8 Probabilités

Objectifs :

- Somme de variables aléatoires, linéarité de l'espérance
- Linéarité de la variance dans une succession d'épreuves indépendantes
- Épreuve de Bernoulli, schéma de Bernoulli ; loi de probabilité associée
- Espérance et variance d'une variable aléatoire suivant une loi binomiale
- Représentation graphique d'une loi binomiale $\mathscr{B}(n;p)$ et programmation
- Échantillon, espérance et variance d'une somme de variables indépendantes identiques
- Inégalité de Bienaymé-Tchebychev, inégalité de concentration et loi des grands nombres

Aperçu historique : L'étude des probabilités telles que nous les connaissons est apparue avec la notion de risque au XII^{e} siècle, pour l'établissement de contrats commerciaux, puis au XVI^{e} siècle avec la généralisation des contrats d'assurance maritime. Mais le véritable début de la théorie des probabilités date des travaux de Pierre de Fermat (1601-1665), de Blaise Pascal (1623-1662) et de Christian Huygens (1629-1695) dans la deuxième moitié du XVII^{e} siècle : ils s'intéressaient principalement à la théorie de jeux. Le terme probabilités prend alors son sens actuel, et Jacques Bernoulli (1654-1705)) traite ce sujet comme une théorie mathématique. Professeur à l'université de Bâle à partir de 1682, il rédige *Ars Conjectandi*, « l'art de conjecturer » (ouvrage publié à titre posthume par son neveu Nicolas en 1713) où il consolide et développe la théorie des probabilités, énonçant notamment, pour la 1re fois, la loi des grands nombres. Il y expose également sa théorie des permutations et des combinaisons, qui constitue aujourd'hui les fondements de la combinatoire.

Abraham de Moivre (1667-1754), français exilé à Londres, publie en 1718 *The doctrine of chances* (la Théorie du hasard) où il donne, entre autre, une première définition de l'indépendance statistique. Thomas Bayes (1702-1761) publie à titre posthume son *Essay Towards Solving a Problem in the Doctrine of Chances* en 1763 qui contient sa fameuse loi. Au XVIII^{e} siècle, Gabriel Cramer (1704-1752) donne un cours sur la logique probabiliste. Jean Le Rond d'Alembert (1717-1783), encyclopédiste notoire, s'est intéressé aux théories probabilistes de son époque mais c'est plus tard qu'elles acquièrent leur cohérence avec la synthèse magistrale de Pierre-Simon de Laplace (1749-1827), le grand savant de l'époque napoléonienne, qui publie en 1812 sa Théorie analytique des probabilités.

L'inégalité de Bienaymé-Tchebychev est une inégalité de concentration qui permet de montrer qu'une variable aléatoire ne prend que rarement une valeur relativement lointaine de son espérance. Ce résultat ne nécessite la connaissance que de l'espérance et de la variance, s'applique dans des cas variés et conduit à une justification de la loi faible des grands nombres. C'est le mathématicien français Irénée-Jules Bienaymé (1796-1878) qui énonça cette inégalité et le russe Pafnouti Tchebychev (1821-1894) qui la démontra.
Ce n'est donc qu'au XIX^{e} siècle que la théorie moderne des probabilités apparaît ; prolongée au début du XX^{e} siècle par l'axiomatique de Andreï Kolmogorov (1903-1987) qui propose un cadre formel rigoureux pour la théorie des probabilités, permettant enfin que celle-ci devienne une branche des mathématiques à part entière. Kolmogorov publie notamment en 1933 les *Fondements de la théorie des probabilités* où il définit les axiomes de la théorie actuelle des probabilités. La description actuelle, en termes d'univers, s'est donc imposée il y a un peu moins d'un siècle.

1. Variables aléatoires

1.a. Rappels

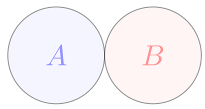
On considère une expérience aléatoire conduisant à n issues, et on note $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ l'ensemble des issues. Ω est « l'univers » des possibles. On définit la probabilité P en donnant la probabilité de chaque issue, c'est-à-dire les nombres $P(\omega_1), P(\omega_2), \ldots, P(\omega_n)$ vérifiant :

- $\forall i \in \{0, 1, \ldots, n\}, 0 \leqslant P(\omega_i) \leqslant 1$
- $\sum_{i=1}^{i=n} P(\omega_i) = P(\omega_1) + P(\omega_2) + \cdots + P(\omega_n) = 1$

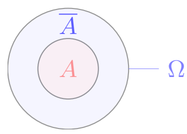
Définition 8.1 Un événement A est une partie de Ω.
La probabilité de A, notée $P(A)$, est la somme des probabilités des issues appartenant à A.

Remarques :

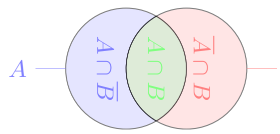
- A est un événement impossible si et seulement si $A = \emptyset$.
  Dans ce cas $P(A) = 0$
- A est un événement certain si et seulement si $A = \Omega$.
  Dans ce cas $P(A) = 1$
- A et B sont incompatibles si $A \cap B$ est impossible, c'est-à-dire si $P(A \cap B) = 0$. Dans ce cas $P(A \cup B) = P(A) + P(B)$.
  L'événement $A \cup B$ est réalisé quand A **ou** B est réalisé. Ce « ou » est, en principe, inclusif mais pour des événements incompatibles, il prend le sens d'un ou exclusif.
- A et B sont des événements contraires si $A \cap B = \emptyset$ et $A \cup B = \Omega$.
  Dans ce cas on note $B = \overline{A}$ et on a $P(\overline{A}) = 1 - P(A)$.
- pour des événements quelconques A et B, on a
  $$P(A \cup B) + P(A \cap B) = P(A) + P(B)$$
  Analysons la situation en termes d'ensembles incompatibles :
  $A = (A \cap B) \cup (A \cap \overline{B})$ et $B = (A \cap B) \cup (\overline{A} \cap B)$,
  d'où $A \cup B = [(A \cap \overline{B}) \cup (A \cap B)] \cup (\overline{A} \cap B)$
  et donc $P(A \cup B) = P(A) + P(\overline{A} \cap B)$,
  mais, comme $P(\overline{A} \cap B) = P(B) - P(A \cap B)$,
  on a finalement $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.



A et B incompatibles



A et $\overline{A}$ contraires



A et B compatibles

Définition 8.2 Lorsque toutes les issues ont la même probabilité – on dit qu'il y a équiprobabilité des issues – s'il y a n issues possibles, la probabilité de chacune est $\frac{1}{n}$.

Remarque :
On reconnait une situation d'équiprobabilité au fait que les issues sont interchangeables. Par exemple, un dé cubique bien équilibré a six faces interchangeables. Que l'une ou l'autre d'entre elles apparaisse lors d'un tirage, cela arrive avec la même probabilité de $\frac{1}{6}$. Quand on lance une pièce équilibrée, on obtient une face ou l'autre avec la même probabilité de $\frac{1}{2}$ car les deux faces sont interchangeables.

Propriété 8.1 Dans le cas d'une situation d'équiprobabilité à n issues possibles, un événement A qui se réalise pour k issues a une probabilité de $P(A) = \frac{k}{n}$. On résume cela en écrivant

$$P(A) = \frac{\text{nombre de cas favorables à A}}{\text{nombre de cas possibles}}$$

La propriété 8.1 suppose que l'on puisse dénombrer les cas favorables et les cas possibles. Le dénombrement n'est pas toujours aisé et manipule parfois de très grands nombres. Le 1er chapitre de ce cours développe les notions utiles pour dénombrer.

DÉFINITION 8.3 (VARIABLE ALÉATOIRE) Si, à chaque issue ω_i de Ω, on peut associer un réel $X(\omega_i)$, on dit que l'on a défini une variable aléatoire X sur Ω. Les probabilités associées aux différentes valeurs de X est appelée « loi » de la variable X.

Remarques :

- Une variable aléatoire est une fonction définie sur Ω ; ce n'est donc pas une variable (ce sont les issues qui sont variables ici) et ce n'est pas non plus aléatoire puisque chaque issue ω_i est associée à la valeur $X(\omega_i)$ de façon bien déterminée (le hasard n'a pas sa place dans cette association).
- Une variable aléatoire permet de définir des évènements. Si l'expérience aléatoire est un jeu dans lequel un joueur gagne ou perd de l'argent, on écrira $X = 5$ l'évènement réalisé lorsque le gain du joueur est 5 (l'unité monétaire est généralement précisée par l'énoncé). L'évènement $X \leqslant 0$ correspond à une perte, tandis que $10 \leqslant X < 20$ correspond à un gain appartenant à l'intervalle $[10; 20[$. etc.
- La notation ensembliste est modifiée puisqu'en notant $X = x$, en réalité, on désigne l'ensemble des issues ω_i telles que $X(\omega_i) = x$, c'est-à-dire $\{\omega_i \in \Omega, X(\omega_i) = x\}$.
- La loi de probabilité de X est une fonction $f : k \mapsto P(X = k)$. Cette fonction agit sur l'ensemble, noté $X(\Omega)$, des valeurs possibles de X. On désigne parfois cette fonction par le terme de « loi de distribution » de X ou encore de « densité » de X.
- La somme de toutes les probabilités $P(X = k)$ quand k décrit l'ensemble $X(\Omega)$ est égal à 1, puisque les évènements $X = k$ sont deux à deux disjoints, leur réunion formant Ω.

DÉFINITION 8.4 (PARAMÈTRES) Soient X une variable aléatoire telle que $X(\Omega) = \{x_1, x_2, \ldots, x_n\}$. En notant $p_i = P(X = x_i)$ les différentes valeurs de la loi de probabilité de X :

- L'espérance de X est le réel, noté $E(X)$, défini par $E(X) = \sum\limits_{i=1}^{n} p_i x_i$
- La variance de X est le réel positif, noté $V(X)$, défini par $V(X) = \sum\limits_{i=1}^{n} p_i \Big(x_i - E(X)\Big)^2$
- L'écart-type de X est le réel positif, noté $\sigma(X)$, défini par $\sigma(X) = \sqrt{V(X)}$

Remarques :

- L'espérance mathématique d'une variable aléatoire est un *paramètre central*, équivalent probabiliste de la notion statistique de moyenne. De même, la variance et l'écart-type sont des mesures de la *dispersion* que l'on calcule aussi en statistiques. Si on réalise l'expérience aléatoire un grand nombre de fois, les paramètres statistiques (fréquences, moyenne) se rapprochent des résultats théoriques (probabilités, espérance).
- Dans un jeu où X est associé aux gains possibles, $E(X)$ représente le gain moyen que l'on peut espérer. Le jeu est dit équitable si $E(X) = 0$, favorable au joueur si $E(X) > 0$ et défavorable au joueur si $E(X) < 0$.

PROPRIÉTÉ 8.2 Soient X une variable aléatoire et a et b deux réels. Les paramètres de la variable aléatoire $X' = aX + b$ sont $E(X') = E(aX + b) = aE(X) + b$ et $V(X') = V(aX + b) = a^2 V(X)$.

