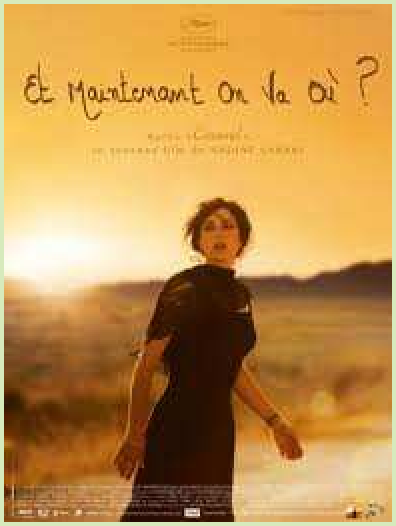
Chapitre 9

Loi Binomiale et Echantillonnage



Hors Sujet

Titre : « Et maintenant on va où ? »
Auteur : Nadine Labaki
Présentation succinte de l'auteur : Nominé pour plusieurs prix « *Et maintenant on va où ?* » , de **Nadine Labaki**, a remporté le prix du Jury Oecuménique – Mention Spéciale à Cannes. Après « *Caramel* » , film déjà primé, la réalisatrice revient sur un thème extrêmement contemporain : la cohabitation religieuse.

Le film s'ouvre sur une scène originale, où, des femmes, drapées de noir, endeuillées, entament une danse esthétiquement précise et émotionnellement forte. Les femmes nous regardent comme si elles assistaient à un spectacle et que donc nous étions le film ce qui renforce l'aspect contemporain du sujet. Il s'agit des femmes d'un village, divisé en deux confessions avec d'un côté les musulmans regroupés derrière leur Cheikh, et de l'autre, les catholiques et leur Prêtre. Lors d'une soirée où ils sont tous regroupés devant la télévision, les habitants regardent le journal télévisé qui leur apprend des affrontements entre catholiques et musulmans. Dès lors, les tensions du passé sont ravivées et les hommes ne jurent plus que par leurs religions laissant les femmes abattues par cette situation, mais cherchant des solutions plus originales les unes que les autres pour garder la paix..

Document réalisé à l'aide de LaTeX
Auteur : C. Aupérin
Site : wicky-math.fr.nf
Lycée : Jules Fil (Carcassonne)

Table des matières

« *Si les gens ne croient pas que les mathématiques sont simples, c'est seulement parce qu'ils ne réalisent pas combien la vie est compliquée !* »

JOHN LOUIS VON NEUMANN (RÉALISTE...)

LEÇON 9

Loi Binomiale et Echantillonnage

Au fil du temps

La théorie des probabilités est jeune par rapport aux autres grandes disciplines mathématiques comme la géométrie, l'arithmétique et l'algèbre : elle est née au XVII^e siècle, dans la correspondance entre Blaise Pascal (dont nous verrons l'importance dans ce chapitre), philosophe, théologien et mathématicien, et Pierre de Fermat, avocat et mathématicien « amateur » comme il se qualifiait lui-même, alors qu'il est l'une des grandes figures mathmétiques modernes (deux théorèmes d'arithmétiques portent son nom, dont l'un ne fut démontré qu'en 1994).

Si les probabilités ont attendu si longtemps avant de voir le jour, c'est peut-être à cause de l'étrangeté philosophique qu'elles véhiculent, à savoir l'idée qu'un événement du monde physique soit pensé non pas dans les termes « il se produit ou il ne se produit pas », mais « il peut se produire avec $x\%$ de chances »...
Si on dit qu'il y a $35,4\%$ de chances qu'il pleuve demain à Carcassonne à 11h, cela signifie-t-il qu'il pleuvra réellement à Carcassonne mais seulement à $35,4\%$ et qu'il fera en même temps soleil à $64,6\%$? Cela semble un pur non sens ... Ou cela sigifie-t-il que la dynamique atmosphérique porte en elle une indétermination physique qui interdit de prévoir « à 100% » ? Ou enfin qu'il n'y a aucune indétermination physique mais que nous, êtres pensants, n'avons pas assez d'informations et de moyens de calculs pour prédire parfaitement son évolution ? En fait, ces trois façons d'interpréter un résultat de probabilité sont valables, chacune dans un certain domaine.
Ainsi, en mécanique quantique, une particule peut se trouver « ici à $35,4\%$ et là-bas à $64,6\%$ » tant qu'on n'a pas mesuré concrètement sa position (et non pas « ici ou là-bas » !). Les physiciens admettent qu'avant une mesure, une particule unique peut être en plusieurs lieux simultanément...
En ce qui concerne les deux autres interprétations, indétermination physique objective ou manque d'informations du physicien, le débat est ouvert depuis 1958, date à laquelle James Maxwell introduit les probabilités en physique. Pour lui, la température d'un gaz est liée aux probabilités de mouvement des milliards de particules qui le composent ... Très choquant pour les physiciens de l'époque : comment un phénomène aussi réel et objectif que la température d'un gaz peut-il dépendre d'un état de connaissance, c'est-à-dire d'une donnée subjective ? Pourtant, cette théorie est l'une des grandes réussites de la physique moderne ...
Ainsi, si la théorie mathématique des probabilités est aujourd'hui bien comprise et acceptée, dès qu'on cherche à en comprendre le sens physique, l'étonnement reprend le dessus et motive des générations de futurs chercheurs.

