Lycée Jules Ferry, Cannes
Année 2020-2021

Classe préparatoire TSI 2e année
Mathématiques - D. Broizat

Chapitre 13
Variables aléatoires discrètes

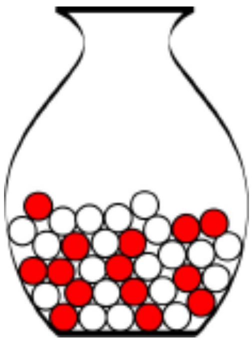

Table des matières

On aura toujours en contexte un espace probabilisé $(\Omega, \mathbb{P})$ (souvent fini ou dénombrable).

I Généralités sur les variables aléatoires discrètes

1) Définitions et notations

Définition 1 (Variable aléatoire réelle discrète).
Étant donné un espace probabilisé $(\Omega, \mathbb{P})$, on appelle **variable aléatoire**
....
réelle discrète *sur Ω toute application $X : \Omega \to \mathbb{R}$ telle que l'image $X(\Omega)$*
....
est un ensemble fini ou dénombrable.
....

Rappel
L'image d'une application est l'ensemble des valeurs atteintes.
Si $X : \Omega \to \mathbb{R}$, on a donc $X(\Omega) = \{X(\omega),\ \omega \in \Omega\} = \{x \in \mathbb{R},\ \exists \omega \in \Omega,\ x = X(\omega)\}$.

Remarque
- On peut généraliser en définissant des variables aléatoires discrètes $X : \Omega \to E$, où E est un ensemble quelconque.
- On peut très bien avoir Ω infini indénombrable mais $X(\Omega)$ fini ou dénombrable.

Notation
Si l'image de X est finie, on la notera $X(\Omega) = \{x_1, \cdots, x_n\}$ avec $n \in \mathbb{N}^*$.
Si l'image de X est dénombrable, on la notera $X(\Omega) = \{x_k,\ k \in \mathbb{N}\}$.
Dans tous les cas, on peut noter $X(\Omega) = \{x_i,\ i \in I\}$ avec I un ensemble d'indices fini ou dénombrable.

Exemple
On lance un dé à 6 faces équilibré. On gagne 1 euro lorsque le dé donne un résultat pair, et où l'on perd 1 euro lorsque le résultat est impair.
Modéliser cette expérience par une variable aléatoire discrète.
On peut modéliser cette expérience par une variable aléatoire réelle discrète X sur l'univers $\Omega = \{1, \cdots, 6\}$ muni de l'équiprobabilité $\mathbb{P}$ et $X : \Omega \to \{-1, 1\}$ définie par $X(2) = X(4) = X(6) = 1$ et $X(1) = X(3) = X(5) = -1$.
Ici, Ω est fini (6 éléments) et l'image $X(\Omega)$ est finie (2 éléments).

Exemple

On lance simultanément deux dés équilibrés. On effectue la somme des faces obtenues.
Modéliser cette expérience par une variable aléatoire discrète.

On peut modéliser cette expérience par une variable aléatoire réelle discrète S sur l'univers $\Omega = \{1, \cdots, 6\}^2$, muni de l'équiprobabilité $\mathbb{P}$ et $S : \Omega \to \{2, \ldots, 12\}$ définie par $S(n, m) = n + m$ pour tout $(n, m) \in \{1, \ldots, 6\}^2$.

Ici, Ω est fini (36 éléments) et l'image $S(\Omega)$ est finie (11 éléments).

Exemple

On lance indéfiniment une pièce équilibrée. On note le numéro de lancer où *Pile* apparaît pour la première fois.

Modéliser cette expérience par une variable aléatoire discrète.

Soit Ω l'ensemble des suites $(a_n)_{n\in\mathbb{N}^*}$ à valeurs dans l'ensemble $\{P, F\}$. Cet univers est infini et indénombrable.

On admet qu'on peut munir cet univers d'une probabilité $\mathbb{P}$.

On pose alors $X : \Omega \to \mathbb{N}^* \cup \{+\infty\}$ définie par $X((a_n)_{n\in\mathbb{N}^*}) = \inf\{n \in \mathbb{N}^*,\ a_n = P\}$.

Ici, l'image $X(\Omega) = \mathbb{N}^* \cup \{+\infty\}$ est infinie et dénombrable, car en bijection avec $\mathbb{N}$.

Notation
Si $X : \Omega \to E$ est une variable aléatoire dans un ensemble quelconque E, alors pour toute partie $A \subset E$, on peut considérer l'événement

$$(X \in A) = X^{-1}(A) = \{\omega \in \Omega,\ X(\omega) \in A\}$$

formé des antécédents de la partie A par l'application X.
Dessin :

Si de plus X est à valeurs réelles, alors on peut considérer les événements :

- $(X = a) = X^{-1}(\{a\}) = \{\omega \in \Omega,\ X(\omega) = a\}$.
  C'est l'ensemble des éventualités $\omega \in \Omega$ pour lesquelles la variable aléatoire X prend la valeur a.
- $(X \leq a) = X^{-1}(]-\infty, a]) = \{\omega \in \Omega,\ X(\omega) \leq a\}$.

- $(X \geq a) = X^{-1}([a, +\infty[) = \{\omega \in \Omega,\ X(\omega) \geq a\}$.
- $(X < a) = X^{-1}(]-\infty, a[) = \{\omega \in \Omega,\ X(\omega) < a\}$.
- $(X > a) = X^{-1}(]a, +\infty[) = \{\omega \in \Omega,\ X(\omega) > a\}$.
- Pour $b > a$, $(a \leq X \leq b) = \{\omega \in \Omega,\ a \leq X(\omega) \leq b\}$, etc.