DÉMONSTRATION Ces propriétés découlent des propriétés de la multiplication et de l'addition, ainsi que des définitions, notamment $\sum p_i = 1$, $\sum p_i x_i = E(X)$ et $\sum p_i (x_i - E(X))^2 = V(X)$:

$$\begin{aligned} E(aX+b) &= (ax_1+b)\times p_1 + (ax_2+b)\times p_2 + \ldots + (ax_n+b)\times p_n \\ &= (ax_1p_1 + ax_2p_2 + \ldots + ax_np_n) + (bp_1 + bp_2 + \ldots + bp_n) \\ &= a(x_1p_1 + x_2p_2 + \ldots + x_np_n) + b(p_1 + p_2 + \ldots + p_n) \\ &= aE(X) + b \times 1 = aE(X) + b \end{aligned}$$

$$\begin{aligned} V(aX+b) &= \sum_{i=1}^{n} p_i(ax_i+b-E(aX+b))^2 = \sum_{i=1}^{n} p_i(ax_i+b-aE(X)-b)^2 \\ &= \sum_{i=1}^{n} p_i(ax_i-aE(X))^2 = \sum_{i=1}^{n} p_i \times a^2 \times (x_i-E(X))^2 \\ &= a^2 \times \sum_{i=1}^{n} p_i \times (x_i-E(X))^2 = a^2V(X) \end{aligned}$$

Conséquence : On a donc aussi $\sigma(aX+b) = \sqrt{V(aX+b)} = \sqrt{a^2V(X)} = |a|\sqrt{V(X)} = |a|\sigma(X)$

Propriété 8.3 (Autre formule pour la variance) Soit X une variable aléatoire.
La variance de X peut se calculer avec la formule $V(X) = E(X^2) - \Big(E(X)\Big)^2$

Démonstration

$$\begin{aligned} V(X) &= \sum_{i=1}^{n} p_i\Big(x_i - E(X)\Big)^2 = \sum_{i=1}^{n} p_i x_i^2 - 2\sum_{i=1}^{n} p_i x_i E(X) + \sum_{i=1}^{n} p_i \Big(E(X)\Big)^2 \\ &= E(X^2) - 2\sum_{i=1}^{n} p_i x_i E(X) + \sum_{i=1}^{n} p_i E(X) \times E(X) = E(X^2) - \sum_{i=1}^{n} \Big(p_i E(X)\,(2x_i - E(X))\Big) \\ &= E(X^2) - E(X)\left(\sum_{i=1}^{n} 2p_i x_i - \sum_{i=1}^{n} p_i E(X)\right) = E(X^2) - E(X)\left(2E(X) - E(X)\sum_{i=1}^{n} p_i\right) \\ &= E(X^2) - E(X)\,(2E(X) - E(X) \times 1) = E(X^2) - E(X) \times E(X) = E(X^2) - \Big(E(X)\Big)^2 \end{aligned}$$

Remarque : Cette propriété est un cas particulier d'un théorème connu sous le nom de « König-Huygens » qui dit que $E((X-a)^2) = V(X) + (E(X)-a)^2$.
En effet, en prenant $a = 0$, on obtient $E(X^2) = V(X) + \Big(E(X)\Big)^2$.
Pour démontrer ce théorème, on peut partir de l'égalité $x_i - a = x_i - E(X) + E(X) - a$ qui, élevée au carré s'écrit $(x_i-a)^2 = (x_i - E(X))^2 + (E(X)-a)^2 + 2 \times (x_i - E(X))(E(X)-a)$.
Multiplions ensuite cette nouvelle égalité par p_i et sommons les quantités obtenues pour toutes les valeurs de i entre 1 et n. On obtient :

$$E((X-a)^2) = V(X) + (E(X)-a)^2 + 2 \times (E(X)-a) \times \sum_{i=1}^{n} p_i(x_i - E(X))$$

Le dernier terme de cette somme est nul car
$\sum\limits_{i=1}^{n} p_i(x_i - E(X)) = \sum\limits_{i=1}^{n} p_i x_i - E(X)\sum\limits_{i=1}^{n} p_i = E(X) - E(X) = 0$, d'où le résultat.

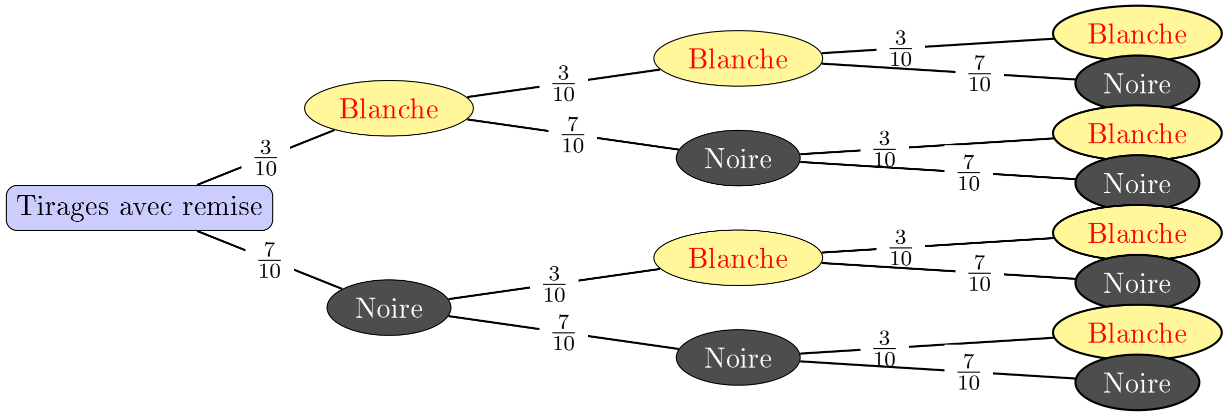
Cette propriété facilite le calcul algorithmique de $V(X)$ puisque l'on n'a pas besoin de connaître « à l'avance » la valeur de $E(X)$: on calcule en parallèle $\sum p_i x_i$ (pour l'obtention de $E(X)$) et $\sum p_i x_i^2$ (pour l'obtention de $E(X^2)$). Supposons que, à l'issue d'un tirage de deux dés, on cherche la variance de la somme X des dés. On peut calculer simultanément :
$E(X) = \sum\limits_{i=1}^{12} p_i x_i = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + \cdots + 12 \times \frac{1}{36} = \frac{252}{36} = 7$
$E(X^2) = \sum\limits_{i=1}^{12} p_i x_i^2 = 2^2 \times \frac{1}{36} + 3^2 \times \frac{2}{36} + 4^2 \times \frac{3}{36} + \cdots + 12^2 \times \frac{1}{36} = \frac{1974}{36}$
Ensuite, on effectue $E(X^2) - \Big(E(X)\Big)^2 = \frac{1974}{36} - \left(\frac{252}{36}\right)^2 = \frac{7560}{36^2} = \frac{35}{6}$.

1.b. Expériences aléatoires successives

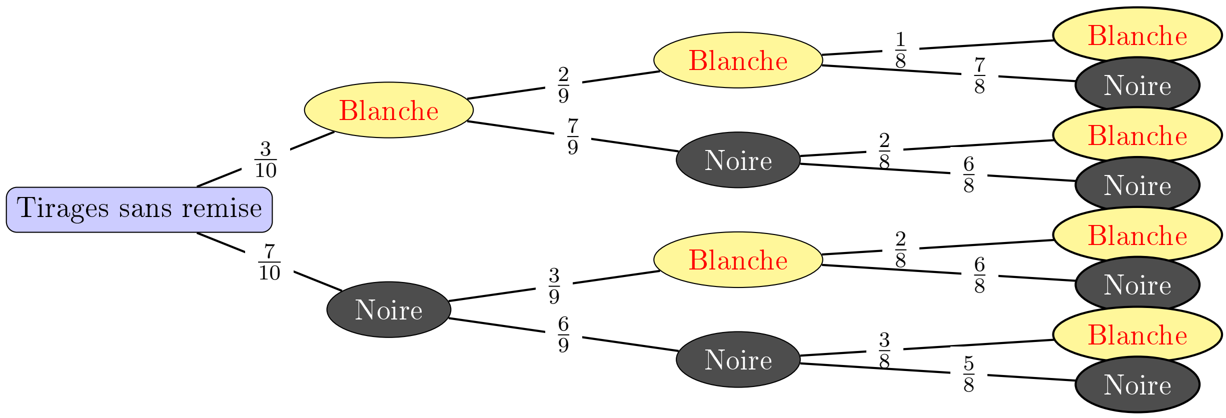
Définition 8.5 Deux expériences sont indépendantes si le résultat de la première n'influence pas le résultat de la seconde, c'est-à-dire qu'il ne modifie pas la loi de probabilité de celle-ci.

Exemples :

- Je lance un dé à six faces (dé cubique) en même temps qu'un dé à quatre faces (dé tétraédrique). Ces deux expériences simultanées sont indépendantes car le résultat du dé cubique n'influence pas le résultat du dé tétraédrique. Même chose si je lance deux dés cubiques : chacun des dés indique un résultat indépendant de l'autre. Le processus expérimental importe peu : je peux lancer les dés ensemble ou séparément, sans changer la loi de probabilité des résultats.
- Si je lance une pièce de monnaie jusqu'à obtenir un « pile », les tirages successifs sont indépendants, la pièce n'ayant pas de mémoire : ce n'est pas parce que j'ai déjà obtenu 5 fois « face » que j'ai plus de chance de faire « pile » au sixième lancé. Les lancés sont indépendants. L'existence du sixième lancé, par contre, dépend du résultat du cinquième puisqu'on ne l'effectue qu'à condition d'avoir obtenu un « face. »
- Dans un vidage d'urne (situation souvent choisie car elle permet de modéliser des problèmes de probabilités très divers), si on remet la boule tirée après avoir noté ce qui nous intéresse (sa couleur, sa taille, son numéro éventuels), les tirages successifs sont indépendants.
  L'urne ne changeant pas de composition, la loi de probabilité reste inchangée d'une étape à l'autre.



Si on ne remet pas la boule qui a été tirée, chacun des tirages influe sur la loi de probabilité de la boule suivante. Les tirages ne sont pas indépendants.



Définition 8.6 (Issues d'une succession d'EAI) Soient n EAI successives (EAI : Expériences Aléatoires Indépendantes) d'univers Ω_1, Ω_2, ..., Ω_n. L'univers résultant de cette succession d'expériences aléatoires est le produit cartésien $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$.
Les issues de Ω sont les n-uplets $(x_1, x_2, \ldots, x_n)$ où $\forall i \in [\![1, n]\!],\ \ x_i \in \Omega_i$.

Remarques : Une succession d'expériences aléatoires est représentée par un graphe étiqueté appelé *arbre de probabilités* : les nœuds (sommets) sont les issues de ces expériences et les étiquettes numériques inscrites sur les branches (arêtes, arcs) sont les probabilités de ces évènements.
Sur un arbre, les issues sont souvent appelées « chemin ». Que les expériences soient dépendantes ou indépendantes les unes des autres, ces arbres vérifient toujours trois règles :

- Règle d'univers : la somme des probabilités des branches partant d'un même nœud vaut 1.
  En partant d'un même nœud, on représente les différentes issues d'une expérience aléatoire. Ces issues sont incompatibles et forment l'univers Ω des possibles de cette expérience. La somme des probabilités vaut donc 1.
- Règle multiplicative : la probabilité d'un chemin est le produit des probabilités inscrites sur ses branches
- Règle additive : la probabilité d'un évènement réalisé par plusieurs chemins est la somme des probabilités de ces chemins

Propriété 8.4 (Règle multiplicative) Dans une succession de n EAI, la probabilité d'une issue $(x_1, \ldots, x_n)$ est égale au produit des probabilités des issues de ses composantes $x_1,\ \ldots,\ x_n$.

Démonstration Considérons la succession de deux épreuves et notons $P(B/A)$ la probabilité de réaliser B à la seconde sachant que A a été réalisé à la première. La probabilité du chemin $A - B$, c'est-à-dire de l'issue (A, B) (le nœud A suivi du nœud B) est celle de l'évènement $A \cap B$. Elle vaut $P(A \cap B) = P(A) \times P(B/A)$, soit le produit des probabilités inscrites sur les branches de ce chemin (revoir les probabilités conditionnelles dans le cours de 1^{re}).
Dans le cas où les expériences successives sont indépendantes, on a $P(A/B) = P(A)$ et par conséquent $P(A \cap B) = P(A) \times P(B)$, la loi multiplicative se vérifie toujours.
La généralisation à n EAI successives se justifie de la même façon.

Propriété 8.5 (Règle additive) Dans une succession de n EAI, la probabilité d'un évènement réalisé par plusieurs chemins est la somme des probabilités de ces chemins.

Démonstration La définition 8.1 de la probabilité d'un évènement implique que si plusieurs chemins réalisent un évènement particulier, alors la probabilité de cet évènement est la somme des probabilités de ces chemins, ceux-ci étant naturellement incompatibles.