C'est en XVIII^e siècle que les statistiques endossent quant à elles leur premier grand rôle dans la vie moderne : celui d'un outil de prévision. Le mathématicien Antoine Deparcieu établit dans un livre le « profil » de la mortalité de populations à partir de données statistiques. En se servant des méthodes d'échantillonnage, de calculs de moyennes et d'écart-type, il crée les premières « tables de mortalité » permettant d'évaluer le risque moyen de mort d'un individu en fonction de son profil. Ce risque est alors directement transformé en pécule, rente versée à quelqu'un durant toute sa vie en contrepartie de l'acquisition de son bien à sa mort. Avec Deparcieu, la statistique fait son entrée dans le monde de l'économie.
Mais ce n'est qu'au XIX^e siècle que la statistique finit de prendre la place qu'est la sienne aujourd'hui : celle d'une science mathématique mais aussi humaine, omniprésente dans le débat publique. C'est Adolphe Quételet, astronome belge, qui intègre dans un ouvrage toutes les lois de probabilités développées depuis Pascal et Fermat : mesure des erreurs, loi binomiale que vous allez découvrir ici, etc.
Il tente d'établir des lois statistiques des suicides et des crimes en fonction de paramètres comme l'origine sociale, l'âge, le sexe, le climat, le niveau d'études, le revenu, etc. Mais Quételet se voit reprocher de faire de l'homme un être dont le comportement est prédéterminé par des lois mathématiques...
De plus, avec les statistiques sociales, une question se pose : sont-ce les statistiques qui déterminent nos comportements ou l'inverse ? Par exemple, si le taux de meurtriers dans la population est de 5% par an, cela signifie-t-il qu'il existe une sorte de loi qui nous dépasse et qui « oblige » 5% de personnes à se transformer en meutriers ?
Ce type de questionnement hantera tout le XX^e siècle et conduira à de tragiques dérapages où l'on voudra « neutraliser » dès le berceau tout homme né avec les « paramètres statistiques du crime »...

I) Schéma de Bernoulli

I-1 Loi de Bernoulli

Travail de l'élève **1.** Un vieux professeur d'histoire-géo à moitié aveugle, se décrivant lui-même « entre deux âges et très séduisant », décide de séquestrer des gens pour le fun. (« Lol! » dixit le vieux prof d'histoire-géo à moitié aveugle) Il en a kidnappé cinq : deux filles qu'il n'arrive pas à distinguer et qu'il appelle toutes les deux Lolo, son fils Hugo et ses collègues Dédé et Dudu.

1. Il choisit au hasard l'un d'entre eux et le séquestre.
Déterminer la loi de probabilité de l'expérience réalisée.
Est-on dans un cadre d'équiprobabilité ?

2. Le vieux professeur d'histoire-géo à moitié aveugle considère que son expérience est un Succès lorsqu'il séquestre l'une des Lolos, un Echec sinon.
On appelle X la variable aléatoire égale à 1 en cas de Succès et 0 sinon.

- **a.** Déterminer la loi de probabilité de X.
- **b.** Calculer l'espérance de X et sa variance.

Définition 1 : *Variable aléatoire suivant une loi de Bernoulli*

Une **épreuve de Bernoulli** est une expérience aléatoire ne comportant que 2 issues (Succès et Echec). On note p la probabilité de succès.
Soit X la variable aléatoire qui est égale à 1 en cas de succès et 0 sinon. Alors, on dit que X suit une **loi de Bernoulli de paramètre** p. On note $X \hookrightarrow B(1;p)$.

Exemple :

- Pile ou Face.
- Lancer un dé et regarder si l'on obtient un 5 ou non.
- Aimer ou non Mireille Matthieu.
- Etre ou ne pas être.
- Séquestrer ou non l'une des Lolos.

Propriété 1 : *Espérance et variance d'une variable aléatoire suivant une loi de Bernoulli*

Si $X \hookrightarrow B(1;p)$ alors :

$$E(X) = p \qquad \text{et} \qquad V(X) = p(1-p)$$

Preuve

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

$$V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p)$$

I-2 Répétition d'épreuves de Bernoulli, identiques et indépendantes

Travail de l'élève **2.** On reprend le contexte de l'activité précédente.

1. Le vieux professeur d'histoire-géo à moitié aveugle choisit maintenant au hasard, successivement et sans remise, deux des personnes kidnappés.

a. Calculer la probabilité p de l'événement : « la personne séquestrée choisie en premier est l'une des Lolos »

b. On sait désormais que le vieux professeur d'histoire-géo à moitié aveugle a choisi l'une des Lolos en premier.
Calculer alors la probabilité q de l'événement : « la personne séquestrée choisie en deuxième est l'une des Lolos ».

c. Les probabilités p et q sont-elles égales ? Expliquer ce résultat.

2. Désormais, le vieux professeur d'histoire-géo choisit au hasard, successivement et avec remise deux personnes parmi les six.

a. Reprendre les questions **a.** , **b.** et **c.** de la question **1.**

b. Le vieux professeur d'histoire-géo à moitié aveugle considère toujours que son expérience est un Succès lorsqu'il séquestre l'une des Lolos, un Echec sinon. On appelle X la variable aléatoire qui compte le nombre de Succès.
Déterminer la loi de probabilité de X, puis calculer l'espérance de X et sa variance.