Exemple
On additionne les résultats de deux dés équilibrés.
On a la variable aléatoire "somme" $S : \{1, \cdots, 6\}^2 \to \{2, \cdots, 12\}$.
Décrire les événements $A = (S = 4)$ et $B = (9 \leq S < 11)$.
On a

$$A = (S = 4) = \{(1,3), (2,2), (3,1)\},$$

$$B = (S = 9) \cup (S = 10) = \{(3,6), (4,5), (5,4), (6,3), (4,6), (5,5), (6,4)\}.$$

Proposition 2.
Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire discrète. Alors les événements $(X = x)_{x \in X(\Omega)}$ forment un système complet d'événements de Ω.

Preuve : Rappelons que pour tout $x \in X(\Omega)$:

$$(X = x) = \{\omega \in \Omega,\ X(\omega) = x\}.$$

- Soit $x_1, x_2 \in X(\Omega)$ avec $x_1 \neq x_2$. Alors les événements $(X = x_1)$ et $(X = x_2)$ sont incompatibles, car sinon, il existe $\omega \in \Omega$ tel que $X(\omega) = x_1$ et $X(\omega) = x_2$, donc $x_1 = x_2$, ce qui est absurde.
- Montrons que $\displaystyle\bigcup_{x \in X(\Omega)} (X = x) = \Omega$. L'inclusion $\subset$ est vraie par définition.
  Réciproquement, étant donné $\omega \in \Omega$, on pose $a = X(\omega)$ et on a alors $\omega \in (X = a)$, donc a fortiori $\omega \in \displaystyle\bigcup_{x \in X(\Omega)} (X = x)$

$\square$

2) Loi d'une variable aléatoire discrète

Définition 3 (Loi d'une variable aléatoire discrète).
Étant donnée une variable aléatoire discrète $X : \Omega \to \mathbb{R}$, on appelle **loi de X** *la fonction $\mathbb{P}_X : \mathcal{P}(X(\Omega)) \to [0,1]$ définie par*

$$\forall A \subset X(\Omega), \quad \mathbb{P}_X(A) = \mathbb{P}(X \in A).$$

Proposition 4 (La loi est une mesure de probabilité sur l'image).
Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle discrète. Alors, la loi $\mathbb{P}_X$ est une probabilité sur l'univers $X(\Omega)$.

Preuve :

- Par définition, $\mathbb{P}_X$ est à valeurs dans $[0,1]$.
- $\mathbb{P}_X\big(X(\Omega)\big) = \mathbb{P}\big(X \in X(\Omega)\big) = \mathbb{P}(\Omega) = 1$ car $\forall \omega \in \Omega$, $X(\omega) \in X(\Omega)$.
- Soit $(A_n)_{n \in \mathbb{N}}$ une suite de parties deux à deux disjointes de $X(\Omega)$.
  Les événements $(X \in A_n)$ sont deux à deux disjoints – car chaque valeur $X(\omega)$ est

dans un seul des A_n – et

$$\left(X \in \bigcup_{n\in\mathbb{N}} A_n\right) = \{\omega \in \Omega,\ \exists n \in \mathbb{N},\ X(\omega) \in A_n\} = \bigcup_{n\in\mathbb{N}} (X \in A_n),$$

donc, par additivité dénombrable de la probabilité $\mathbb{P}$, on obtient

$$\mathbb{P}_X\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mathbb{P}\left(X \in \bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(X \in A_n) = \sum_{n=0}^{+\infty} \mathbb{P}_X(A_n).$$

□

Remarque

- Et ainsi, plutôt que de désigner $\mathbb{P}_X$ simplement par "loi de X", on peut aussi employer la terminologie "loi de probabilité de X".
- Pour une variable aléatoire discrète $X : \Omega \to \mathbb{R}$, on a donc **deux mesures de probabilité** : la probabilité $\mathbb{P}$ sur Ω, et la loi $\mathbb{P}_X$ sur l'image $X(\Omega)$.
- Une fois qu'on a l'espace probabilisé $\Big(X(\Omega), \mathbb{P}_X\Big)$, on peut très bien "oublier" sur quel espace Ω a été introduit X. C'est pour cela que, régulièrement, les énoncés ne s'embêtent pas à décrire l'espace probabilisé sur lequel X est définie, et se contentent de dire qu'on admet qu'il en existe un …

Proposition 5 (Détermination de la loi d'une v.a. discrète).
Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire discrète. La loi de probabilité $\mathbb{P}_X$ est
uniquement déterminée par ses valeurs $\mathbb{P}_X(\{x\}) = \mathbb{P}(X = x)$ où x décrit $X(\Omega)$.

Preuve : La probabilité $\mathbb{P}_X$ sur l'univers $X(\Omega)$ est uniquement déterminée par les probabilités des événements élémentaires $\{x\}$ (voir CH.8). □

Méthode
Pour déterminer la loi d'une variable aléatoire discrète $X : \Omega \to \mathbb{R}$:

- on commence par déterminer $X(\Omega)$ (l'ensemble des valeurs prises par X).
- on calcule ensuite $\mathbb{P}(X = x)$ pour toute valeur $x \in X(\Omega)$.