Exemple 1 – Une urne contient 3 boules blanches et 7 noires. On tire successivement de cette urne 3 boules, avec remise de la boule tirée. Les tirages sont donc indépendants. Cette succession de trois EAI indépendantes a été représentée par l'arbre « Tirages avec remise » ci-dessus. On peut également raisonner à partir des univers : pour chaque EAI $\Omega_1 = \Omega_2 = \Omega_3 = \{B, N\}$ d'où l'univers pour leur succession :
$\Omega = \Omega_1 \times \Omega_2 \times \Omega_3 = \{(B,B,B),(B,B,N),(B,N,B),(B,N,N),(N,B,B),(N,B,N),(N,N,B),(N,N,N)\}$
Déterminons la loi de la variable aléatoire X comptant le nombre de boules blanches obtenues :
➡ $X = 0$ est l'évènement « obtenir 3 boules noires » qui n'est réalisé que par l'issue (N, N, N).
$P(X = 0) = \left(\frac{7}{10}\right)^3 = \frac{343}{1000} = 0,343$.
➡ $X = 1$ est l'évènement « obtenir 2 boules noires et 1 blanche ».
Cet évènement est l'ensemble $\{(B, N, N), (N, B, N), (N, N, B)\}$.
$P(X = 1) = 3 \times \frac{3}{10} \times \left(\frac{7}{10}\right)^2 = \frac{441}{1000} = 0,441$.
➡ De même, on montre que $P(X = 2) = \frac{189}{1000} = 0,189$ et $P(X = 3) = \frac{27}{1000} = 0,027$.

Vérifions ces résultats et calculons les paramètres de X : $\sum\limits_{k=0}^{3} P(X = k) = \frac{343+441+189+27}{1000} = 1$
$E(X) = \frac{0\times373+1\times441+2\times189+3\times27}{1000} = \frac{900}{1000} = 0,9$ et $E(X^2) = \frac{0\times373+1\times441+4\times189+9\times27}{1000} = \frac{1440}{1000} = 1,44$.
D'où $V(X) = E(X^2) - E(X^2) = 1,44 - 0,9^2 = 0,63$ et $\sigma(X) = \sqrt{0,63} \approx 0,79$.

1.c. Espérance et variance de $X+Y$

Les variables aléatoires envisagées sont finies[1] et définies sur un même univers Ω.
De plus, pour le calcul de $V(X+Y)$, les variables sont indépendantes.

PROPRIÉTÉ 8.6 (ESPÉRANCE DE $X+Y$) Soient X et Y deux variables aléatoires.
L'espérance de la variable $X+Y$ est $E(X+Y)=E(X)+E(Y)$

DÉMONSTRATION On utilisera des notations particulières pour simplifier :

- $X(\Omega)=\{x_i\}_{i\in[\![0,n]\!]}$ et $Y(\Omega)=\{x_j\}_{j\in[\![0,p]\!]}$ sont les ensembles des valeurs prises par X et Y
- les $p_{ij}=P((X=x_i)\cap(Y=y_j))$ sont les probabilités *conjointes* du couple (X,Y)
- les $p_i=\sum\limits_{j=0}^{n}p_{ij}$ et $p_j=\sum\limits_{i=0}^{p}p_{ij}$ sont les probabilités *marginales* pour X et Y
- les couples (i,j) prennent leurs valeurs dans l'ensemble $\Pi=[\![0,n]\!]\times[\![0,p]\!]$

On déduit les espérances des variables X et Y des probabilités conjointes :
$E(X)=\sum\limits_{i=0}^{n}x_ip_i=\sum\limits_{i=0}^{n}x_i(\sum\limits_{j=0}^{p}p_{ij})=\sum\limits_{(i,j)\in\Pi}x_ip_{ij}$ et $E(Y)=\sum\limits_{j=0}^{p}y_jp_j=\sum\limits_{j=0}^{p}y_j(\sum\limits_{i=0}^{n}p_{ij})=\sum\limits_{(i,j)\in\Pi}y_jp_{ij}$
Par conséquent $E(X)+E(Y)=\sum\limits_{(i,j)\in\Pi}x_ip_{ij}+\sum\limits_{(i,j)\in\Pi}y_jp_{ij}=\sum\limits_{(i,j)\in\Pi}(x_i+y_j)p_{ij}$

$Z=X+Y$ est la variable aléatoire qui admet l'espérance $E(Z)=\sum\limits_{z\in Z(\Omega)}zP(Z=z)$.
Comme l'évènement $X+Y=z$ peut s'obtenir de plusieurs façons, regroupons les couples d'indices (i,j) qui conduisent à une même valeur de x_i+y_j dans un ensemble :
$\forall z\in Z(\Omega),\ \ I_z=\{(i,j)\in\Pi,x_i+y_j=z\}$
Ces ensembles I_z formant une *partition* de Π, on en déduit
$E(X)+E(Y)=\sum\limits_{z\in Z(\Omega)}(\sum\limits_{(i,j)\in I_z}(x_i+y_j)p_{ij})=\sum\limits_{z\in Z(\Omega)}(\sum\limits_{(i,j)\in I_z}zp_{ij})=\sum\limits_{z\in Z(\Omega)}z(\sum\limits_{(i,j)\in I_z}p_{ij})$
Mais comme $\sum\limits_{(i,j)\in I_z}p_{ij}=P(X+Y=z)=P(Z=z)$
on en conclut $E(X)+E(Y)=\sum\limits_{z\in Z(\Omega)}zP(Z=z)=E(Z)$

PROPRIÉTÉ 8.7 (VARIANCE DE $X+Y$) Soient X et Y deux variables aléatoires indépendantes.
La variance de la variable $X+Y$ est $V(X+Y)=V(X)+V(Y)$

DÉMONSTRATION Par définition de la variance $V(X+Y)=E\left((X+Y-E(X+Y))^2\right)$
Or $E(X+Y)=E(X)+E(Y)$, donc $V(X+Y)=E\left([(X-E(X))+(Y-E(Y))]^2\right)$
Développons $V(X+Y)=E\left((X-E(X))^2+(Y-E(Y))^2+2(X-E(X))(Y-E(Y)\right)$
On a donc $V(X+Y)=V(X)+V(Y)+E\left(2(X-E(X))(Y-E(Y)\right)=V(X)+V(Y)+2\mathrm{cov}(X,Y)$
La quantité $E\left((X-E(X))(Y-E(Y)\right)$ est appelée covariance de X et Y. On la note $\mathrm{cov}(X,Y)$.
D'après les propriétés de linéarité de l'espérance :
$\mathrm{cov}(X,Y)=E\left(XY-XE(Y)-YE(X)+E(X)E(Y)\right)=E(XY)-E(X)E(Y)$

Montrons que la covariance est nulle lorsque les deux variables sont indépendantes :
$E(XY)=\sum\limits_{(i,j)\in\Pi}x_iy_ip_{ij}$, et l'indépendance se traduisant par $p_{ij}=p_ip_j$, on obtient
$E(XY)=\sum\limits_{(i,j)\in\Pi}x_iy_ip_ip_j=\sum\limits_{(i,j)\in\Pi}(x_ip_i)(y_ip_j)=\sum\limits_{i\in[\![0,n]\!]}x_ip_i\times\sum\limits_{i\in[\![0,p]\!]}y_ip_j=E(X)E(Y)$.
Ainsi $\mathrm{cov}(X,Y)=E(XY)-E(X)E(Y)=0$.
Par conséquent, si X et Y sont indépendantes on a
$V(X+Y)=V(X)+V(Y)+2\mathrm{cov}(X,Y)=V(X)+V(Y)$

1. X est une variable aléatoire finie si le nombre d'issues, c'est-à-dire $\mathrm{card}X(\Omega)$, est fini.

Remarques :

- L'espérance possède les deux propriétés de linéarité : $E(X+Y)=E(X)+E(Y)$ et, pour tout réel a, $E(aX)=aE(X)$ (cela a été montré dans la propriété 8.2). Par conséquent pour tout couple de réels (a,b) on a $E(aX+bY)=aE(X)+bE(Y)$. La variance ne possède que la linéarité additive ($V(X+Y)=V(X)+V(Y)$) et encore, cela n'est vrai que lorsque les deux variables sont indépendantes. Par rapport à la multiplication par un réel a, on a $V(aX)=a^2V(X)$ (cela a été aussi été montré dans la propriété 8.2).
- Si X est une variable aléatoire, la variable $X-E(X)$ est *centrée*, c'est-à-dire $E(X-E(X))=0$ et $V(X-E(X))=V(X)$; la variable $\frac{X-E(X)}{\sigma_X}$ est centrée et *réduite*, c'est-à-dire $E(\frac{X-E(X)}{\sigma_X})=0$ et $V(\frac{X-E(X)}{\sigma_X})=1$.
- La covariance de deux variables indépendantes est nulle mais la réciproque est fausse. Une covariance nulle n'indique pas nécessairement que les variables sont indépendantes (pour qu'elles le soient, on doit avoir $p_{ij}=p_ip_j$ pour tout couple $(i,j)\in\Pi$).

Définition 8.7 (Échantillon) Soit $m\geqslant 2$ un entier.
Le m-uplet $(X_1,X_2,\dots,X_m)$ de m variables aléatoires identiques (définies sur un même univers Ω et ayant une même loi de probabilité) et indépendantes est un *échantillon* de taille m.

Propriété 8.8 (Variables associées à un échantillon) Soit un échantillon $(X_1,\dots,X_m)$ de variables d'espérance $E(X_i)=\bar{x}$ et de variance $V(X_i)=\sigma^2$ pour tout $i\in[\![1,m]\!]$.

- La variable $S_m=X_1+X_2+\cdots+X_m$ a pour caractéristiques : $E(S_m)=m\bar{x}$ et $V(S_m)=n\sigma^2$
- La variable $M_m=\frac{X_1+X_2+\cdots+X_m}{m}$ a pour caractéristiques : $E(M_m)=\bar{x}$ et $V(M_m)=\frac{\sigma^2}{m}$

Démonstration La linéarité additive de E et V implique que :
$E(S_m)=E(X_1+X_2+\cdots+X_m)=E(X_1)+E(X_1)+\cdots+E(X_m)=m\bar{x}$
$V(S_m)=V(X_1+X_2+\cdots+X_m)=V(X_1)+V(X_1)+\cdots+V(X_m)=m\sigma^2$
La linéarité multiplicative de E implique que : $E(M_m)=E(\frac{S_m}{m})=\frac{E(S_m)}{m}=\frac{m\bar{x}}{m}=\bar{x}$
Par contre $V(M_m)=V(\frac{S_m}{m})=\frac{V(S_m)}{m^2}=\frac{m\sigma^2}{m^2}=\frac{\sigma^2}{m}$

2. Loi binomiale

2.a. Épreuve de Bernoulli

Définition 8.8 Une expérience aléatoire admettant deux issues est appelée épreuve de Bernoulli.

- l'une des issues est appelée « succès », notée S ; sa probabilité est p.
- l'autre issue est appelée « échec », notée $\overline{S}$; sa probabilité est $1-p$

On dit que la variable aléatoire X prenant la valeur 1 en cas de succès et la valeur 0 en cas d'échec suit une *loi de Bernoulli* de paramètre p, notée $b(p)$. On appelle X une variable de Bernoulli.

Propriété 8.9 Soit X une variable aléatoire de Bernoulli de paramètre p.
Les paramètres de X sont : $E(X)=p$; $V(X)=p\times(1-p)$; $\sigma(X)=\sqrt{V(X)}=\sqrt{p\times(1-p)}$

Démonstration On a : $E(X)=0\times(1-p)+1\times p=p$.
De même : $V(X)=E(X^2)-(E(X))^2=0^2\times(1-p)+1^2\times p-p^2=p(1-p)$

Exemple : Tirer une boule de l'urne précédente qui contient 3 boules blanches et 7 noires constitue une épreuve de Bernoulli. Si, on convient d'appeler succès le tirage d'une boule blanche, il s'agit d'une épreuve de Bernoulli de paramètre $p = \frac{3}{10}$. La variable aléatoire de Bernoulli X de paramètre p qui prend la valeur 1 en cas de succès et 0 en cas d'échec a une espérance $E(X) = p = \frac{3}{10}$ et une variance $V(X) = p(1-p) = \frac{21}{100}$.
Si, au contraire, on convient d'appeler succès le tirage d'une boule noire, il s'agit d'une épreuve de Bernoulli de paramètre $q = \frac{7}{10}$. La variable aléatoire de Bernoulli X' de paramètre q qui prend la valeur 1 en cas de succès et 0 en cas d'échec a une espérance $E(X') = q = \frac{7}{10}$ et une variance $V(X') = q(1-q) = \frac{21}{100}$.

2.b. Schéma de Bernoulli

Définition 8.9 Soit une épreuve de Bernoulli de paramètre p et soit $n \geqslant 1$ un entier.
La répétition de n fois cette même épreuve de Bernoulli, de façon indépendante, constitue une expérience aléatoire appelée « schéma de Bernoulli » de paramètres n et p.