Définition 2 :
On dit que deux expériences aléatoires sont **indépendantes** lorsque les résultats de l'une n'influencent pas les probabilités des issues de l'autre.

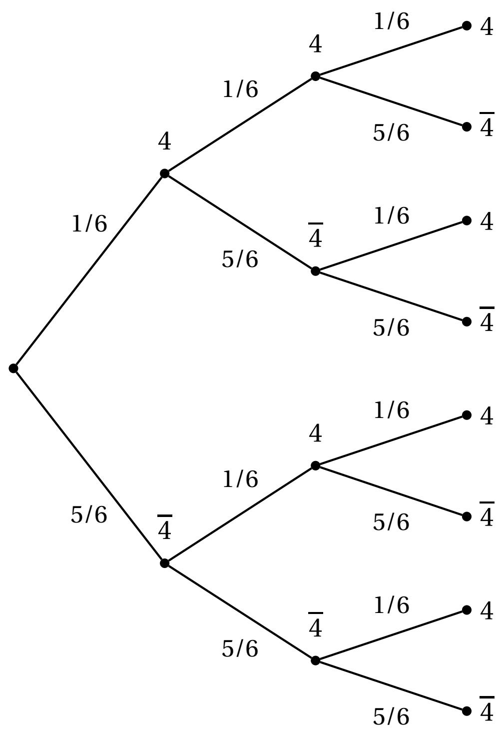
Remarque : Pour résoudre des problèmes étudiant une répétition de n expériences aléatoires indépendantes, on utilisera un arbre, avec les règles classiques, comme nous l'avons déjà fait dans le chapitre de probabilités précédents.

Exemple :
On lance n dés ($n \geq 1$). On note A l'événement « obtenir au moins un 4 (sur l'ensemble des n lancers) ».

1. Décrire l'événement $\overline{A}$ à l'aide d'une phrase.
2. Faire un arbre et calculer $p(A)$ dans le cas où $n = 3$.
3. Dans cette question, on suppose n quelconque. Exprimer $p(A)$ en fonction de n.
4. A l'aide d'un tableau de valeurs, déterminer le nombre de dés qu'il faut lancer pour que la probabilité d'obtenir au moins un quatre soit supérieure à $\dfrac{3}{4}$.

Solution :

1. $\overline{A}$ est l'événement « Ne pas obtenir de 4 ».
2.



Par conséquent

$$P(A) = 1 - P(\overrightarrow{A}) = 1 - \left(\frac{5}{6}\right)^3 = 1 - \frac{125}{216} = \frac{91}{216}$$

3. En raisonnement de la même manière on obtient :

$$P(A) = 1 - P(\overline{A}) = 1 - \left(\frac{5}{6}\right)^n$$

4. On cherche le plus petit entier n tel que $P(A) \geq \frac{3}{4} \Longleftrightarrow 1 - \left(\frac{5}{6}\right)^n \geq \frac{3}{4}$.
   Nous ne savons pas encore résoudre ce type d'inéquation, mais nous l'apprendrons l'an prochain ...
   A la calculatrice, on trouve que pour $n = 7$, $P(A) \simeq 0.72$ et que pour $n = 8$ on a $P(A) \simeq 0.77$.
   Nous devons donc lancer au moins 8 fois le dé pour être sûr à 75% d'obtenir au moins un 4.

Définition 3 : *Schéma de Bernoulli et loi binomiale*

Soit $n \in \mathbb{N}^*$. La répétition de n épreuves de Bernoulli, identiques et indépendantes, de paramètre p, s'appelle un **schéma de Bernoulli de paramètres n et p.**

Soit X la variable aléatoire comptant le nombre de succès (X est à valeurs dans $\{0; 1; \ldots; n\}$). Alors, on dit que X suit une **loi binomiale de paramètres n et p**. On note $X \hookrightarrow B(n; p)$.

Exemple :

Reprenons la situation de l'exemple précédent dans le cas de 3 dés. Notons X la variable aléatoire qui compte le nombre de 4 obtenus. Remarquons que X est à valeurs dans $\{0;1;2;3\}$.
Calculons la probabilité d'obtenir exactement deux 4, ie $P(X=2)$. D'après les règles sur les arbres, on a :

$$\begin{array}{rcl} P(X=2) & = & P(4\cap\overline{4}\cap 4) \quad + \quad P(\overline{4}\cap 4\cap 4) \quad + \quad P(4\cap 4\cap\overline{4}) \\ & = & \frac{1}{6}\times\frac{5}{6}\times\frac{1}{6} \quad + \quad \frac{5}{6}\times\frac{1}{6}\times\frac{1}{6} \quad + \quad \frac{1}{6}\times\frac{1}{6}\times\frac{5}{6} \\ & = & 3\times\left(\frac{1}{6}\right)^2\times\frac{5}{6} \quad = \quad \frac{5}{72} \end{array}$$

Dans la partie suivante, nous allons généraliser cette méthode.

II) Loi Binomiale

II-1 Coefficients binômiaux

Travail de l'élève **3.** Le vieux professeur d'histoire-géo à moitié aveugle est désormais complètement aveugle et il a décidé de piquer ses otages au cure-dent, juste pour le fun. (« Lol ! » dixit le vieux prof d'histoire-géo complètement aveugle).
L'expérience consiste à choisir au hasard l'un des cinq otages, puis tenter de le piquer au cure-dent. La probabilité qu'il touche sa cible est de $\frac{2}{3}$ quand il s'agit d'un garçon et de $\frac{1}{2}$ quand il s'agit d'une fille (les jumelles étant plus menues que Dédé et plus petites que Hugo et Dudu, il est plus difficile de les atteindre).