Exemple (Somme des résultats des lancers de deux dés)
Calcul de la loi de $S : \{1, \cdots, 6\}^2 \to \{2, \cdots, 12\}$ définie par $S(\omega_1, \omega_2) = \omega_1 + \omega_2$.

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

x	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(S = x)$											

Définition 6 (Suivre la même loi).
Soient X et Y deux variables aléatoires discrètes définies sur le même espace probabilisé $(\Omega, \mathbb{P})$. On dit que X et Y suivent la même loi si $X(\Omega) = Y(\Omega)$ et si $\forall x \in X(\Omega),\ \mathbb{P}(X = x) = \mathbb{P}(Y = x)$.

Notation
On notera $X \sim Y$ pour dire que X et Y suivent la même loi.

Remarque
A toute variable aléatoire correspond une unique loi mais des variables aléatoires différentes peuvent avoir la même loi.
Ainsi, lors du lancer de deux dés équilibrés, si l'on note D_1 la variable aléatoire résultat du premier dé et D_2 celle du résultat du deuxième dé, D_1 et D_2 suivent la même loi.

Théorème 7 (Caractérisation de la loi d'une VA discrète).
Soit I un ensemble fini ou dénombrable, et $(x_i)_{i \in I}$, $(p_i)_{i \in I}$ deux familles de réels.
Si pour tout $i \in I$, $p_i \geqslant 0$ et si $\sum_{i \in I} p_i = 1$, alors il existe un espace probabilisé $(\Omega, \mathbb{P})$ et une variable aléatoire réelle discrète $X : \Omega \to \mathbb{R}$ dont la loi est définie par $\forall i \in I,\ \mathbb{P}(X = x_i) = p_i$.

Méthode
Ce théorème permet, lorsqu'on nous donne des valeurs pour p_i et x_i, de déterminer si ces valeurs sont la loi d'une variable aléatoire réelle discrète.

Exemple
Montrer qu'il existe une variable aléatoire $X : \Omega \to \mathbb{N}^*$ dont la loi est donnée par :

$$\forall n \in \mathbb{N}^*, \qquad \mathbb{P}(X = n) = \frac{1}{n(n+1)}.$$

On a $\forall n \geq 1,\ \frac{1}{n(n+1)} \geq 0$ et

$$\sum_{k=1}^{+\infty} \frac{1}{k(k+1)} = \lim_{N\to+\infty} \sum_{k=1}^{N} \left(\frac{1}{k} - \frac{1}{k+1}\right) = \lim_{N\to+\infty} \left(1 - \frac{1}{N+1}\right) = 1.$$

3) Fonction de répartition

Définition 8 (Fonction de répartition d'une VA discrète).
Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle discrète. On appelle
fonction de répartition *de X la fonction $F_X : \mathbb{R} \to [0;1]$ définie par*
$\forall x \in \mathbb{R}, \quad F_X(x) = \mathbb{P}(X \leq x)$.

Remarque
- Une fonction de répartition est **croissante**, car si $x \leq y$, on a $]-\infty; x] \subset]-\infty; y]$, donc
$$F_X(x) = \mathbb{P}(X \in]-\infty; x]) \leq \mathbb{P}(X \in]-\infty; y]) = F_X(y).$$
- La fonction de répartition d'une variable aléatoire discrète est une fonction en escalier, et on a toujours $\lim_{x\to-\infty} F_X(x) = 0$ et $\lim_{x\to+\infty} F_X(x) = 1$.

Proposition 9 (Loi d'une VA discrète à partir de sa fct de répartition).
Soit $X : \Omega \to \mathbb{R}$. Notons $X(\Omega) = \{x_i,\ i \in I\}$ (avec I fini ou dénombrable).
Si les x_i sont rangés par ordre croissant, alors pour tout $i \in I$ tel que $i-1 \in I$,
on a $\mathbb{P}(X = x_i) = F_X(x_i) - F_X(x_{i-1})$. .

Preuve : Vu que les $(x_i)_{i\in I}$ sont rangés par ordre croissant, on a

$$(X \leqslant x_i) = (X = x_i) \cup (X \leqslant x_{i-1}).$$

De plus, les événements $(X = x_i)$ et $(X \leqslant x_{i-1})$ sont incompatibles donc :

$$\mathbb{P}(X \leqslant x_i) = \mathbb{P}(X = x_i) + \mathbb{P}(X \leqslant x_{i-1}) \Longleftrightarrow \mathbb{P}(X = x_i) = F_X(x_i) - F_X(x_{i-1}).$$

□

Exemple
Un sac contient 4 boules numérotées de 1 à 4. On tire deux boules avec remise. On note X_1 le numéro de la première boule, X_2 le numéro de la seconde boule, et Y le plus grand des deux numéros obtenus.
Déterminer la fonction de répartition de Y, puis en déduire la loi de Y.

On remarque, tout d'abord, que Y prend les valeurs 1, 2, 3 ou 4, donc $Y(\Omega) = \{1, 2, 3, 4\}$.

De plus, on remarque qu'il est plus facile de calculer $\mathbb{P}(Y \leqslant k)$ que $\mathbb{P}(Y = k)$. Par exemple,

$$\begin{aligned}
[Y = 3] = {} & ([X_1 = 1] \cap [X_2 = 3]) \cup ([X_1 = 2] \cap [X_2 = 3]) \cup ([X_1 = 3] \cap [X_2 = 3]) \cup \\
& ([X_1 = 3] \cap [X_2 = 2]) \cup ([X_1 = 3] \cap [X_2 = 1]);\\
[Y \leqslant 3] = {} & [X_1 \leqslant 3] \cap [X_2 \leqslant 3].
\end{aligned}$$

Notons F_Y la fonction de répartition de la variable aléatoire réelle Y.