Remarque : L'exemple précédent, où on tire successivement et avec remise, trois boules d'une urne qui en contient 3 blanches et 7 noires est un schéma de Bernoulli, du fait que les tirages sont indépendants les uns des autres. Si on considère comme « succès » le tirage d'une boule blanche, il s'agit d'un schéma de Bernoulli de paramètres $n = 3$ et $p = \frac{3}{10}$.

Dans un schéma de Bernoulli, si on s'intéresse à l'évènement E_k : « obtenir k succès », celui-ci est réalisé par différents chemins qui ont tous la même probabilité : $p^k \times (1-p)^{n-k}$.
Il suffit de dénombrer ces chemins pour déterminer la probabilité de l'évènement E_k.

Définition 8.10 Dans le cas d'une répétition de n épreuves de Bernoulli identiques et indépendantes, on note $\binom{n}{k}$ ou C_n^k le nombre de chemins aboutissant à k succès exactement ($0 \leqslant k \leqslant n$). L'entier $\binom{n}{k}$ est appelé « coefficient binomial » et lu « k parmi n ».

Dans le cas d'un schéma de Bernoulli à 3 épreuves comme dans l'exemple du tirage de trois boules avec remise, les coefficients sont $\binom{3}{0} = 1$, $\binom{3}{1} = 3$, $\binom{3}{2} = 3$ et $\binom{3}{3} = 1$.
Ces coefficients sont ceux du développement $(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$.
Les propriétés et usages de ces « coefficients binomiaux » font l'objet d'une partie du chapitre 1.

Définition 8.11 (loi binomiale) Lors d'un schéma de Bernoulli de paramètres n et p, la variable aléatoire X comptant le nombre de succès suit une loi de probabilité appelée *loi binomiale* de paramètres n et p, notée $\mathscr{B}(n, p)$.

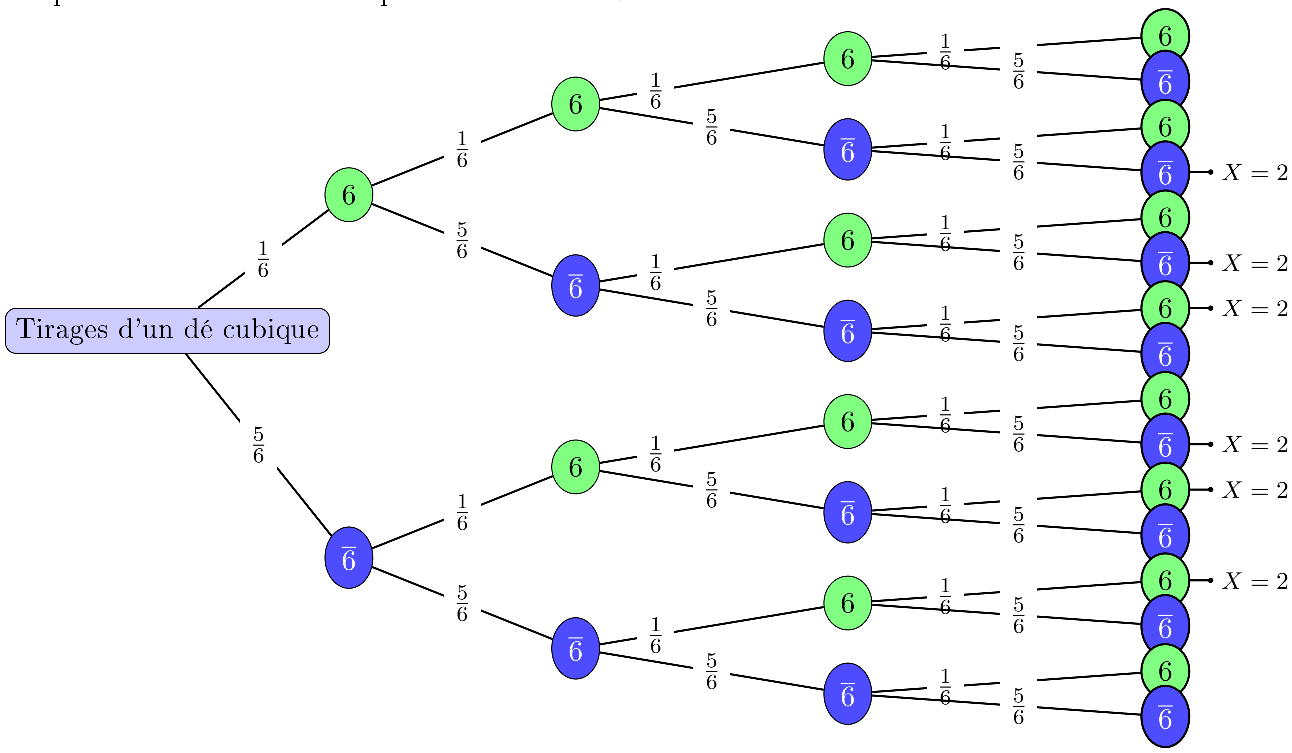
Propriété 8.10 Si X est une variable aléatoire suivant une loi binomiale $\mathscr{B}(n, p)$ alors on a :
- La variable X prend ses valeurs dans $X(\Omega) = \{0, 1, \ldots, n\} = [\![0, n]\!]$
- $\forall k \in \{0, 1, 2, ..., n\}, \quad P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

Démonstration $X(\Omega) = [\![0, n]\!]$ par définition de la variable X.

$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ par définition des coefficients binomiaux qui dénombrent les chemins qui contiennent k succès et qui, du fait de la propriété 8.4, ont la probabilité $p^k (1-p)^{n-k}$ d'advenir.

Exemple 2 – On lance quatre fois de suite un dé équilibré à six faces. On appelle succès l'obtention d'un 6 et on considère la variable aléatoire qui comptabilise le nombre de succès.
Cette loi suit une loi binomiale $\mathscr{B}(4, \frac{1}{6})$ car le dé est équilibré et n'a pas de mémoire.

On peut construire un arbre qui contient $2^4 = 16$ chemins.



1 seul chemin mène à 0 ou 4 succès ; 4 mènent à 1 ou 3 succès. Il en reste donc $16 - (1 + 1 + 4 + 4) = 6$ qui réalisent 2 succès (j'ai marqué ces six chemins sur l'illustration). Les coefficients binomiaux pour un schéma de Bernoulli de paramètre $n = 4$ sont $\binom{4}{0} = 1$, $\binom{4}{1} = 4$, $\binom{4}{2} = 6$, $\binom{4}{3} = 4$ et $\binom{4}{4} = 1$.
La loi de probabilité de la variable X comptabilisant les succès est donnée par la formule suivante :
$P(X = k) = \binom{4}{k} \times \left(\frac{1}{6}\right)^k \times \left(\frac{5}{6}\right)^{4-k}$
Donnons explicitement les résultats pour toutes les valeurs $k \in X(\Omega)$.

k	0	1	2	3	4
$P(X = k)$	$= \left(\frac{5}{6}\right)^4$ $\approx 0,4823$	$= 4 \times \left(\frac{1}{6}\right) \times \left(\frac{5}{6}\right)^3$ $\approx 0,3858$	$= 6 \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^2$ $\approx 0,1157$	$= 4 \times \left(\frac{1}{6}\right)^3 \times \left(\frac{5}{6}\right)$ $\approx 0,0154$	$= \left(\frac{1}{6}\right)^4$ $\approx 0,0008$

Calculons l'espérance et la variance de X :
$E(X) = 0 \times \left(\frac{5}{6}\right)^4 + 1 \times 4\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^3 + 2 \times 6\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2 + 3 \times 4\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right) + 4 \times \left(\frac{1}{6}\right)^4 = \frac{2}{3}$
$V(X) = \left[0 \times \left(\frac{5}{6}\right)^4 + 1 \times 4\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^3 + 4 \times 6\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2 + 9 \times 4\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right) + 16 \times \left(\frac{1}{6}\right)^4\right] - \left[\frac{2}{3}\right]^2 = \frac{5}{9}$

Propriété 8.11 (Paramètres de la loi binomiale)
Si X est une variable aléatoire $\mathscr{B}(n, p)$ alors :
$E(X) = np \quad ; \quad V(X) = np(1-p) \quad ; \quad \sigma(X) = \sqrt{V(X)} = \sqrt{np(1-p)}$

Démonstration Cette propriété découle de la linéarité additive de l'espérance (propriété 8.6) :
Si Y_1 et Y_2 sont deux variables définies sur une même expérience alors $E(Y_1 + Y_2) = E(Y_1) + E(Y_2)$.
Cette propriété s'étend naturellement à n variables de Bernoulli de même loi et d'espérances $E(Y_1) = E(Y_2) = \cdots = E(Y_n) = p$.
La variable $X = Y_1 + Y_2 + \cdots + Y_n$ compte les succès d'un schéma de Bernoulli à n épreuves.
Appliquons à la somme de ces n variables de Bernoulli la linéarité additive :
$E(X) = E(Y_1 + Y_2 + \cdots + Y_n) = E(Y_1) + E(Y_2) + \cdots + E(Y_n) = np$.

La variance a également cette propriété de linéarité additive lorsque les épreuves sont indépendantes.
Pour deux variables indépendantes de même loi on a $V(Y_1 + Y_2) = V(Y_1) + V(Y_2)$.
Cette propriété s'étend aux n variables indépendantes d'un schéma de Bernoulli de variances $V(Y_1) = V(Y_2) = \cdots = V(Y_n) = p(1-p)$.
On en déduit $V(X) = V(Y_1 + Y_2 + \cdots + Y_n) = V(Y_1) + V(Y_2) + \cdots + V(Y_n) = np(1-p)$.

Remarque :
Il est possible de démontrer directement ces résultats, à l'aide d'une propriété des coefficients binomiaux qui constituera notre lemme : $\forall n \in \mathbb{N},\ \ \forall k \in \mathbb{N},\ \ 1 \leqslant k \leqslant n \Longrightarrow k\binom{n}{k} = n\binom{n-1}{k-1}$.
En effet, $\binom{n}{k} = k \times \frac{n!}{k!\times(n-k)!} = \frac{n(n-1)!}{(k-1)!\times(n-1+1-k)!} = \frac{n(n-1)!}{(k-1)!\times((n-1)-(k-1)!} = n\binom{n-1}{k-1}$.

$\boxed{\text{Calcul de } E(X)}$ en utilisant le lemme et en notant, à partir d'ici, q à la place de $1-p$:
$E(X) = \sum\limits_{k=0}^{n} k\binom{n}{k}p^kq^{n-k} = \sum\limits_{k=1}^{n} k\binom{n}{k}p^kq^{n-k} = \sum\limits_{k=1}^{n} n\binom{n-1}{k-1}p^kq^{n-k}$ (terme nul pour $k=0$).
Posons $k' = k-1$ et mettons n puis p en facteur :
$E(X) = n\sum\limits_{k'=0}^{n-1} \binom{n-1}{k'}p^{k'+1}q^{n-(k'+1)} = np\sum\limits_{k'=0}^{n-1} \binom{n-1}{k'}p^{k'}q^{n-(k'+1)} = np\sum\limits_{k'=0}^{n-1} \binom{n-1}{k'}p^{k'}q^{n-1-k'}$.
On remarque alors que $\sum\limits_{k'=0}^{n-1} \binom{n-1}{k'}p^kq^{n-1-k}$ est l'écriture du développement du binôme $(p+q)^{n-1}$ et comme $p+q=1$ cette somme vaut 1 et donc $E(X) = np$.