1. Quelle est la probabilité qu'il pique Hugo avec un cure-dent ?
2. Le vieux professeur d'histoire-géo complètement aveugle répète n fois cette expérience, de manière indépendante. Il considère cette fois qu'il s'agit d'un Succès lorsqu'il pique son fils Hugo avec un cure-dent. On appelle X_n la variable aléatoire qui compte le nombre de succès S.
   - **a.** Pour $n=2$, construire un arbre décrivant la situation. Combien de chemins dans cette arbre, mènent à l'événement $(X_2=0)$? Même question pour les événements $(X_2=1)$ et $(X_2=2)$.
     Déterminer alors la loi de probabilité de X_2.
   - **b.** Pour $n=3$, construire un arbre décrivant la situation. Combien de chemins dans cette arbre, mènent à l'événement $(X_3=0)$? Même question pour les événements $(X_3=1)$, $(X_3=2)$ et $(X_3=3)$.
     Déterminer alors la loi de probabilité de X_3.
   - **c.** Pour $n=4$, construire un arbre décrivant la situation. Combien de chemins dans cette arbre, mènent à l'événement $(X_4=0)$? Même question pour les événements $(X_4=1)$, $(X_4=2)$, $(X_4=3)$ et $(X_4=4)$.
     Déterminer alors la loi de probabilité de X_4.
   - **d.** Pour un entier naturel n et pour tout entier k tel que $0\leq k\leq n$, on note $\begin{pmatrix} n \\ k \end{pmatrix}$ (« k parmi n ») le nombre de chemins conduisant à l'événement $(X_n=k)$ dans l'arbre correspondant à n expériences.
     Exprimer la probabilité $P(X_n=k)$ en utilisant la notation $\begin{pmatrix} n \\ k \end{pmatrix}$.
3. **a.** Grâce aux résultats des questions 2.a. à 2.c., compléter les lignes correspondantes du tableau ci-contre, donnant les valeurs de $\begin{pmatrix} n \\ k \end{pmatrix}$ en fonction de n et de k.

n \ k	0	1	2	3	4	5	6
0							
1							
2							
3							
4							
5							
6							

b. A l'aide de la définition de $\binom{n}{k}$ et de votre tête, compléter les lignes $n = 0$ et $n = 1$.

c. A l'aide de la calculatrice, compléter les lignes $n = 5$ et $n = 6$.

d. On a fait apparaître en couleur trois séries de 3 cellules.
Trouver une relation simple entre ces 3 cellules.

e. Conjecturer une relation entre $\binom{n}{k}$, $\binom{n}{k+1}$ et $\binom{n+1}{k+1}$

Définition 4 :
On considère un schéma de Bernoulli de paramètres n et p représenté par un arbre, et k un entier naturel tel que $0 \leq k \leq n$.
On appelle coefficient binomial et on note $\binom{n}{k}$ le nombre de chemins de l'arbre réalisant k succès parmi les n épreuves de Bernoulli.

Exemples :

Il est clair que : $\binom{n}{0} = \binom{n}{n} = 1$ $\binom{n}{1} = \binom{n}{n-1} = n.$

A la calculatrice, pour obtenir $\binom{10}{2}$ on écrit :

TI 82 : 10 `Combinaison` 2
Combinaison se trouve en appuyant sur [Math] + [PBR]

TI 89 : `nCr` (10,2) ou en français `nbrComb` (10,2)
nCr se trouve en appuyant sur [Math] + [7]

Casio 35+ : 10 `nCr` 2
nCr se trouve en appuyant sur [OPTN] + [PROB]

$\binom{4}{0} = \ldots$ $\binom{4}{1} = \ldots$ $\binom{4}{2} = \ldots$ $\binom{4}{3} = \ldots$ $\binom{4}{4} = \ldots$

Remarquons la symétrie dans les valeurs des coefficients ...

Théorème 1 :

Soit X une variable aléatoire suivant une loi binomiale de paramètres n et p.
Pour tout $k \in \{0;1;\ldots;n\}$:

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

Preuve

La probabilité d'avoir k succès suivis de $n-k$ échecs est :

$$p^k(1-p)^{n-k}$$

Mais les succès et les échecs n'apparaissent pas nécessairement dans cet ordre.... Cependant, on sait par définition qu'il y a exactement $\binom{n}{k}$ chemins de l'arbre qui contiennent k succès parmi les n épreuves (donc $n-k$ échecs).
On en déduit :

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

Remarques :

- Si on note q la probabilité d'échec alors $P(X=k)=\binom{n}{k}p^kq^{n-k}$
- La probabilité d'avoir n succès est : $P(X=n)=p^n$
- La probabilité de n'avoir aucun succès est : $P(X=0)=q^n$
  Par conséquent, on retrouve des résultats déjà connus, comme la probabilité d'avoir au moins un succès est :

$$P(X\geq 1)=1-P(X=0)=1-q^n$$

Exercices du livre : Repère

n° 76-77-89 p 250

II-2 Propriété des coefficients binomiaux

Propriété 2 :

Symétrie : Pour tous entiers naturels n et k tels que $0\leq k\leq n$ on a

$$\binom{n}{k}=\binom{n}{n-k}$$

Triangle de Pascal : Pour tous entiers naturels n et k tels que $0\leq k\leq n-1$ on a

$$\binom{n}{k}+\binom{n}{k+1}=\binom{n+1}{k+1}$$

Preuve

On considère un schéma de Bernoulli à n épreuves représenté par un arbre, et k un entier naturel tel que $0 \leq k \leq n$.