Nous allons donc calculer $F_Y(1)$, $F_Y(2)$, $F_Y(3)$ et $F_Y(4)$ puis en déduire la loi de Y grâce à la propriété précédente.

- $F_Y(1) = \mathbb{P}(Y \leqslant 1) = \mathbb{P}(Y = 1) = \mathbb{P}([X_1 = 1] \cap [X_2 = 1])$. Les tirages se font avec remise donc ils sont indépendants et, ainsi, les événements $[X_1 = 1]$ et $[X_2 = 1]$ sont indépendants.

  On a donc $F_Y(1) = \mathbb{P}(X_1 = 1) \times \mathbb{P}(X_2 = 1) = \dfrac{1}{16}$.
- Grâce au même argument d'indépendance,

  $$F_Y(2) = \mathbb{P}(Y \leqslant 2) = \mathbb{P}([X_1 \leqslant 2] \cap [X_2 \leqslant 2]) = \mathbb{P}(X_1 \leqslant 2) \times \mathbb{P}(X_2 \leqslant 2) = \frac{2}{4} \times \frac{2}{4} = \frac{1}{4}.$$
- De même, $F_Y(3) = \dfrac{3}{4} \times \dfrac{3}{4} = \dfrac{9}{16}$.

- Et enfin, l'événement $[Y \leqslant 4]$ est certain donc $F_Y(4) = 1$.

On peut maintenant en déduire la loi de la variable Y :

$$\mathbb{P}(Y=1) = F_Y(1) = \frac{1}{16}; \quad \mathbb{P}(Y=2) = \mathbb{P}(Y \leqslant 2) - \mathbb{P}(Y \leqslant 1) = \frac{3}{16};$$

$$\mathbb{P}(Y=3) = \mathbb{P}(Y \leqslant 3) - \mathbb{P}(Y \leqslant 2) = \frac{5}{16}; \quad \mathbb{P}(Y=4) = \mathbb{P}(Y \leqslant 4) - \mathbb{P}(Y \leqslant 3) = \frac{7}{16}.$$

4) Lois usuelles à image finie

Des rappels de TSI 1 pour commencer, avec des lois prenant un nombre fini de valeurs ($X(\Omega)$ est fini) :

a) Loi uniforme

Définition 10 (Loi uniforme).
Etant donné un ensemble fini E, on dit qu'une variable aléatoire X définie sur Ω ....
suit **la loi uniforme sur *E*** *si $X(\Omega) = E$ et pour tout $x \in E$, $\mathbb{P}(X = x) = \frac{1}{\#E}$.*
....
On note alors $X \sim \mathcal{U}(E)$.
....

Remarque (Lien avec l'équiprobabilité)
$X : \Omega \to E$ suit une loi uniforme ssi tous les singletons $\{x\}$ de l'ensemble fini E ont même probabilité $\mathbb{P}_X(\{x\})$. La loi uniforme est donc l'équiprobabilité sur E.

Exemple
- Si on lance un dé équilibré et on note X le résultat obtenu, alors X suit la loi uniforme sur $\Omega = \{1, \cdots, 6\}$.
- La variable "somme de deux dés équilibrés" S ne suit pas une loi uniforme, puisque par exemple, $\mathbb{P}(S = 2) \neq \mathbb{P}(S = 3)$.

b) Loi de Bernoulli

Définition 11 (Loi de Bernoulli).
Soit un réel $p \in [0;1]$. On dit qu'une variable aléatoire définie sur X ....
suit la **loi de Bernoulli de paramètre p** *si $X(\Omega) = \{0;1\}$ ainsi que* ....
$\mathbb{P}(X=1) = p$ *et* $\mathbb{P}(X=0) = 1-p$. *On note alors $X \sim \mathcal{B}(p)$* . ...

Remarque (Schéma "succès-échec")
Si l'on prend comme convention que $X = 1$ désigne le succès d'une expérience et $X = 0$ l'échec de celle-ci, alors cette loi est naturellement celle codant la probabilité de succès d'une expérience.

Exemple (Pile ou Face)
Si $\Omega = \{P, F\}$ et si l'on définit $X : \Omega \to \{0, 1\}$ par $X(P) = 1$ et $X(F) = 0$, alors X code la probabilité d'obtention d'un *Pile*.

- Si la pièce est équilibrée, alors X suit la loi de Bernoulli $\mathcal{B}\left(\frac{1}{2}\right)$, qui coïncide dans ce cas particulier avec la loi uniforme $\mathcal{U}\big(\{0,1\}\big)$.
- Si on suppose que *Pile* a une probabilité de sortir de $\dfrac{2}{3}$, alors $X \sim \mathcal{B}\left(\frac{2}{3}\right)$.

c) Loi binomiale

Définition 12 (Loi binomiale).
Soient un entier $n \in \mathbb{N}$ et un réel $p \in [0,1]$. On dit qu'une variable aléatoire ....
X définie sur Ω suit la **loi binomiale de paramètres (n,p)** ....
si $X(\Omega) = \{0, \cdots, n\}$ *et* $\forall k \in \{0, \ldots, n\}, \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$. ....
On note alors $X \sim \mathcal{B}(n,p)$ . ...