$\boxed{\text{Calcul de } V(X)}$: Pour pouvoir utiliser deux fois successivement le lemme, il est astucieux d'écrire $E(X^2)$ sous la forme $E(X(X-1)) + E(X)$. Cette transformation d'écriture est possible d'après cette propriété de l'espérance : si φ et ψ sont deux fonctions réelles et X une variable aléatoire alors $E(\varphi(X)+\psi(X)) = E(\varphi(X)) + E(\psi(X))$. En effet, en supposant que $\text{card}(\Omega) = n$, on a

$$E(\varphi(X)+\psi(X)) = E((\varphi+\psi)(X)) = \sum\limits_{k=0}^{n}(\varphi+\psi)(x_k)P(X=x_k) = \underbrace{\sum_{k=0}^{n}\varphi(x_k)P(X=x_k)}_{E(\varphi(X))} + \underbrace{\sum_{k=0}^{n}\psi(x_k)P(X=x_k)}_{E(\psi(X))}$$

En choisissant $\varphi : x \longmapsto x(x-1)$ et $\psi : x \longmapsto x$, on a bien $(\varphi+\psi)(x_k) = x(x-1) + x = x^2$.
Revenons à notre démonstration :
$E(X(X-1)) = \sum\limits_{k=0}^{n} k(k-1)\binom{n}{k}p^kq^{n-k} = \sum\limits_{k=2}^{n} k(k-1)\binom{n}{k}p^kq^{n-k}$ (termes nuls pour $k=0$ et 1)
Appliquons deux fois le lemme : $k(k-1)\binom{n}{k} = n(k-1)\binom{n-1}{k-1} = n(n-1)\binom{n-2}{k-2}$
Par conséquent, en mettant $n(n-1)$ en facteur :
$E(X(X-1)) = n(n-1)\sum\limits_{k=2}^{n} \binom{n-2}{k-2}p^kq^{n-k}$
Posons $k' = k-2$ et mettons p^2 en facteur :
$E(X(X-1)) = n(n-1)p^2\sum\limits_{k'=0}^{n-2} \binom{n-2}{k'}p^{k'}q^{n-(k'+2)} = n(n-1)p^2\sum\limits_{k'=0}^{n-2} \binom{n-2}{k'}p^{k'}q^{(n-2)-k'}$
On remarque alors que $\sum\limits_{k'=0}^{n-2} \binom{n-2}{k'}p^{k'}q^{(n-2)-k'}$ est l'écriture du développement du binôme $(p+q)^{n-2}$ qui vaut 1 et donc $E(X(X-1)) = n(n-1)p^2$.
Comme $V(X) = E(X^2) - (E(X))^2 = E(X(X-1)) + E(X) - (E(X))^2$, on en déduit
$V(X) = n(n-1)p^2 + np - n^2p^2 = np(1-p)$

Utilisation de la calculatrice :

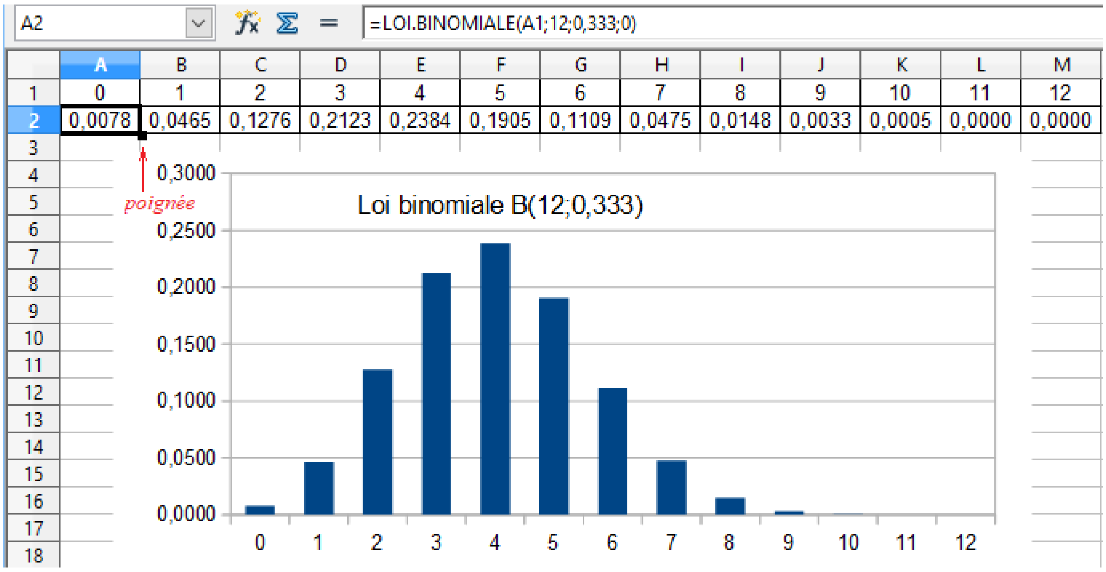
	TI	Casio	Numworks
$\binom{n}{k}$	touche math option PRB choisir 3 (combinaison)	menu RUN option PROB(F3) choisir nCr	menu Calculs, Boîte à outils (5ᵉ touche), Probabilités puis Dénombrement, $\binom{n}{k}$, entrer n et k
$P(X=k)$	touche 2nde puis var dans DISTRIB choisir binomFdp(, compléter avec les valeurs de n, p, k et)	menu STAT options DIST(F5) puis BINM(F5). Compléter l'option Bpd(F1) « probability distribution » avec les valeurs de k, n, p	menu Probabilités, choisir Binomiale, entrer n,p, puis Suivant, sélectionner la 4ᵉ icône
$P(X \leqslant k)$	dans l'onglet DISTRIB, choisir binomFRép(compléter avec les valeurs de n, p, k	Compléter l'option Bcd(F2) « cumulative distribution » avec les valeurs de k, n, p	menu Probabilités, choisir Binomiale, entrer n,p, puis Suivant, sélectionner la 1ʳᵉ icône

La calculatrice permet de déterminer les coefficients binomiaux $\binom{n}{k}$, les valeurs de $P(X=k)$ et aussi $P(X \leqslant k)$ pour une variable aléatoire X qui suit une loi binomiale $\mathscr{B}(n,p)$, avec $n \in \mathbb{N}^*$, $p \in [0;1]$ et k entier tel que $0 \leqslant k \leqslant n$. En se mettant dans le mode « Table », on peut également obtenir un tableau de ces différentes valeurs : on remplace k par X, et on écrit Y1=binomFdp(n,p,X) sur TI ou Bpd(X, n, p) sur Casio. Sur Numworks, le tableau n'est pas disponible, mais les différentes valeurs sont représentées graphiquement par un diagramme en bâton.

Utilisation du tableur : Avec un tableur (*calc* d'OpenOffice), on a accès aux valeurs utiles :
- les coefficients binomiaux $\binom{n}{k}$: taper la formule « =COMBIN(n ;k) »
- les valeurs de $P(X = k)$: taper la formule « =LOI.BINOMIALE(k ;n ;p ;0) »
- les valeurs de $P(X \leqslant k)$: taper la formule « =LOI.BINOMIALE(k ;n ;p ;1) »

Remarques : avec le tableur *Excel* de Microsoft, il suffit de remplacer les codes 0 et 1 par `False` et `True`. En faisant varier k entre 0 et n dans une feuille de calculs, on obtient simplement le tableau des $P(X = k)$ d'une loi binomiale. Ce tableau fourni à l'assistant graphique conduit à la représentation de cette loi par un diagramme en bâtons.

Pour représenter la loi binomiale $\mathscr{B}(12; 0,333)$ à l'aide du tableur : commencer par entrer les valeurs de k sur la ligne 1 (de 0 à 12), puis entrer la formule en A2 : « =LOI.BINOMIALE(A1 ;12 ;0,333 ;0) ». Recopier ensuite cette formule en tirant la poignée vers la droite (voir l'illustration).
Pour obtenir le diagramme, sélectionner la plage des données (de la cellule A1 à la cellule M2, plage notée A1 :M2), puis lancer l'assistant graphique (icône Diagramme) où on sélectionne le type de diagramme « Colonne ». Déclarer les données « en ligne » et la « Première ligne comme étiquette ».



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	0	1	2	3	4	5	6	7	8	9	10	11	12
2	0,0078	0,0465	0,1276	0,2123	0,2384	0,1905	0,1109	0,0475	0,0148	0,0033	0,0005	0,0000	0,0000

2.c. Prise de décision

PROPRIÉTÉ 8.12 (INTERVALLE DE FLUCTUATION) On considère une population sur laquelle on étudie la proportion p d'individus ayant un caractère C avec $p \in [0,2; 0,8]$ et dans laquelle on prélève un échantillon de taille $n \geqslant 25$.
Dans ces conditions, la fréquence f des individus de l'échantillon ayant le caractère C fluctue, dans 95% des cas, à l'intérieur de l'intervalle $\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$, dit *intervalle de fluctuation*.

Remarques :
Cela signifie qu'en choisissant un échantillon de taille n au hasard, la probabilité de l'évènement « $f \in \left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$ » est supérieure ou égale à $0,95$ sous réserve de satisfaire les conditions sur n et p. Reformulons cette situation dans le contexte de la loi binomiale : le tirage au hasard d'un individu dans une population qui peut présenter un caractère C avec une probabilité p est assimilable à une épreuve de Bernoulli de paramètre p où le succès correspond à « avoir le caractère C ». Le prélèvement d'un échantillon de taille n dans cette population s'assimile alors à un schéma de Bernoulli de paramètres n et p et la variable aléatoire X qui compte le nombre de succès suit la loi binomiale $\mathscr{B}(n, p)$. La variable aléatoire $F = \frac{X}{n}$ correspond à la fréquence du succès.

Définition 8.12 Soit X une variable qui suit une loi binomiale $\mathscr{B}(n, p)$ et $F = \frac{X}{n}$ la variable aléatoire qui représente la fréquence du succès. Un intervalle de fluctuation de F au seuil de 95% est un intervalle de la forme $\left[\frac{a}{n}; \frac{b}{n}\right]$ où a et b sont des entiers compris entre 0 et n, tels que

$$P(\frac{a}{n} \leqslant F \leqslant \frac{b}{n}) \geqslant 0,95$$

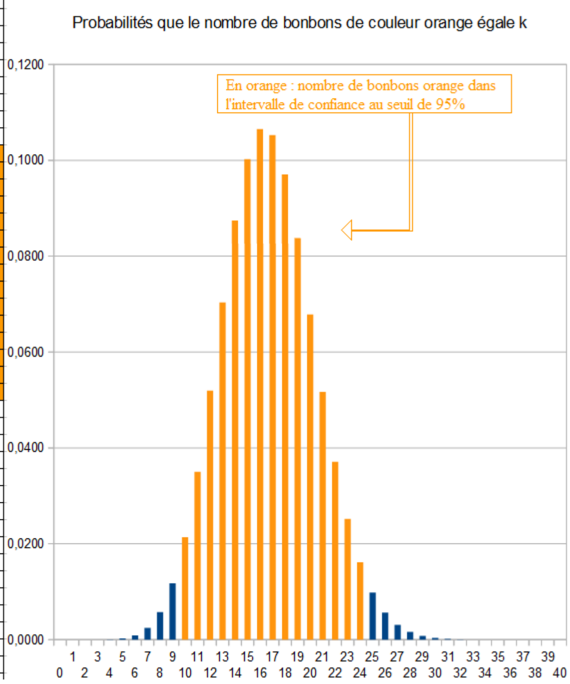
Dans la pratique :
La condition $P(\frac{a}{n} \leqslant F \leqslant \frac{b}{n}) \geqslant 0,95$ équivaut à $P(a \leqslant X \leqslant b) \geqslant 0,95$. Afin d'obtenir l'intervalle de fluctuation de plus petite amplitude, on cherche les plus petits entiers a et b tels que :

$$P\left(X \leqslant a\right) \geqslant 0,025 \text{ et } P\left(X \leqslant b\right) \geqslant 0,975$$

Ces conditions satisfaites, on aura bien $P(a \leqslant X \leqslant b) \geqslant 0,95$ car
$P\left(X \leqslant b\right) \geqslant 0,975 \Longrightarrow P\left(X > b\right) < 0,025$ et $P\left(X \leqslant a\right) \geqslant 0,025 \Longrightarrow P\left(X < a\right) < 0,025$.
Comme $P\left((X < a) \cup (X > b)\right) = P\left(X < a\right) + P\left(X > b\right) < 0,025 + 0,025 = 0,05$,
on a $P(a \leqslant X \leqslant b) = P\left(\overline{(X < a) \cup (X > b)}\right) = 1 - P\left((X < a) \cup (X > b)\right) \geqslant 0,95$.