1. Compter le nombre de chemins de l'arbre contenant n succès reivent par complémentarité à compter le nombre de chemins de l'arbre contenant $n-k$ échecs.
   Par symétrie des rôles des succès et des échecs on a donc :
   $$\binom{n}{k} = \binom{n}{n-k}$$

2. On considère un schéma de Bernoulli à $n+1$ épreuves représenté par un arbre, et k un entier naturel tel que $0 \leq k \leq n-1$.
   $\binom{n+1}{k+1}$ est, par définition, le nombre de chemins réalisant $k+1$ succès lors des $n+1$ épreuves.
   On peut distinguer deux façons d'obtenir ce type de chemin :
   – Ceux finissent par un succès, ie qui contiennent k succès parmi les n premières épreuves : il y en a $\binom{n}{k}$
   – Ceux finissent par un échec, ie qui contiennent $k+1$ succès parmi les n premières épreuves : il y en a $\binom{n}{k+1}$
   On obtient alors que
   $$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$$

n \ k	0	1	2	3	4
0					
1					
2					
3					
4					

k	$k+1$

n	$n+1$

k					
$k+1$					

n					
$n+1$					

Exercices du livre : Repère

n° 66-67-68 p 249

II-3 Espérance et variance de la loi $B(n,p)$

Propriété 3 : ***Espérance et variance d'une variable aléatoire suivant une loi de Bernoulli***

Si X suit une loi binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0;1]$, alors :

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

Exemple :

Calculer l'espérance et la variance de X_4 dans l'activité du vieux professeur d'histoire-géo complètement aveugle.

Exercice 1 **:**

Dans lequel des cas suivants X suit-elle une loi binomiale ? Si oui, donner les paramètres de la loi et calculer l'espérance et la variance de X.

1. Dans une classe, on tire au sort sans remise 5 élèves, X est le nombre d'élèves abonné à Star'Ac mag dans le lot tiré au sort.
2. Dans un sac de 20 billes contenant 7 noires et 13 blanches, on tire avec remise 3 d'entre elles, X étant le nombre de billes noires obtenues.
3. On lance 10 dés, X est le nombre de 5 obtenus.
4. Un circuit comprend 2 lampes en série. Pour chacune d'elle, la probabilité qu'elle fonctionne est de 0.03. X est le nombres de lampes qui s'allument lorsqu'on appuie sur l'interrupteur.
   Même question avec cette fois des lampes en parallèles.

Exercices du livre : Repère

n° **48-50-52**-54-55-56-**58** p 247

III) Echantillonnage

Cette partie vous introduit dans le monde des *Statistiques inférentielles* (**Inférer** : tirer une conséquence de quelque proposition, de quelque fait, etc.) auquel vous êtes en fait constamment confronté(e) en tant que citoyen(ne) d'une société hautement médiatisée. La statistique inférentielle est en effet la science qui permet de « *modéliser une partie observable du réel comme résultant d'un phénoméne aléatoire pour lequel on envisage non pas une mais toute une famille de lois de probabilités possibles* » (J.P. Raoult - dossier APMEP- Déc 2005)

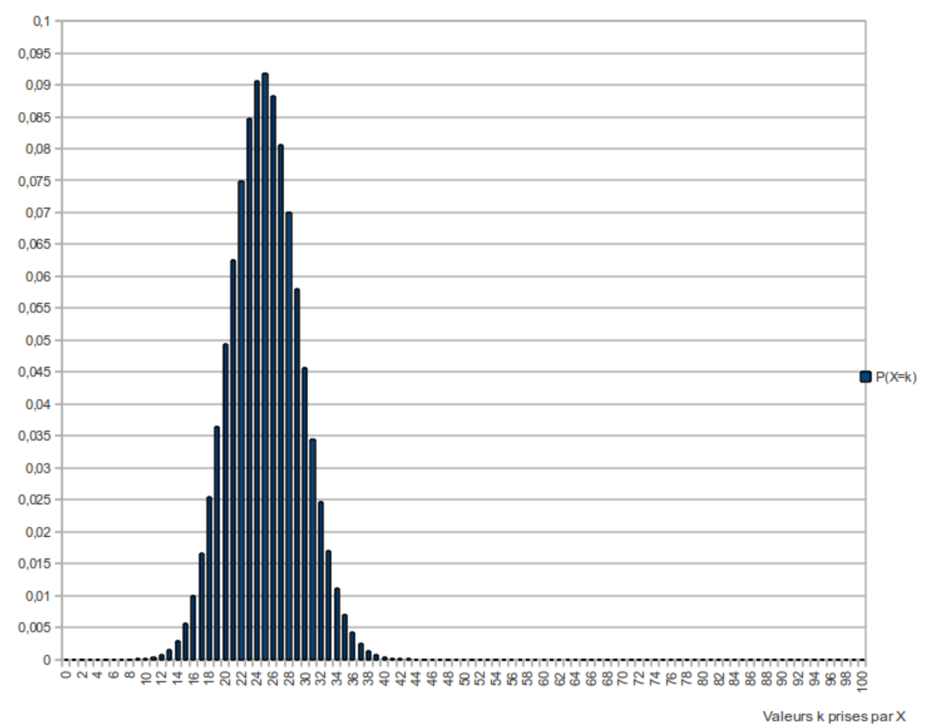
C'est un sujet très intéressant, exigeant une réflexion approfondie sur les notions abordées, les conclusions à émettre, entraînant un débat intéressant, permettant d'avoir une démarche scientifique partant d'une expérimentation. Malheureusement, nous sommes loin de disposer du temps nécessaire. Nous nous contenterons donc d'une préparation pure et simple

