de la définition d'un « intervalle de fluctuation ». Si X est une loi binomiale $\mathcal{B}(n, p)$, nous noterons la moyenne empirique (ou fréquence observée du caractère S étudié) par \bar{X}_n .

Définition 14.5.1 (Intervalle de fluctuation). Si $X \sim \mathcal{B}(n, p)$ avec $n \geq 1$ et $p \in [0, 1]$, un intervalle de fluctuation (avec un seuil d'erreur de 5% sur un échantillon aléatoire de taille n) est donné par :

$$I_n = \left[\frac{k_1}{n}; \frac{k_2}{n} \right] \quad \text{avec} \quad \begin{cases} k_1 : \text{le plus petit entier tel que} & \mathbb{P}(X < k_1) < 0,025 \\ k_2 : \text{le plus petit entier tel que} & \mathbb{P}(X > k_2) < 0,025 \end{cases}$$

Ainsi, il y a 95% de chance que $\bar{X}_n \in I_n$.⁴

Remarque. 1. La deuxième condition peut-être remplacée de manière équivalente par $\mathbb{P}(X \leq k_2) \geq 0,975$.

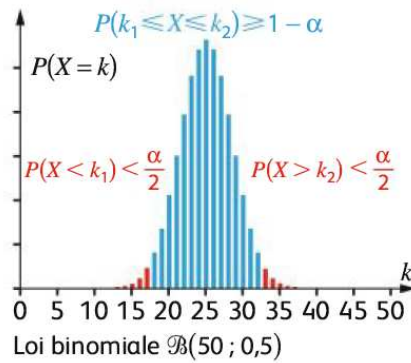
2. Il est possible de remplacer la marge d'erreur de 5% par un nombre $\alpha \in]0; 1[$. Dans ce cas, la définition de k_1 et k_2 est changée en

$$I_n = \left[\frac{k_1}{n}; \frac{k_2}{n} \right] \quad \text{avec} \quad \begin{cases} k_1 : \text{le plus petit entier tel que} & \mathbb{P}(X < k_1) < \frac{\alpha}{2} \\ k_2 : \text{le plus petit entier tel que} & \mathbb{P}(X > k_2) < \frac{\alpha}{2} \end{cases}$$

Graphiquement, les entiers k_1 et k_2 sont choisis de sorte que les aires en rouge soient inférieures à $\frac{\alpha}{2}$:

3. L'intervalle $I_n = \left[\frac{k_1}{n}; \frac{k_2}{n} \right]$ est appelé intervalle de fluctuation avec seuil de 5%. De manière générale, I est un intervalle de fluctuation (de seuil $\alpha \in]0, 1[$ pour une variable aléatoire X si $\mathbb{P}(X \in I) \geq 1 - \alpha$; lorsque $I = [a, b]$ avec $\mathbb{P}(X < a) < \frac{\alpha}{2}$ et $\mathbb{P}(X > b) < \frac{\alpha}{2}$, l'intervalle de fluctuation $[a; b]$ est dit centré.

4. Il suffit d'observer que $\mathbb{P}\left(\frac{X}{n} \in I_n\right) = \mathbb{P}(k_1 \leq X \leq k_2) = 1 - \mathbb{P}(X > k_2) - \mathbb{P}(X < k_1)$ puis de procéder à une minoration.



Nous pouvons alors énoncer la règle de décision suivante : laquelle qui apporte une solution au problème (14.5.1).

Proposition 88 (Règle de décision). *Dans le cadre décrit précédemment, étant donnée la fréquence observée \bar{X}_n , nous avons l'alternative suivante :*

1. si $\bar{X}_n \in I_n$ alors, avec une marge d'erreur de 5%, l'hypothèse H_0 n'est pas rejetée .
2. si $\bar{X}_n \notin I_n$ alors, avec une marge d'erreur de 5%, l'hypothèse H_0 est rejetée .

Remarque. D'un point de vue pratique, les **tests statistiques sont plutôt là pour rejeter des hypothèses plutôt que les accepter**. Cela revient à dire : si je modélise mon aléa pour une certaine loi de probabilité, est-ce que les **fréquences observées sont réalistes ou non** ? Comme mentionné plutôt, la marge d'erreur de 5% peut-être remplacée par $\alpha \in]0; 1[$.

Voyons cela en application sur un exemple.

Exemple 14.5.2. Reprenons l'exemple de la pièce et testons

$$H_0 : p = \frac{1}{2} \quad \text{contre} \quad H_1 : p \neq \frac{1}{2}.$$

Supposons qu'après 50 lancers, nous ayons observé 19 piles. D'où $f = \frac{19}{50}$. A l'aide de la calculatrice nous déterminons la valeur de k_1 et de k_2 (associée à une loi binomiale de paramètre $n = 50$ et $p = \frac{1}{2}$). Nous trouvons

$$k_1 = 18 \quad \text{et} \quad k_2 = 32.$$

Par suite, $I_n = \left[\frac{18}{50}; \frac{32}{50} \right]$. Puisque $f \in I_n$, nous ne rejetons pas l'hypothèse H_0 . Autrement dit, d'après les observations, il n'est pas aberrant de dire que la pièce est équilibrée (mais nous ne savons toujours pas si c'est le cas). En revanche, si le nombre de piles observés était plus grand ou plus petit (par exemple, 10 piles ou 40 piles) nous aurions rejeté l'hypothèse H_0 car la réalisation de cet événement était très peu probable.