Exemple 3 – On constitue des sachets de 100 bonbons `M&M's` à partir d'un mélange contenant des bonbons des 6 couleurs en proportion égale ($f = \frac{1}{6}$ pour chaque couleur). Á l'aide du tableur, on veut déterminer un intervalle de fluctuation au seuil de 95% de la fréquence d'une couleur, disons orange, dans un sachet. Le nombre X de bonbons de couleur orange suit la loi binomiale $\mathscr{B}(100; \frac{1}{6})$. La fréquence des bonbons de couleur orange est donné par la variable aléatoire $F = \frac{X}{100}$. On cherche deux entiers a et b tels que $P(\frac{a}{100} \leqslant F \leqslant \frac{b}{100}) \geqslant 0,95$ autrement dit tels que $P(a \leqslant X \leqslant b) \geqslant 0,95$. Ci-dessous une copie d'écran du tableur avec les valeurs prise par la variable X, les probabilités $P(X = k)$ et les probabilités cumulées $P(X \leqslant k)$.

k	P(X=k)	P(X<=k)
0	0,0000	0,0000
1	0,0000	0,0000
2	0,0000	0,0000
3	0,0000	0,0000
4	0,0001	0,0001
5	0,0003	0,0004
6	0,0009	0,0013
7	0,0025	0,0038
8	0,0058	0,0095
9	0,0118	0,0213
10	0,0214	0,0427
11	0,0350	0,0777
12	0,0520	0,1297
13	0,0703	0,2000
14	0,0874	0,2874
15	0,1002	0,3877
16	0,1065	0,4942
17	0,1052	0,5994
18	0,0971	0,6965
19	0,0838	0,7803
20	0,0679	0,8481
21	0,0517	0,8998
22	0,0371	0,9369
23	0,0252	0,9621
24	0,0162	0,9783
25	0,0098	0,9881
26	0,0057	0,9938
27	0,0031	0,9969
28	0,0016	0,9985
29	0,0008	0,9993
30	0,0004	0,9997
31	0,0002	0,9999
32	0,0001	1,0000
33	0,0000	1,0000
34	0,0000	1,0000
35	0,0000	1,0000
36	0,0000	1,0000
37	0,0000	1,0000
38	0,0000	1,0000
39	0,0000	1,0000
40	0,0000	1,0000



La partie colorée en orange montre que :

- $P(X \leqslant 10) \geqslant 0,025$ et $a = 10$ est la 1re valeur de k où $P(X \leqslant k) \geqslant 0,025$ se produit
- $P(X \leqslant 24) \geqslant 0,975$ et $b = 24$ est la 1re valeur de k où $P(X \leqslant k) \geqslant 0,975$ se produit

Donc $P(10 \leqslant X \leqslant 24) \geqslant 0,95$ ou encore $P(\frac{10}{100} \leqslant F \leqslant \frac{24}{100}) \geqslant 0,95$ et donc $\left[\frac{10}{100}; \frac{24}{100}\right]$ est un intervalle de fluctuation au seuil de 95% de la fréquence de la couleur orange dans les sachets de 100 `M&M's`. Le risque de voir la fréquence de la couleur orange sortir de cet intervalle est inférieur à 5%.

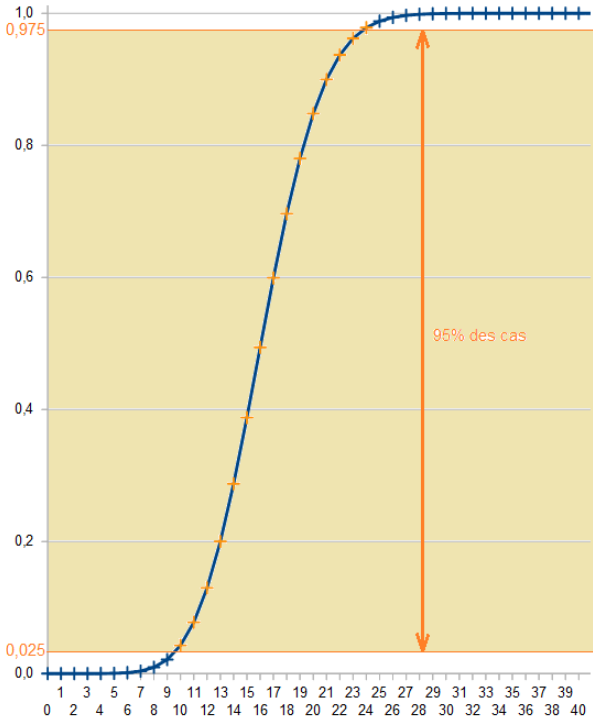
Un échantillon de 100 bonbons contenant moins de 10 bonbons de couleur orange ou plus de 24 bonbons de couleur orange ne serait pas représentatif de cette distribution au seuil de 95%.

Comparons ce résultat à l'intervalle de fluctuation prévu par la propriété 8.12 :
$IF_{prop} = \left[\frac{1}{6} - \frac{1}{\sqrt{100}}; \frac{1}{6} + \frac{1}{\sqrt{100}}\right] = \left[\frac{1}{15}; \frac{4}{15}\right] \approx [0,0667; 0,2667]$. L'intervalle de fluctuation donné par la loi binomiale est $IF_{binom} = [0,1; 0,24]$. On constate que IF_{binom} a une amplitude plus faible que IF_{prop} et que $IF_{binom} \subset IF_{prop}$: les fluctuations prévues par la loi binomiale sont moins dispersées que celles que donne la propriété 8.12.

La représentation cumulative des probabilités associées à la loi binomiale $\mathscr{B}(100; \frac{1}{6})$ est donnée ci-contre. Il s'agit de la représentation de la fonction dont les valeurs sont fournies par le tableau ci-dessus :

$$k \mapsto P(X \leqslant k)$$

Les droites d'équation $y = 0,025$ et $y = 0,975$ correspondant aux valeurs de coupure $2,5\%$ et $97,5\%$, délimitent une zone (colorée sur le graphique) qui contient exactement 95% de la population. Du fait que les nombres a et b sont des entiers, les valeurs de coupure retenues sont légèrement décalées et le pourcentage de la population contenu entre ces valeurs dépasse toujours légèrement 95%. Ici, la somme des probabilités de $P(X = 10)$ à $P(X = 24)$ est environ égale à $95,7\%$.



Règle de décision :
Dans une population, on ne connaît pas toujours avec exactitude la proportion de la population ayant un caractère C mais on peut faire une hypothèse sur cette proportion, c'est-à-dire se fixer une valeur p de cette proportion, et soumettre cette hypothèse au test d'échantillonnage. Notons f la fréquence observée dans un échantillon de taille n de cette population. Comme on sait que 95% des échantillons doivent avoir une fréquence comprise dans l'intervalle de fluctuation IF déterminé à partir de n et p, on peut établir la règle de décision suivante :

> Si la fréquence f de l'échantillon n'appartient pas à IF, on rejette l'hypothèse faite sur p avec un risque d'erreur de 5%. Dans le cas contraire, on affirme que l'hypothèse est crédible car compatible avec les données d'échantillonnage.

Application : cette règle examine si la variabilité d'un échantillonnage peut s'expliquer par le simple fait du hasard. Lorsque la fréquence f mesurée dans l'échantillon sort de l'intervalle de fluctuation, il faut chercher une autre explication que le hasard pour justifier l'écart entre la fréquence observée et la probabilité hypothétique p attribuée au caractère C.
Si on trouve qu'un sachet de 100 bonbons `M&M's` ne contient que 7 bonbons oranges ou 25 bonbons rouges, d'après ce qui a été vu plus haut, il faut rejeter l'hypothèse que les 6 couleurs sont équiprobables dans ce genre de sachet. La simple fluctuation due au hasard ne permet pas d'expliquer ces valeurs. En affirmant cela, on assume un risque d'erreur de 5%.
Avec les mêmes données numériques, pour déterminer si un dé cubique est bien équilibré on peut faire l'expérience de le tirer 100 fois et de noter le nombre de fois où l'on a obtenu chacune des faces. Selon notre étude, si une quelconque des 6 faces est sortie plus de 24 fois ou moins de 10 fois, alors on trouver ce résultat « anormal » au seuil de 95% et rejeter l'hypothèse d'équiprobabilité des faces.

2.d. Simulations

Une expérience aléatoire amène un « succès » avec une probabilité égale à p, nous voulons simuler avec un programme la constitution de N échantillons de taille n pour chacun desquels on déterminera la valeur k prise par la variable aléatoire X qui compte le nombre de succès. Le programme doit pouvoir estimer les probabilités $P(X = k)$ et $P(X \leqslant k)$ ainsi que les paramètres de la loi de X.

Programme P0 : échantillon

Le programme génère un échantillon de n individus dont certains sont des succès (notés S) et les autres des échecs (notés E). On affiche ici l'échantillon ainsi que la valeur prise par la variable aléatoire X comptant le nombre de succès.

```
from random import random

def echantillon(n,p):
    k=0
    for i in range(n):
        if random()<p:
            k+=1
            print("S", end=" ")
        else : print("E", end=" ")
    return k

n,p=16,1/6
print("Génération d'un échantillon de {} individus :".format(n))
print(" -> Nombre de succès={}".format(echantillon(n,p)))
```

```
Génération d'un échantillon de 16 individus :
E E E E E E E E E E S E S E S S  -> Nombre de succès=4
```

Programme P1 : loi de probabilité

Une valeur particulière de $K \in [\![0, n]\!]$) étant donnée, le programme comptabilise :

- le nombre C_1 d'échantillons dans lequel $X = K$
- le nombre C_2 d'échantillons dans lequel $X \leqslant K$

Le programme affiche alors les valeurs empiriques de $P(X = K) = \frac{C_1}{N}$ et $P(X \leqslant K) = \frac{C_2}{N}$.

Programme P2 : paramètres de la loi

Le programme détermine les paramètres (espérance, variance, écart-type) de la variable aléatoire X :

- l'espérance est calculée comme la moyenne des valeurs de k
- la variance est calculée avec la formule de König-Huygens comme la différence entre la moyenne des k^2 et le carré de la moyenne des k

Je reprends les valeurs numériques de l'exemple 3 afin de pouvoir comparer les valeurs empiriques obtenues par la simulation et les valeurs théoriques obtenues par les formules.

P1

```
from random import random

def echantillon(n,p):
    k=0
    for i in range(n):
        if random()<p:k+=1
    return k

def loi(N,n,p,K):
    C1,C2=0,0
    for i in range(N):
        X=echantillon(n,p)
        if X==K:C1+=1
        if X<=K:C2+=1
    return C1/N,C2/N

N,n,p,K=100000,100,1/6,15
f1,f2=loi(N,n,p,K)
print("P(X={})={}".format(K,f1))
print("P(X<={})={}".format(K,f2))
```

```
P(X=5)=0.00021     P(X=10)=0.02224
P(X<=5)=0.00032    P(X<=10)=0.04388
```

P2

```
from random import random
from math import sqrt
def echantillon(n,p):
    k=0
    for i in range(n):
        if random()<p:k+=1
    return k

def parametre(N,n,p):
    X1,X2=0,0
    for i in range(N):
        X=echantillon(n,p)
        X1+=X
        X2+=X**2
    return X1/N,X2/N-(X1/N)**2

N,n,p=100000,100,1/6
E,V=parametre(N,n,p)
E,V,S=round(E,3),round(V,3),round(sqrt(V),3)
print("E(X)={},V(X)={},sigma(X)={}".format(E,V,S))
```

```
E(X)=16.68,V(X)=13.817,sigma(X)=3.717
```

L'intérêt de ce programme est de donner une estimation empirique des caractéristiques de cette loi de probabilité. On a obtenu ici, en simulant 100 000 échantillons, que l'espérance et l'écart-type valent approximativement $E(X) = 16,68$ et $\sigma(X) = 3,717$.

Cette loi de probabilité est une loi Binomiale $\mathscr{B}(100; \frac{1}{6})$) dont les paramètres sont $E(X) = \frac{100}{6} \approx 16,67$ et $\sigma(X) = \sqrt{\frac{500}{36}} \approx 3,727$.

2.e. Autres lois

Il existe de nombreuses autres lois de probabilité en dehors de la loi binomiale. En voici deux :

2.e.1. Loi géométrique tronquée

DÉFINITION 8.13 Soit $n \neq 0$ un entier et p la probabilité de succès d'une épreuve de Bernoulli. On appelle « loi géométrique tronquée » de paramètres n et p, la loi de probabilité d'une variable X donnant le nombre d'essais nécessaires pour obtenir le premier succès, étant entendu que le nombre d'essais ne peut pas dépasser la valeur n.