aux exercices qui sont tous identiques, mis à part le contexte expérimental.
Dans tous les cas, il s'agira d'étudier les résultats d'une série de n expériences ayant deux issues possibles (Succès ou Echec), dans laquelle on pense pouvoir affirmer que $P(\text{Succès}) = p$.
La question qui se pose alors est : **Comment décider que cette affirmation est « raisonnable » ou pas ? ?**
Mais regardons en détail un exemple.

Ponpon dispose d'une capsule de coca et il pense que la probabilité d'obtenir la face bombée visible est $p = \dfrac{1}{4}$. Il souhaite savoir s'il a raison. Pour cela, il la lance 100 fois (à la main, c'est fastidieux, mais Ponpon est curieux et courageux !). Il a obtenu 68 fois la face creuse visible et 32 fois la face bombée.
Ponpon a alors envie de penser qu'il s'est trompé et que $p = \dfrac{1}{3}$. Mais il continue sa quête de la vérité et ne s'arrête pas là.
Il pense à utiliser une variable aléatoire X suivant la loi $B\left(100, \dfrac{1}{4}\right)$, qui compte le nombre de fois où l'on obtient le côté bombé de la capsule avec son hypothèse $p = \dfrac{1}{4}$. Avec un tableur, il calcule chacune des probabilités $P(X = k)$, pour $0 \leq k \leq 100$, et les représente sous forme de diagramme.
Et oui ! Ponpon est vraiment doué ! Vous, avec vos connaissances de l'an dernier, vous auriez plutôt simulé l'expérience aléatoire de 100 lancers, et observé la fluctuation d'échantillonnage ...
Bref, voici sa feuille de calcul *(remarquez la formule utilisée)*.

B2 =LOI.BINOMIALE(A2;F1;I1;0)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	k	P(X=k)	P(X<=k)		n	100		p	0,25				
2	0	0,0000000000	0,0000000000003										
3	1	0,0000000000	0,0000000000110										
4	2	0,0000000002	0,0000000001874										
5	3	0,0000000019	0,0000000021082										
6	4	0,0000000155	0,0000000176343										
7	5	0,0000000994	0,0000001170015										
8	6	0,0000005244	0,0000006414395										
9	7	0,0000023475	0,0000029889239										
10	8	0,0000090965	0,0000120854260										
11	9	0,0000309955	0,0000430809145										
12	10	0,0000940196	0,0001371005632										
13	11	0,0002564172	0,0003935177867										
14	12	0,0006339204	0,0010274381449										
15	13	0,0014303844	0,0024578225428										
16	14	0,0029629391	0,0054207616528										
17	15	0,0056625059	0,0110832675075										
18	16	0,0100273541	0,0211106216251										
19	17	0,0165156421	0,0376262637012										
20	18	0,0253851536	0,0630114172626										
21	19	0,0365189928	0,0995304101053										
22	20	0,0493006403	0,1488310504430										
23	21	0,0626039877	0,2114350381734										
24	22	0,0749350762	0,2863701143961										
25	23	0,0847092166	0,3710793309957										
26	24	0,0905918011	0,4616711320814										
27	25	0,0917996918	0,5534708238482										
28	26	0,0882689344	0,6417397582394										
29	27	0,0806407549	0,7223805131153										
30	28	0,0700806560	0,7924611691384										
31	29	0,0579977843	0,8504589534334										
32	30	0,0457538076	0,8962127610439										
33	31	0,0344383498	0,9306511108583										
34	32	0,0247525639	0,9554036747873										
35	33	0,0170017611	0,9724054358699										
36	34	0,0111678235	0,9835732593261										
37	35	0,0070197747	0,9905930340700										
38	36	0,0042248644	0,9948178984992										
39	37	0,0024359579	0,9972538563683										
40	38	0,0013461872	0,9986000436118										
41	39	0,0007133642	0,9993134077920										
42	40	0,0003626268	0,9996760345837										
43	41	0,0001768911	0,9998529257016										
44	42	0,0000828300	0,9999357556695										



Ponpon constate déjà que la probabilité d'obtenir exactement 25 fois le côté bombé est inférieur à 0.1, ce qui lui semble bien peu.

De plus, il constate qu'une certaine probabilité de ces 100 lancers donne 32 fois le côté bombé (pas forcément négligeable au vue de la probabilité ci-dessus).
Cependant, comme Ponpon ne présume pas de ses connaissances statistiques (qu'il est modeste !), il demande à Chacha son avis :

Quels peuvent-être les critères de décisions ? Quelles sont alors les erreurs commises ?