Remarque

- On a $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$ donc on a bien $\sum_{0 \leq k \leq n} \mathbb{P}(X = k) = 1$, ce qui confirme que $\mathbb{P}_X$ est une loi de probabilité sur $\{0, 1, \cdots, n\}$.
- La loi binomiale $\mathcal{B}(1,p)$ coïncide avec la loi de Bernoulli $\mathcal{B}(p)$.

Exemple (Tirages successifs avec remise)
On dispose d'urne contenant 10 boules indiscernables, 6 rouges et 4 bleues. On effectue 3 tirages avec remise. On note X la variable aléatoire renvoyant le nombre de boules rouges obtenues au total. Déterminer la loi de X.

Pour tout $1 \leq k \leq 3$, on note R_k l'événement "obtenir une boule rouge au k^e tirage". Les événements $(R_k)_{1\leq k\leq 3}$ sont mutuellement indépendants (vu que les tirages sont avec remise).
On a $X(\Omega) = \{0, 1, 2, 3\}$, et

$$\mathbb{P}(X = 0) = \mathbb{P}(\overline{R_1} \cap \overline{R_2} \cap \overline{R_3}) = \left(\frac{4}{10}\right)^3 = \left(\frac{2}{5}\right)^3,$$

$$\mathbb{P}(X = 1) = \mathbb{P}((R_1 \cap \overline{R_2} \cap \overline{R_3}) \cup (\overline{R_1} \cap R_2 \cap \overline{R_3}) \cup (\overline{R_1} \cap \overline{R_2} \cap R_3)) = 3 \times \left(\frac{3}{5}\right)^1 \times \left(\frac{2}{5}\right)^2,$$

$$\mathbb{P}(X = 2) = \mathbb{P}((R_1 \cap R_2 \cap \overline{R_3}) \cup (R_1 \cap \overline{R_2} \cap R_3) \cup (\overline{R_1} \cap R_2 \cap R_3)) = 3 \times \left(\frac{3}{5}\right)^2 \times \left(\frac{2}{5}\right)^1,$$

$$\mathbb{P}(X = 3) = \mathbb{P}(R_1 \cap R_2 \cap R_3) = \left(\frac{3}{5}\right)^3.$$

Finalement, on a $X \sim \mathcal{B}\big(3, \frac{3}{5}\big)$.

Proposition 13 (Réalisation concrète de la loi binomiale).
Soit X la variable aléatoire comptant le **nombre de succès** *dans la répétition de n expériences aléatoires à deux issues, supposées mutuellement indépendantes, et de même probabilité de succès $p \in [0; 1]$. Alors X suit la loi binomiale $\mathcal{B}(n, p)$.*

Preuve : Les valeurs de X sont comprises dans $\{0, \ldots, n\}$.
Pour chaque indice $i \in \{1, \cdots, n\}$ notons A_i l'événement "la i^e répétition de l'expérience a donné un succès". Les événements $A_1, \cdots, A_n$ sont mutuellement indépendants et chacun de probabilité p.
Pour $k \in \{0, \cdots, n\}$ fixé, décomposons l'événement $(X = k)$:
$(X = k)$ est réalisé ssi exactement k événements parmi $A_1, \cdots, A_n$ sont réalisés et $n-k$ ne le sont pas.

Formellement, cela s'écrit :

$$(X = k) = \bigcup_{1 \leq i_1 < \cdots < i_k \leq n} \left(A_{i_1} \cap \cdots \cap A_{i_k} \cap \left(\bigcap_{j \notin \{i_1, \cdots, i_k\}} \overline{A_j} \right) \right)$$

$(X = k)$ est donc réunion disjointe de $\binom{n}{k}$ événements de même probabilité $p^k(1-p)^{n-k}$, donc

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

$\square$

Exemple
On lance une pièce équilibrée n fois. Le nombre X de *Pile* obtenus suit la loi binomiale $\mathcal{B}(n,p)$ avec $p = \frac{1}{2}$.

5) Lois usuelles à image dénombrable

Passons maintenant aux nouveautés du programme de TSI 2, avec deux lois prenant une quantité dénombrable de valeurs .

a) Loi géométrique

Définition 14 (Loi géométrique).
Soit un réel $p \in]0;1[$. On dit qu'une variable aléatoire X définie sur Ω
....
suit la **loi géométrique de paramètre p** *si $X(\Omega) = \mathbb{N}^*$ et*
....
$\forall k \in \mathbb{N}^*,\ \mathbb{P}(X = k) = (1-p)^{k-1}p$. On note alors $X \sim \mathcal{G}(p)$.
....

ATTENTION !
Ici, p est différent de 0 et de 1.

Remarque
On a bien

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

ce qui confirme que la loi géométrique est une probabilité sur $\mathbb{N}^*$.

Exemple (Tirages successifs avec remise jusqu'à ...)
Une urne contient b boules blanches et n boules noires, toutes indiscernables.
On tire une boule au hasard avec remise jusqu'à obtenir une boule noire.
On note X le numéro du tirage où cette boule noire apparaît.

1. Déterminer la loi de X.
2. Quelle est la probabilité qu'au moins k essais soient nécessaires pour que la boule noire apparaisse ?

Pour $i \in \mathbb{N}^*$, notons B_i l'événement "on tire une boule blanche au i^e tirage".
Vu la composition de l'urne et les tirages avec remise, les événements $(B_i)_{i\in\mathbb{N}^*}$ sont mutuellement indépendants, chacun de probabilité $\frac{b}{b+n}$.