Remarque : Il n'est pas difficile de déterminer $P(X = k)$ lorsque $1 \leqslant k \leqslant n$. Comme il doit y avoir $k-1$ échecs (de probabilité $1-p$) suivis d'un succès, on a $P(X = k) = p(1-p)^{k-1}$. Lorsqu'il n'y a pas de succès en n essais, la variable aléatoire prend la valeur 0 ; ainsi, on a $P(X = 0) = (1-p)^n$.
Montrons que la somme des probabilités de cette loi vaut bien 1, en notant $q = 1-p$:

$$\sum_{k=1}^{n} pq^{k-1} + q^n = p\sum_{k=1}^{n} q^{k-1} + q^n = p \times 1 \times \frac{1-q^n}{1-q} + q^n = p \times \frac{1-q^n}{p} + q^n = 1 - q^n + q^n = 1$$

Programme P3 : loi et paramètres

Ce programme utilise une liste L afin de donner les valeurs empiriques de $P(X = k)$ pour $0 \leqslant k \leqslant n$. La simulation détermine un rang k du premier succès ce qui, à chaque fois, incrémente de 1 le compteur $L(k)$. Lorsque les N valeurs ont été ainsi enregistrées, il n'y a plus qu'à calculer les fréquences $\frac{L(k)}{N}$ correspondant aux probabilités $P(X = k)$.
Pour le calcul des paramètres, on procède à peu près comme pour la loi binomiale sauf qu'il faut additionner les produits $k \times L(k)$ pour déterminer $E(X)$ et $k^2 \times L(k)$ pour déterminer $E(X^2)$.

```python
from random import random
from math import sqrt

def echantillon(n,p):
    k=0
    for i in range(1,n+1):
        if random()<p:
            k=i
            break
    return k

def loi(N,n,p):
    L=(n+1)*[0]
    for i in range(N):
        L[echantillon(n,p)]+=1
    return L
```

```python
N,n,p=100000,3,1/4
X1,X2=0,0
Liste=loi(N,n,p)
for i in range(n+1):
    P=round(Liste[i]/N,3)
    print("P(X={})={}".format(i,P))
    X1+=Liste[i]*i
    X2+=Liste[i]*i**2
E,V=round(X1/N,3),round(X2/N-(X1/N)**2,3)
S=round(sqrt(V),3)
print("E(X)={},V(X)={},sigma(X)={}".format(E,V,S))
```

```
P(X=0)=0.424
P(X=1)=0.248
P(X=2)=0.188
P(X=3)=0.14
E(X)=1.043,V(X)=1.17,sigma(X)=1.082
```

EXEMPLE 4 – Un client cherche à joindre par téléphone un service de dépannage. La probabilité que son appel soit pris sans attente est $p = \frac{1}{4}$. Si son appel n'est pas pris sans attente, ce client raccroche et rappelle le service. Il fait au maximum trois tentatives ($n = 3$).
On note X la variable aléatoire donnant le rang de la tentative aboutissant sans attente. Si les trois tentatives ont échoué, on convient que $X = 0$.
La loi de probabilité de X est une loi géométrique tronquée de paramètres n et p. D'après les formules établies ci-dessus dans la remarque, les probabilités sont calculées dans le tableau suivant.

k	0	1	2	3
$P(X = k)$	$0,75^3$ $= 0,421875$	$0,25$ $= 0,25$	$0,25 \times 0,75$ $= 0,1875$	$0,25 \times 0,75^2$ $= 0,140625$

Calculons les paramètres de cette loi :
$E(X) = 0 \times 0,421875 + 1 \times 0,25 + 2 \times 0,1875 + 3 \times 0,140625 = 1,046875$
$V(X) = E(X^2) - (E(X))^2 = 1 \times 0,25 + 4 \times 0,1875 + 9 \times 0,140625 - 1,046875^2 \approx 1,169678$
$\sigma(X) = \sqrt{V(X)} \approx \sqrt{1,169678} \approx 1,081516$
Ces valeurs théoriques sont approchées par les valeurs empiriques d'autant mieux que l'on prend un N grand. Avec $N = 100000$, on constate que le résultat d'exécution confirme cela.

2.e.2. Loi hypergéométrique

DÉFINITION 8.14 Soient $M \neq 0$ un entier, m et n deux entiers inférieurs à M.
On appelle « loi hypergéométrique » de paramètres M, n et $\frac{m}{M}$, la loi de probabilité d'une variable X donnant le nombre d'individus possédant un certain caractère, dans un tirage simultané de n individus d'une population de taille M où il y a m individus possédant ce caractère.

Remarque : On peut déterminer $P(X = k)$ lorsque $0 \leqslant k \leqslant n$ en effectuant deux dénombrements, dans le cadre d'une situation d'équiprobabilité :

- le nombre de cas possibles est le nombre de choix de n éléments parmi M, soit $\binom{M}{n}$
- le nombre de cas favorables est le produit du nombre de choix de k éléments parmi m ($\binom{m}{k}$) par le nombre de choix de $n-k$ éléments parmi $M-m$ ($\binom{M-m}{n-k}$)

Ainsi, on a $P(X = k) = \frac{\binom{m}{k} \times \binom{M-m}{n-k}}{\binom{M}{n}}$

Programme P4 : loi et paramètres

Ce programme utilise une liste L afin de donner les valeurs empiriques de $P(X = k)$ pour $0 \leqslant k \leqslant n$. La simulation détermine le nombre d'essais qui ont amenés k succès ce qui, à chaque fois, incrémente de 1 le compteur $L(k)$. Le reste est similaire à la loi géométrique tronquée.

```
from random import random
from math import sqrt

def echantillon(M,n,m):
    k=0
    for i in range(n):
        if random()<m/M:
            k+=1
            m-=1
        M-=1
    return k

def loi(N,M,n,m):
    L=(n+1)*[0]
    for i in range(N):
        L[echantillon(M,n,m)]+=1
    return L
```

```
N,M,n,m=100000,32,5,4
X1,X2=0,0
Liste=loi(N,M,n,m)
for i in range(n+1):
    P=round(Liste[i]/N,4)
    print("P(X={})={}".format(i,P))
    X1+=Liste[i]*i
    X2+=Liste[i]*i**2
E,V=round(X1/N,4),round(X2/N-(X1/N)**2,4)
S=round(sqrt(V),4)
print("E(X)={},V(X)={},sigma(X)={}".format(E,V,S))
```

```
P(X=0)=0.5127
P(X=1)=0.3889
P(X=2)=0.0913
P(X=3)=0.007
P(X=4)=0.0001
P(X=5)=0.0
E(X)=0.5931,V(X)=0.4677,sigma(X)=0.6839
```

EXEMPLE 5 – Au poker menteur, célèbre jeu de cartes à deux joueurs, on distribue 5 cartes à l'un deux qui annonce « brelan de rois » (présence de 3 rois exactement sur les 5 cartes tirées).
Quelle est la probabilité que le joueur ne mente pas s'il n'y a pas de jokers dans le jeu de 32 cartes ?
La variable X (le nombre de rois tirés) suit une loi hypergéométrique de paramètres 32, 5 et $\frac{4}{32}$.
D'après la formule établie ci-dessus, la probabilité cherchée est

$$P(X = 3) = \frac{\binom{4}{3} \times \binom{28}{2}}{\binom{32}{5}} = \frac{27}{3596} \approx 0,0075$$

La valeur empirique calculée par notre programme s'approche bien de cette valeur.
Sur cet exemple, on peut vérifier une propriété de cette loi : son espérance est égale à celle de la loi binomiale, soit $E(X) = np$ égal ici à $\frac{nm}{M} = \frac{5\times4}{32} = \frac{5}{8} = 0,625$.
Quelle est la probabilité que le joueur ne mente pas s'il y a un joker dans le jeu de $32+1$ cartes ?
Dans ce cas, on considère que le joker est un roi. Il s'agit d'une loi hypergéométrique de paramètres 33, 5 et $\frac{5}{33}$ et $P(X = 3) = \frac{\binom{5}{3}\times\binom{28}{2}}{\binom{33}{5}} = \frac{315}{19778} \approx 0,016$, ce que confirme approximativement le programme.

3. Loi des grands nombres

3.a. L'inégalité de Bienaymé-Tchebychev

PROPRIÉTÉ 8.13 (INÉGALITÉ DE MARKOV) Soit X une variable aléatoire finie qui ne prend que des valeurs positives (telle que $X(\Omega) \subset \mathbb{R}^+$) et soit ϵ un réel strictement positif.

$$P(X \geqslant \epsilon) \leqslant \frac{E(X)}{\epsilon}$$

DÉMONSTRATION Par définition de l'espérance $E(X) = \sum\limits_{k \in X(\Omega)} P(X = k) \times k$.
On a $\sum\limits_{k \in X(\Omega)} P(X = k) \times k = \sum\limits_{\substack{k \in X(\Omega) \\ k < \epsilon}} P(X = k) \times k + \sum\limits_{\substack{k \in X(\Omega) \\ k \geqslant \epsilon}} P(X = k) \times k$.
Comme $k \geqslant 0$ et $P(X = k) > 0$, la somme $\sum\limits_{\substack{k \in X(\Omega) \\ k < \epsilon}} P(X = k) \times k$ est positive.
Par conséquent $\sum\limits_{k \in X(\Omega)} P(X = k) \times k \geqslant \sum\limits_{\substack{k \in X(\Omega) \\ k \geqslant \epsilon}} P(X = k) \times k$ et donc $E(X) \geqslant \sum\limits_{\substack{k \in X(\Omega) \\ k \geqslant \epsilon}} P(X = k) \times k$.
Mais $k \geqslant \epsilon \Longrightarrow P(X = k) \times k \geqslant P(X = k) \times \epsilon$.
On en déduit $E(X) \geqslant \sum\limits_{\substack{k \in X(\Omega) \\ k \geqslant \epsilon}} P(X = k) \times \epsilon$, soit $E(X) \geqslant \epsilon \sum\limits_{\substack{k \in X(\Omega) \\ k \geqslant \epsilon}} P(X = k)$.
Par définition $\sum\limits_{\substack{k \in X(\Omega) \\ k \geqslant \epsilon}} P(X = k) = P(X \geqslant k)$.
Finalement $E(X) \geqslant \epsilon P(X \geqslant k)$ d'où, puisque $\epsilon > 0$, la conclusion $P(X \geqslant \epsilon) \leqslant \frac{E(X)}{\epsilon}$

Remarques :

- La double majoration opérée ici est sévère (on néglige toutes les valeurs de $k < \epsilon$ et ensuite on minore tous les $k \geqslant \epsilon$ par ϵ) ce qui peut parfois conduire à un résultat peu intéressant ; néanmoins, on va utiliser ce résultat pour en établir un autre, l'inégalité de Bienaymé-Tchebychev.
- Dans l'exemple 4 où $E(X) = 1,046875$,
  - avec $\epsilon = 2$ on a $P(X \geqslant 2) \leqslant \frac{1,046875}{2} \approx 0,52$ alors que $P(X \geqslant 2) = 0,328125$
  - avec $\epsilon = 3$ on a $P(X \geqslant 3) \leqslant \frac{1,046875}{3} \approx 0,35$ alors que $P(X \geqslant 3) = 0,140625$
  - avec $\epsilon = 4$ on a $P(X \geqslant 4) \leqslant \frac{1,046875}{4} \approx 0,26$ alors que $P(X \geqslant 4) = 0$

PROPRIÉTÉ 8.14 (INÉGALITÉ DE BIENAYMÉ-TCHEBYCHEV) Soit X une variable aléatoire finie et soient ϵ un réel et n un entier, tous deux strictement positifs.

$$P(|X - E(X)| \geqslant \epsilon) \leqslant \frac{V(X)}{\epsilon^2} \text{ et, en particulier, } P(|X - E(X)| \geqslant k\sigma(X)) \leqslant \frac{1}{k^2}$$

DÉMONSTRATION Il suffit d'appliquer l'inégalité de Markov a la variable $(X - E(X))^2$:
$P((X - E(X))^2 \geqslant \epsilon^2) \leqslant \frac{E((X-E(X))^2)}{\epsilon^2}$.
Par définition de la variance $V(X) = E((X - E(X))^2)$.
$P((X - E(X))^2 \geqslant \epsilon^2)$ est la probabilité de l'évènement $(X - E(X))^2 \geqslant \epsilon^2 \Longleftrightarrow |X - E(X)| \geqslant \epsilon$ du fait de la croissance de la fonction racine carrée sur $\mathbb{R}^+$ (et du fait que $|\epsilon| = \epsilon$).
Finalement on a bien $P(|X - E(X)| \geqslant \epsilon) \leqslant \frac{V(X)}{\epsilon^2}$.