Voici leur discussion :

- Pour examiner l'hypothèse $p = \frac{1}{4}$, tu as bien fait de réaliser l'expérience un certain nombre de fois.
  La fréquence d'apparition du dessus de la capsule est de $f = 0.32$ (tu me suis là ? parce que j'ai réduit compter directement).
- Mais évidemment que je te suis, je ne suis pas aussi nul que Thomas quand même !
- Ok, ok, je te taquinais juste.
  Bref, si les valeurs de f et p sont « trop éloignées », on rejettera l'hypothèse.
- Oui mais ca veut dire quoi « trop éloignées » ?
- J'y viens. Avant je voudrais juste que tu comprennes une chose. En fait, on ne peut jamais être sûr de rien, mais d'un point de vue statistique, on peut avoir une idée de la réponse, avec une certaine marge d'erreur.
- Pffff ! Mais c'est nul ton truc, autant répondre au hasard !
- Ponpon, écoute-moi au lieu de bouder. C'est exactement parce que les gens ne s'intéresse pas à la suite qu'ils ne comprennent pas les sondages et pensent que les statisticiens se plantent les 3/4 du temps. Pourtant c'est faux. Alors écoute, tout ceci éveillera ta conscience.
  Je reprends et je m'explique plus clairement. Notre éventuel rejet ne peut pas être fait à coup sûr, c'est pour cela que l'on se fixe une marge d'erreur à l'avance (jamais donnée dans les sondages ceci dit, alors qu'elle est essentielle ...). On parle aussi de seuil.
  Dans notre cas, on fixera le seuil à 95%, ce qui signifie que la probabilité de rejeter l'hypothèse $p = \frac{1}{4}$ alors qu'elle est vraie, sera inférieure à 5%. Ok ?
- Pffff ! Mais pourquoi pas moins ? On ne peut pas 0%, ça j'ai compris, mais pourquoi pas 1% ?
- Non Ponpon, tu n'as pas bien compris. 5% est la probabilité de rejeter l'hypothèse faite alors qu'elle est vraie. Mais évidemment, il y a aussi des cas où on ne rejettera pas l'hypothèse alors qu'elle est fausse !
  Si tu réduis le seuil à 1%, tu rejetteras encore moins de fois ton hypothèse, vraie ou pas, donc a priori beaucoup trop souvent pour accorder du crédit à ta réponse.
  Soit dit en passant, tu peux aussi faire ton test à 0% d'erreur. L'explication est alors plus flagrante encore : cela veut dire que tu ne rejettes jamais ton hypothèse (ce qui t'évite effectivement de te planter en la rejetant alors qu'elle est vraie). Donc que tu acceptes aussi toutes les fausses. Alors ? Utile comme test ? ?
- Non, tu as raison (comme toujours). Bon alors, ce test à 5% d'erreur, comment on le fait ?
- Je veux pointer une dernière nuance de vocabulaire.
- Pffff ! Moi je voulais une réponse par oui ou non au départ ...
- Et rester inculte ! Non, il faut comprendre pour progresser. Donc, je reviens sur ma dernière phrase. J'ai dit « tu acceptes toutes les fausses ». En fait, il s'agit d'un abus de langage de ma part (et même d'une erreur honteuse, mais c'était pour que tu comprennes bien). On n'accepte jamais l'hypothèse faite. On ne sait jamais là non plus si on a raison ou tort. Je ne sais d'ailleurs pas si l'on peut déterminer la marge d'erreur dans ce sens, mias c'est un autre débat.
  On se contentera donc de ne pas rejeter notre hypothèse, autrement dit, de considérer qu'elle est « raisonnable » (ou pas lorsqu'on la rejette).
- Ok. Bon, on le fait ce test ? ?
- C'est parti ! Voici mon cours !

En classe de seconde, les élèves ont observé que sur un grand nombre d'échantillons de taille n simulés (avec $n \leq 25$), 95% au moins fournissent une fréquence d'apparition f appartenant à l'intervalle $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$ (appelé intervalle de fluctuation des fréquences à 95%), où p est la probabilité d'apparition du Succès (et $0.2 \leq p \leq 0.8$).

Mais en classe de première, on peut désormais utiliser la loi Binomiale (ce à quoi Ponpon a bien sûr pensé), et donner un intervalle de fluctuation encore plus précis.
On appelle X la variable aléatoire qui compte le nombre de Succès. On a alors $X \hookrightarrow B\left(n;p\right)$.

Définition 5 :

L'intervalle de fluctuation à 95% associé à une variable aléatoire X suivant une loi $B(n;p)$ est l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$ où a et b sont les deux entiers naturels définis par :
- a est le plus petit entier tel que $P(X \leq a) > 0.025$
- b est le plus petit entier tel que $P(X \leq b) \leq 0.975$

Remarque : Il est clair que cet intervalle contient bien au moins 95% des fréquences, puisqu'on a rejeté les 5% les plus « obsolètes », à savoir 2.5% à chaque extrême des valeurs possibles prises par X.
Pour calculer a et b, on utilisera souvent un tableur, ou alors les résultats seront donnés dans l'énoncé.