1. On a $X(\Omega) = \mathbb{N}^*$, et pour tout $k \in \mathbb{N}^*$:
$$(X = k) = B_1 \cap \cdots \cap B_{k-1} \cap \overline{B_k},$$
donc
$$\mathbb{P}(X = k) = \left(\frac{b}{b+n}\right)^{k-1} \frac{n}{b+n}.$$
Donc X suit la loi géométrique $\mathcal{G}(p)$ avec $p = \frac{n}{b+n}$.

2. Pour tout $k \in \mathbb{N}^*$:
$$\mathbb{P}(X \geq k) = \sum_{n=k}^{+\infty} \mathbb{P}(X = n) = \sum_{n=k}^{+\infty} (1-p)^{n-1} p = p \sum_{n=0}^{+\infty} (1-p)^{n+k-1} = (1-p)^{k-1}.$$
On a donc $\mathbb{P}(X \geq k) = \left(\frac{b}{b+n}\right)^{k-1}$.

Proposition 15 (Réalisation concrète de la loi géométrique).
Soit X la variable aléatoire donnant le **rang du premier succès** *dans la répétition d'expériences aléatoires à deux issues, supposées mutuellement indépendantes, et de même probabilité de succès $p \in]0;1[$. Alors X suit la loi géométrique $\mathcal{G}(p)$.*

Preuve : Pour chaque indice $i \in \mathbb{N}^*$ notons A_i l'événement "la i^e répétition de l'expérience a donné un succès". Les événements $(A_i)_{i \geq 1}$ sont mutuellement indépendants et chacun de probabilité p.
Pour $k \in \mathbb{N}^*$ fixé, on a

$$(X = k) = \overline{A_1} \cap \cdots \cap \overline{A_{k-1}} \cap A_k,$$

donc par indépendance mutuelle :

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

□

b) Loi de Poisson

Définition 16 (Loi de Poisson).
Soit un réel $\lambda > 0$. On dit qu'une variable aléatoire X définie sur Ω
....
suit **la loi de Poisson de paramètre $\boldsymbol{\lambda}$** *si $X(\Omega) = \mathbb{N}$ et*
....
$\forall k \in \mathbb{N},\ \mathbb{P}(X = k) = e^{-\lambda} \times \dfrac{\lambda^k}{k!}$. *On note alors $X \sim \mathcal{P}(\lambda)$* .
....

ATTENTION !
Bien noter que le paramètre λ est strictement positif.

Remarque
On a bien les $\mathbb{P}(X = k) \geq 0$ et

$$\sum_{k=0}^{+\infty} \mathbb{P}(X = k) = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \times e^{\lambda} = 1$$

(en utilisant le DSE de e^{λ}), ce qui confirme que la loi de Poisson est bien une probabilité sur $\mathbb{N}$.

Le résultat suivant est essentiel pour comprendre l'intérêt de la loi de Poisson :

Proposition 17 (Approximation de la loi binomiale par la loi de Poisson). *Soient un réel $\lambda > 0$, une suite $(p_n)_{n\in\mathbb{N}}$ à valeurs dans $[0,1]$ et $(X_n)_{n\in\mathbb{N}}$ une suite*

de variables aléatoires avec $X_n \sim \mathcal{B}(n, p_n)$.

Si $n \times p_n \underset{n\to+\infty}{\longrightarrow} \lambda$, alors $\forall k \in \mathbb{N}, \quad \mathbb{P}(X_n = k) \underset{n\to+\infty}{\longrightarrow} e^{-\lambda} \times \dfrac{\lambda^k}{k!}$.

Preuve : Soit $k \in \mathbb{N}$. On a, pour tout entier $n \geq k$,

$$\mathbb{P}(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{1}{k!}\left(\frac{n!}{(n-k)!}\right) p_n^k (1-p_n)^{n-k}.$$

Comme k est fixé, on a

$$\frac{n!}{(n-k)!} = n(n-1)\cdots(n-k+1) \underset{n\to+\infty}{\sim} n^k.$$

Donc

$$\frac{n!}{(n-k)!} p_n^k \underset{n\to+\infty}{\sim} n^k p_n^k = (np_n)^k \underset{n\to+\infty}{\sim} \lambda^k$$

(puisque $\lim\limits_{n\to+\infty} np_n = \lambda$ par hypothèse). Enfin, l'hypothèse sur p_n se réécrit $p_n \underset{n\to+\infty}{\sim} \frac{\lambda}{n}$ donc on a le développement limité $p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right)$ et

$$(1-p_n)^{n-k} = e^{(n-k)\ln(1-p_n)} = e^{(n-k)\left(-\frac{\lambda}{n}+o\left(\frac{1}{n}\right)\right)} = e^{(-\lambda+o(1))} \underset{n\to+\infty}{\sim} e^{-\lambda}.$$

Donc, par produit d'équivalents, $\mathbb{P}(X_n = k) \underset{n\to+\infty}{\sim} e^{-\lambda}\dfrac{\lambda^k}{k!}$. $\square$

Remarque

- Ainsi, dans une situation avec $\boldsymbol{n}$ **grand et** $\boldsymbol{p}$ **petit**, on a intérêt à **approcher la loi binomiale $\mathcal{B}(n,p)$ par la loi de Poisson $\mathcal{P}(np)$** car les calculs sont bien plus commodes pour la loi de Poisson (ne serait-ce que calculer le coefficient binomial $\binom{n}{k}$ peut présenter de grosses difficultés pour certaines valeurs de k).
- Mais surtout, ce théorème offre une **interprétation de la loi de Poisson**.
  Elle s'apparente à un nombre de réussites lors de n essais indépendants suivant une loi de Bernoulli de paramètre p, pour n grand et p petit.
  La loi de Poisson s'apparente donc à la loi des événements rares.