Remarque. Notre étude repose uniquement sur l'utilisation des valeurs extrêmes d'une loi binomiale. Il se trouve qu'il est possible d'obtenir des résultats plus performants et plus précis que celui-ci en utilisant d'autres arguments liés à la loi normale et au théorème de la limite centrale.⁵

Il existe d'autres questions que nous n'auront pas le temps d'étudier ici. Par exemple, comment estimer la valeur (inconnue) p à partir d'un échantillon de taille n ? Comment produire un intervalle⁶ contenant cette valeur inconnue avec une marge d'erreur de 5% ? Cependant, à la fin de l'année, nous verrons de quelle manière les inégalités de concentration sont des outils féconds pour aborder ce genre de problématique.

5. Il est possible de montrer que l'intervalle $I_n = \left[p - 1.96 \sqrt{\frac{p(1-p)}{n}}; p + 1.96 \sqrt{\frac{p(1-p)}{n}} \right]$ a 95% de chance de contenir la moyenne empirique \bar{X}_n (à condition que n soit assez grand et p ni trop proche de 0 ou de 1) ; I_n peut être remplacé, après quelques majorations, par l'intervalle $J_n = \left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right]$ qui est plus simple mais moins précis car plus grand que I_n .

6. Il s'agit d'un intervalle de confiance, par opposition à l'intervalle de fluctuation obtenu lorsque la valeur de p est donnée. Cette fois-ci, l'intervalle implique la moyenne empirique (il est donc aléatoire) et contient le paramètre p (inconnu) avec une probabilité supérieure à $1 - \alpha$ pour α un seuil donné.

14.6 Notes historiques

La famille Bernoulli est très importante dans l'histoire des sciences :

- Jacques ou Jakob Bernoulli (1654 – 1705) est un mathématicien et physicien suisse. Il a travaillé dans de nombreux domaines mathématiques. Par exemple, il publie en 1685 une étude dans laquelle il établit des parallèles entre la logique et l'algèbre. Dans la même période, il proposa aussi des résultats en probabilités et en géométrie. Il produit de nombreux traités sur les séries infinies (somme infinie de termes d'une suite donnée), notamment celles liées à des exponentielles apparaissant dans des calculs d'intérêts composés. Il a également résolu des problèmes d'équations différentielles : il s'agit d'équations liant une fonction et ses dérivées et l'objet à déterminer est une fonction (plutôt qu'un nombre x vérifiant une égalité donnée).

Bernoulli a été l'un des promoteurs les plus importants des méthodes formelles de l'analyse mathématique. L'astuce et l'élégance sont rarement présentes dans sa façon de présenter et de rédiger, mais on y trouve un maximum de rigueur.

- Daniel Bernoulli est un médecin, physicien et mathématicien suisse, né à Groningue le 8 février 1700, et mort à Bâle, le 17 mars 1782. Il est le fils de Jean Bernoulli, le neveu de Jacques Bernoulli et le frère de Nicolas Bernoulli et de Jean Bernoulli. Daniel Bernoulli a beaucoup travaillé sur la théorie cinétique des gaz et la mécanique des fluides ainsi que sur des façons de mesurer des risques (financiers)
- Jean Bernoulli, Johann Bernoulli, né le 27 juillet 1667 à Bâle où il est mort le 1er janvier 1748, est un mathématicien et physicien suisse. Il est le frère de Jacques Bernoulli et le père de Daniel, Nicolas et Jean. Il travailla sur des thématiques proches de celles de son frère (en analyse, calcul infinitésimal, mécanique du point, . . .)
- Nicolas II Bernoulli (Groningen, 6 février 1695 - Saint-Petersbourg, 31 juillet 1726) est un mathématicien suisse. Nicolas et ses frères Daniel et Jean avaient pour père Jean Bernoulli, frère de Jacques Bernoulli. Nicolas a travaillé principalement sur les courbes, les équations différentielles, et les probabilités. Il a également contribué à la dynamique des fluides. Il était un ami et contemporain de Leonhard Euler, qui étudia avec son père Jean Bernoulli. Il a été invité par Pierre le Grand à l'Académie des sciences de Russie nouvellement fondée. Huit mois après sa nomination, il est tombé malade avec de la fièvre et est mort. Son poste a été occupé en 1727 par Euler, qui avait été recommandé par les frères Bernoulli. Sa mort prématurée a mis fin à une carrière prometteuse.

Récemment, en 2012, une conjecture difficile, énoncée par Talagrand, portant sur les processus de Bernoulli et l'estimation de la taille de leur supremum a été résolue par Latala et Bednorz au prix d'efforts techniques considérables.