Critère de décision

On veut examiner l'hypothèse $P(\text{Succès}) = p$. Soit f la fréquence d'apparition observée de l'événement Succès dans un échantillon d'expériences répétées de taille n.
On désigne par I l'intervalle de fluctuation des fréquences à 95% associée à la loi binomiale $B(n;p)$.
- Si $f \notin I$, on rejette l'hypothèse avec une probabilité inférieure à 5% de se tromper
- Si $f \in I$, on ne rejette pas l'hypothèse car elle est raisonnable (on ne connait pas le risque de se tromper).

Reprenons l'exemple de Ponpon et de sa capsule.
- Avec le résultat de seconde, il obtient l'intervalle de fluctuation

$$I = \left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right] = \left[\frac{1}{4} - \frac{1}{\sqrt{100}}; \frac{1}{4} + \frac{1}{\sqrt{100}}\right] = [0.15; 0.35]$$

- Sur la copie d'écran ci-dessus, on peut voir que $a = 17$ et $b = 34$. Donc l'intervalle de fluctuation associé est

$$J = \left[\frac{a}{n}; \frac{b}{n}\right] = \left[\frac{17}{100}; \frac{34}{100}\right] = [0.17; 0.34]$$

  Ce qui signifie que si on a bien $p = \frac{1}{4}$, alors dans au moins 95% des expériences à 100 lancers, on obtient entre 17 et 34 fois le côté bombé de la capsule.
- **Comparaison :** Les deux démarches produisent des intervalles très voisins, mais on remarque que $J \in I$, il est donc plus précis.
  De plus, comme J contient au moins 95% des fréquences, ceci est aussi vrai pour I, ce qui conforte le résultat vu en seconde qui était sorti de nul part.
  Enfin, on peut remarquer que I est toujours centré en p, mais ce n'est pas forcément le cas de J, comme sur cet exemple.
- **Conclusion :** Ponpon avait obtenu 32 fois le côté bombé, donc son hypothèse $p = 0.25$ est raisonnable.

Remarque : Pour sa seconde hypothèse $p = \frac{1}{3}$, on obtient grâce à la loi $B\left(100, \frac{1}{3}\right)$ l'intervalle de fluctuation :

$$J = \left[\frac{a}{n}; \frac{b}{n}\right] = \left[\frac{24}{100}; \frac{43}{100}\right] = [0.24; 0.43]$$

Donc son hypothèse est encore raisonnable.

Exercices du livre : Repère

n° 5 p 238 + 97 p 255

III-1 TPs Tableur (à adapter)

Problème 1

Dans la ville imaginaire de Sir (issue d'un roman de Diane Meur, « les villes de la plaine ») on a observé que les familles avaient toutes un, deux ou trois enfants. On souhaite savoir quelle la probabilité qu'une famille croisée au hasard ait un, deux ou trois enfants dans cette population.

On a retrouvé l'algorithme d'un ingénieur siriote, qui connaissait la réponse à la question précédente. Par chance on parvient à faire fonctionner son algorithme sur un ordinateur, par contre on a réussi à le décrypter qu'en partie.

Voici l'algorithme :

```
VARIABLES
  compteur1 EST_DU_TYPE NOMBRE
  compteur2 EST_DU_TYPE NOMBRE
  compteur3 EST_DU_TYPE NOMBRE
  n EST_DU_TYPE NOMBRE
  i EST_DU_TYPE NOMBRE
DEBUT_ALGORITHME
  POUR i ALLANT_DE 1 A 1000
    DEBUT_POUR
    n PREND_LA_VALEUR un nombre entier au hasard entre ? et ?
    SI (impossible à décrypter) ALORS
      DEBUT_SI
      compteur1 PREND_LA_VALEUR compteur1+1
      FIN_SI
    SI (impossible à décrypter) ALORS
      DEBUT_SI
      compteur2 PREND_LA_VALEUR compteur2+1
      FIN_SI
    SI (impossible à décrypter) ALORS
      DEBUT_SI
      compteur3 PREND_LA_VALEUR compteur3+1
      FIN_SI
    FIN_POUR
  AFFICHER "Si on croise 1000 familles alors :"
  AFFICHER compteur1
  AFFICHER " familles ont exactement 1 enfant"
  AFFICHER compteur2
  AFFICHER " familles ont exactement 2 enfants"
  AFFICHER compteur3
  AFFICHER " familles ont exactement 3 enfants"
FIN_ALGORITHME
```

1. Formuler une hypothèse (en observant les résultats obtenus en exécutant ce programme) sur les probabilités que les familles de Sir aient un, deux ou trois enfants.
   Pour cela, compléter le tableau suivant :

Enfants par famille	1	2	3	Total
Probabilités théoriques (hypothèse)				

2. Le résultat d'une simulation est le suivant : Si on croise 1000 familles alors :
   450 familles ont exactement 1 enfant
   325 familles ont exactement 2 enfants
   225 familles ont exactement 3 enfants.
   Pensez-vous que les écarts entre les fréquences observées dans cette simulation et les probabilités théoriques (i.e vos hypothèses) s'expliquent par la seule fluctuation d'échantillonage due au hasard ? Êtes-vous sûr d'avoir raison ?
3. Ecrire un programme sur algobox similaire à celui qui est donné avec vos hypothèses.
4. Observer les résultats de vos simulations avec les résultats de simulations effectuées avec l'ordinateur du professeur. Commenter.
5. Le professeur révèle le programme original, ou pas.