  Rentrent dans cette catégorie, par exemple :
  * le nombre de personnes centenaires dans une population.
  * le nombre de coquilles par page dans un livre.
  * le nombre de faux numéros composés sur un laps de temps.
  * le nombre de particules désintégrées par radioactivité lors d'une période de temps fixée ...

6) Opérations sur les variables aléatoires discrètes

Proposition 18 (Opérations sur les variables aléatoires discrètes).
Soit $X : \Omega \to \mathbb{R}$ et $Y : \Omega \to \mathbb{R}$ deux variables aléatoires discrètes. Alors :

- *la somme $X+Y : \Omega \to \mathbb{R}$ définie par $(X+Y)(\omega) = X(\omega)+Y(\omega)$ est une variable aléatoire discrète ;*
- *pour tout $\lambda \in \mathbb{R}$, le produit $\lambda X : \Omega \to \mathbb{R}$ défini par $(\lambda X)(\omega) = \lambda X(\omega)$ est une variable aléatoire discrète ;*
- *le produit $XY : \Omega \to \mathbb{R}$ défini par $(XY)(\omega) = X(\omega) \times Y(\omega)$ est une variable aléatoire discrète ;*
- *pour toute fonction $f : X(\Omega) \to \mathbb{R}$, la composée $f \circ X : \Omega \to \mathbb{R}$ définie par $(f \circ X)(\omega) = f(X(\omega))$ est une variable aléatoire discrète.*

Preuve : Il faut vérifier dans chaque cas que l'image de la fonction en question est finie ou dénombrable (comme $X(\Omega)$ et $Y(\Omega)$), et c'est admis. $\square$

Notation
La variable aléatoire composée $f \circ X$ se note traditionnellement $f(X)$.
Exemples de variable aléatoire composée : $aX+b$ (avec $(a,b) \in \mathbb{R}^2$), X^2, X^3, ...

II Espérance d'une variable aléatoire discrète

Jusqu'à la fin du chapitre, on suppose toutes les variables aléatoires réelles, discrètes et définies sur un même espace probabilisé $(\Omega, \mathbb{P})$.

1) Définition pour une variable à image finie

Définition 19 (Espérance d'une variable aléatoire à image finie).
Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire à image finie : $X(\Omega) = \{x_1, \cdots, x_n\}$.
. . . .

On appelle **espérance de X** *la somme* $E(X) = \displaystyle\sum_{k=1}^{n} x_k \mathbb{P}(X = x_k)$.
. . . .

Remarque

- L'espérance est **une moyenne pondérée des valeurs prises par X**.
  La pondération est donnée par la probabilité d'apparition de chaque valeur.
- L'espérance de X ne dépend **que de la loi de X**, pas de X elle-même (en particulier elle ne dépend pas de l'univers Ω).

Exemple (Somme des résultats de deux lancers d'un dé équilibré)
On a $E(S) = \displaystyle\sum_{k=2}^{12} k\mathbb{P}(S = k) = 2 * \frac{1}{36} + 3 * \frac{2}{36} + 4 * \frac{3}{36} + 5 * \frac{4}{36} + 6 * \frac{5}{36} + 7 * \frac{6}{36} + 8 * \frac{5}{36} + 9 * \frac{4}{36} + 10 * \frac{3}{36} + 11 * \frac{2}{36} + 12 * \frac{1}{36} = \frac{252}{36} = 7$.

2) Définition pour une variable à image dénombrable

Généralisons maintenant la notion d'espérance pour les variables aléatoires à image dénombrable.

Définition 20 (Espérance d'une variable aléatoire à image dénombrable).
Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire à image dénombrable : $X(\Omega) = \{x_k,\ k \in \mathbb{N}\}$.

- *On dit que* **X admet une espérance finie** *lorsque*

  la série $\displaystyle\sum_{k=0}^{+\infty} x_k \mathbb{P}(X = x_k)$ *converge absolument* .
  . . . .
- *Si c'est le cas, on appelle* **espérance de X** *la somme infinie*

  $E(X) = \displaystyle\sum_{k=0}^{+\infty} x_k \mathbb{P}(X = x_k)$.
  . . . .

ATTENTION !
On parle bien de convergence **absolue**.

Remarque

- Une variable aléatoire à image finie possède toujours une espérance finie, mais pas une variable aléatoire à image dénombrable.
  Il faut vérifier la convergence de la série positive $\sum_{k\geq 0} |x_k| \ \mathbb{P}(X = x_k)$.
- On admet que la convergence absolue et la valeur de la somme infinie $\sum_{k=0}^{+\infty} x_k \mathbb{P}(X = x_k)$ ne dépendent pas de l'ordre d'énumération des éléments x_k de $X(\Omega)$ (mais en général, on choisit la suite (x_k) croissante).

Exemple (Si X est constante presque sûrement)
Si X est **constante presque sûrement**, c'est-à-dire si $\exists c \in \mathbb{R},\ \mathbb{P}(X = c) = 1$, alors X possède une espérance finie et $E(X) = c$.
En effet, en notant $X(\Omega) = \{x_n,\ n \in \mathbb{N}\}$, on a $\mathbb{P}(X = x_k) = 0$ si $x_k \neq c$, donc la série $\displaystyle\sum_{k=0}^{+\infty} x_k \mathbb{P}(X = x_k)$ a tous ses termes nuls sauf un (celui où $x_k = c$) : elle converge donc absolument et on a $E(X) = c\mathbb{P}(X = c) = c$.