Si $\epsilon = k\sigma(X)$ avec $k \in \mathbb{N}^*$, l'inégalité s'écrit $P(|X - E(X)| \geqslant k\sigma(X)) \leqslant \frac{V(X)}{(k\sigma(X))^2}$ soit $P(|X - E(X)| \geqslant k\sigma(X)) \leqslant \frac{1}{k^2}$ puisque $(\sigma(X))^2 = V(X)$.

Remarques :

- La reformulation de l'inégalité de B-T qui prend comme paramètre ϵ un multiple de l'écart-type de la variable donne un majorant très simple ($\frac{1}{k^2}$) de la probabilité que la variable s'écarte de l'espérance de plus de k fois l'écart-type. Cette constatation lui vaut d'être souvent utilisée pour une prise de décision, nous allons l'appliquer dans le cas d'une loi binomiale.
- Dans l'exemple 4 où $E(X) = 1,046875$ et $V(X) \approx 1,169678$,
  - avec $\epsilon = 1$ on a approximativement $P(|X - E(X)| \geqslant 1) \leqslant 1,17$ alors que $|X - E(X)| \geqslant 1 = (X = 0) \cup (X = 3)$
    En effet $|0 - E(X)| \approx 1,05 > 1$, $|1 - E(X)| \approx 0,05 < 1$, ..., $|2 - E(X)| \approx 0,95 < 1$, $|3 - E(X)| \approx 1,95 > 1$
    et donc $P(|X - E(X)| \geqslant 1) = P(X = 0) + P(X = 3) = 0,5625$
  - avec $\epsilon = 2$ on a $P(|X - E(X)| \geqslant 2) \leqslant \frac{1,169678}{4} \approx 0,29$ alors que $|X - E(X)| \geqslant 2 = \varnothing$
    En effet $|0 - E(X)| \approx 1,05 < 2$, $|1 - E(X)| \approx 0,05 < 2$, ..., $|3 - E(X)| \approx 1,95 < 2$
    et donc $P(|X - E(X)| \geqslant 2) = 0$

  Voyons ce qu'apporte la reformulation avec $\sigma(X) \approx 1,082$ et $2\sigma(X) \approx 2,163$:
  - avec $\epsilon = \sigma(X)$ on a $P(|X - E(X)| \geqslant \sigma(X)) \leqslant 1$ alors que $|X - E(X)| \geqslant \sigma(X) = (X < 3)$
    En effet $|0 - E(X)| \approx 1,05 < \sigma(X)$, ..., $|2 - E(X)| \approx 0,95 < \sigma(X)$, $|3 - E(X)| \approx 1,95 > \sigma(X)$
    et donc $P(|X - E(X)| \geqslant \sigma(X)) = 1 - P(X = 3) \approx 0,86$
  - avec $\epsilon = 2\sigma(X)$ on a $P(|X - E(X)| \geqslant 2\sigma(X)) \leqslant \frac{1}{4}$ alors que $|X - E(X)| \geqslant 2\sigma(X) = \varnothing$
    En effet $|0 - E(X)| \approx 1,05 < 2\sigma(X)$, ..., $|3 - E(X)| \approx 1,95 < 2\sigma(X)$
    et donc $P(|X - E(X)| \geqslant 2) = 0$

3.b. L'inégalité de concentration

PROPRIÉTÉ 8.15 (POUR UN ÉCHANTILLON) Soit un échantillon $(X_1, X_2, \ldots, X_m)$ de variables d'espérance $E(X_i) = \bar{x}$ et de variance $V(X_i) = \sigma^2$ pour tout $i \in [\![1, m]\!]$.
Pour la variable $M_m = \frac{X_1+X_2+\cdots+X_m}{m}$, quel que soit $\epsilon > 0$, on a $P(|M_m - \bar{x}| \geqslant \epsilon) \leqslant \frac{\sigma^2}{m\epsilon^2}$

DÉMONSTRATION L'inégalité de Bienaymé-Tchebychev appliquée à la variable M_m s'écrit $P(|M_m - E(M_m)| \geqslant \epsilon) \leqslant \frac{V(M_m)}{\epsilon^2}$
La propriété 8.8 nous indique les paramètres de M_m : $E(M_m) = \bar{x}$ et $V(M_m) = \frac{\sigma^2}{m}$.
Par substitution dans l'inégalité, on obtient $P(|M_m - \bar{x}| \geqslant \epsilon) \leqslant \frac{\sigma^2}{m\epsilon^2}$

Remarques :

- Cette inégalité est parfois appelée *loi des grands nombres* d'après Siméon Denis Poisson qui utilisa cette expression en 1837. Pourquoi parle t-on de grands nombres ? Cette inégalité indique que lorsque la taille de l'échantillon augmente – lorsque m devient un grand nombre – la probabilité que la moyenne des variables s'écarte de plus d'une certaine valeur ϵ devient de plus en plus petite. La limite de $P(|M_m - \bar{x}| \geqslant \epsilon)$ est nulle quand m tend vers l'infini, c'est d'ailleurs plutôt cette dernière affirmation que l'on devrait appeler loi des grands nombres. Cette loi était déjà perçue dans le domaine des jeux de hasard et Jacques Bernoulli en fit une première théorisation relative à la loi binomiale dans son *Ars Conjectandi* de 1737.
- Pourquoi parle t-on de concentration ? Car M_m et $\bar{x}$ sont des paramètres centraux : le premier est la moyenne de l'échantillon et le second l'espérance de toutes les variables impliquées dans cet échantillon. On majore, avec cette inégalité, le risque que l'échantillon ne soit pas assez concentré, c'est-à-dire que sa moyenne s'écarte trop de l'espérance. Celle-ci apparait alors comme la valeur théorique vers laquelle tend la moyenne de l'échantillon.

Propriété 8.16 (Application au schéma de Bernoulli) Soit un échantillon $(X_1, \dots, X_m)$ de variables de Bernoulli de paramètre p. Les variables $S_m = X_1 + \dots + X_m$ et $M_m = \frac{S_m}{m}$ vérifient :

$$\forall \epsilon > 0, \quad P(|S_m - mp| \geqslant \epsilon) \leqslant \frac{mp(1-p)}{\epsilon^2} \text{ et } P(|M_m - p| \geqslant \epsilon) \leqslant \frac{p(1-p)}{m\epsilon^2}$$

Démonstration La variable $S_m = X_1 + X_2 + \dots + X_m$ compte le nombre de succès car $X_i(\Omega) = \{0, 1\}$ pour tout $i \in [\![1, m]\!]$. Elle suit donc une loi binomiale $\mathscr{B}(m, p)$ dont les caractéristiques sont $E(S_m) = mp$ et $V(S_m) = mp(1-p)$.
En appliquant l'inégalité de Bienaymé-Tchebychev à cette variable on obtient $P(|S_m - mp| \geqslant \epsilon) \leqslant \frac{mp(1-p)}{\epsilon^2}$

La variable $M_m = \frac{S_m}{m}$ a pour caractéristiques :
$E(\frac{S_m}{m}) = \frac{E(S_m)}{m} = \frac{mp}{m} = p$ et $V(M_m) = V(\frac{S_m}{m}) = \frac{V(S_m)}{m^2} = \frac{mp(1-p)}{m^2} = \frac{p(1-p)}{m}$
En appliquant l'inégalité de Bienaymé-Tchebychev à cette variable on obtient $P(|M_m - p| \geqslant \epsilon) \leqslant \frac{p(1-p)}{m\epsilon^2}$

__**Intervalle de fluctuation**__ : Un échantillon $(X_1, \dots, X_m)$ de variables de Bernoulli de paramètre p étant donné, la variable $S_m = X_1 + X_2 + \dots + X_m$ compte le nombre de succès et suit une loi $\mathscr{B}(m, p)$. Dans la partie 8.2.3, on a donné comme intervalle de fluctuation de la variable $M_m = \frac{S_m}{m}$ qui évalue la fréquence des succès au seuil de 95% l'intervalle $\left[\frac{a}{m}; \frac{b}{m}\right]$ où a et b sont des entiers compris entre 0 et m, tels que $P(\frac{a}{m} \leqslant M_m \leqslant \frac{b}{m}) \geqslant 0,95$.

L'évènement $|M_m - p| \geqslant \epsilon$ est le contraire de l'évènement $|M_m - p| < \epsilon$ soit $p - \epsilon < M_m < p + \epsilon$.
Par conséquent $P(p - \epsilon < M_m < p + \epsilon) = 1 - P(|M_m - p| \geqslant \epsilon)$.
La propriété 8.16 implique $1 - P(|M_m - p| \geqslant \epsilon) \geqslant 1 - \frac{p(1-p)}{m\epsilon^2}$.
On en déduit $P(p - \epsilon < M_m < p + \epsilon) \geqslant 1 - \frac{p(1-p)}{m\epsilon^2}$.
Pour comparer avec la règle édictée au seuil de 95% dans la partie 8.2.3, choisissons ϵ de manière à avoir $1 - \frac{p(1-p)}{m\epsilon^2} = 0,95$ soit $0,05 = \frac{p(1-p)}{m\epsilon^2} \iff 0,05m\epsilon^2 = p(1-p) \Longrightarrow \epsilon = \sqrt{\frac{p(1-p)}{0,05m}}$.

Appliqué à la situation de l'exemple 3, la loi suivie par S_m est $\mathscr{B}(100, \frac{1}{6})$ on obtient $\epsilon = \sqrt{\frac{\frac{1}{6}\times\frac{5}{6}}{0,05\times 100}} = \sqrt{\frac{5}{5\times 36}} = \sqrt{\frac{1}{36}} = \frac{1}{6} \approx 0,167$.
Par conséquent $P(\frac{1}{6} - \frac{1}{6} < M_m < \frac{1}{6} + \frac{1}{6}) \geqslant 0,95$, approximativement $P(0 < M_m < 0,333) \geqslant 0,95$.
L'intervalle de confiance au seuil de 95% qu'on peut déduire de l'inégalité de Bienaymé-Tchebychev est donc $IF = [0; 0,333]$. Cet intervalle est moins précis que ceux déterminés plus haut mais l'obtenir nécessite peu de calculs et on peut étendre la méthode très facilement à d'autres seuils que 95%, par exemple 90%, 99% ou 99,5%.

Généralisons cette méthode : supposons qu'on cherche à déterminer l'intervalle de fluctuation au seuil de t% avec $t \in [0, 100]$. On doit avoir $P(p - \epsilon < M_m < p + \epsilon) \geqslant \frac{t}{100}$ avec la valeur de ϵ donnée par l'égalité $1 - \frac{p(1-p)}{m\epsilon^2} = \frac{t}{100}$ qui conduit à $\epsilon = 10\sqrt{\frac{p(1-p)}{m(100-t)}}$.
L'intervalle de fluctuation est alors $IF = [p - \epsilon; p + \epsilon]$.
Voici quelques valeurs obtenues à partir de cette formule pour $p = \frac{1}{6}$ et $m = 100$ ou $m = 1000$:

t	ε	p_ε	p+ε
90	0,118	0,049	0,285
91	0,124	0,042	0,291
92	0,132	0,035	0,298
93	0,141	0,026	0,308
94	0,152	0,015	0,319
95	0,167	0,000	0,333
96	0,186	-0,020	0,353
97	0,215	-0,048	0,382
98	0,264	-0,097	0,430
99	0,373	-0,206	0,539
99,5	0,527	-0,360	0,694

m	p
100	0,166666666

t	ε	p_ε	p+ε
90	0,037	0,129	0,204
91	0,039	0,127	0,206
92	0,042	0,125	0,208
93	0,045	0,122	0,211
94	0,048	0,119	0,215
95	0,053	0,114	0,219
96	0,059	0,108	0,226
97	0,068	0,099	0,235
98	0,083	0,083	0,250
99	0,118	0,049	0,285
99,5	0,167	0,000	0,333

m	p
1000	0,166666666