Exemple (Variable ne possédant pas d'espérance finie)
Considérons une variable $X : \Omega \to \mathbb{N}^*$ dont la loi est donnée par :

$$\forall n \in \mathbb{N}^*, \qquad \mathbb{P}_X(\{n\}) = \mathbb{P}(X = n) = \frac{1}{n(n+1)}.$$

(c'est possible car $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{n(n+1)} = 1$, voir p.13).
Une telle variable **ne possède pas une d'espérance finie**.
En effet, on a $X(\Omega) = \{k,\ k \in \mathbb{N}^*\}$, et la série positive $\displaystyle\sum_{k\geq 1} k\mathbb{P}(X = k) = \sum_{k\geq 1} \frac{1}{k+1} = \sum_{k\geq 2} \frac{1}{k}$ diverge.

3) Propriétés de l'espérance

Proposition 21 (Propriétés de l'espérance).
Soient X et Y deux variables aléatoires discrètes sur un même espace probabilisé $(\Omega, \mathbb{P})$ (quelconque) et admettant toutes deux une espérance finie.
L'espérance possède les propriétés suivantes :

(i) ***Linéarité :*** *Pour tout $\lambda \in \mathbb{R}$, la variable $\lambda X + Y$ admet une espérance finie et* $E(\lambda X + Y) = \lambda E(X) + E(Y)$.

(ii) ***Positivité :*** *Si $\boldsymbol{X} : \boldsymbol{\Omega} \to \mathbb{R}_+$, alors on a* $E(X) \geq 0$ *et*
$E(X) = 0 \iff X = 0$ *presque sûrement (c'est-à-dire* $\mathbb{P}(X = 0) = 1$*)* .

(iii) ***Croissance :*** *Si on a $X \leq Y$, c'est-à-dire si pour tout $\omega \in \Omega$, $X(\omega) \leq Y(\omega)$, alors* $E(X) \leq E(Y)$.

Preuve :

(i) Admis (difficile).

(ii) Puisque X est à valeurs positives, on a $x_k \geq 0$ pour tout $k \in I$, donc

$$E(X) = \sum_{k \in I} x_k \underbrace{\mathbb{P}(X = x_k)}_{\geq 0} \geq 0.$$

Ensuite $E(X) = 0 \iff \big(\forall x \in X(\Omega),\ x\mathbb{P}(X = x) = 0\big) \iff \forall x \in X(\Omega) \setminus \{0\},\ \mathbb{P}(X = x) = 0 \iff \mathbb{P}(X = 0) = 1.$

(iii) Si $Y - X \geq 0$, alors $E(Y - X) \geq 0$, donc $E(Y) \geq E(X)$ par linéarité.

$\square$

Remarque
On a en particulier $E(aX + b) = aE(X) + b$ pour tous réels a et b.

Exemple (Somme de deux dés)
La variable aléatoire S sur $\Omega = \{1, \dots, 6\}^2$ est en fait somme des deux variables aléatoires D_1 et D_2 donnant les résultats du premier et deuxième dé (respectivement).
On a $E(D_1) = E(D_2) = \dfrac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \dfrac{7}{2} = 3,5$,
donc $E(S) = E(D_1 + D_2) = E(D_1) + E(D_2) = 7$.

Définition 22 (Variable aléatoire centrée).
On dit d'une variable aléatoire discrète X qu'elle est **centrée** *lorsque $E(X) = 0$.*

Remarque (Comment centrer une variable aléatoire)
Pour toute variable aléatoire discrète $X : \Omega \to \mathbb{R}$ qui possède une espérance finie, la variable aléatoire $X - E(X)$ est centrée.
En effet, par linéarité, $E\big(X - E(X)\big) = E(X) - E\big(E(X)\big) = E(X) - E(X) = 0$, d'après le résultat sur l'espérance d'une constante.

4) Espérance des lois usuelles

Proposition 23 (Espérance des loi uniforme, de Bernoulli et binomiale).

(i) Si $X \sim \mathcal{U}(\{x_1, \dots, x_n\})$, *alors* $E(X) = \frac{1}{n}\sum_{k=1}^{n} x_k$.

(ii) Si $X \sim \mathcal{B}(p)$, *alors* $E(X) = p$.

(iii) Si $X \sim \mathcal{B}(n,p)$, *alors* $E(X) = np$.

Preuve :

(i) $X(\Omega) = \{x_1, \cdots, x_n\}$ et $\mathbb{P}(X = x_k) = \frac{1}{n}$ pour tout $k \in \{1, \cdots, n\}$, donc

$$E(X) = \sum_{k=1}^{n} x_k \mathbb{P}(X = x_k) = \frac{1}{n}\sum_{k=1}^{n} x_k.$$

(ii) $X(\Omega) = \{0; 1\}$, donc $E(X) = 0 * \mathbb{P}(X = 0) + 1 * \mathbb{P}(X = 1) = 0 * (1-p) + 1 * p = p$.

(iii) $X(\Omega) = \{0, \cdots, n\}$ et $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ pour tout $k \in \{0, \cdots, n\}$ (ici $x_k = k$), donc

$$E(X) = \sum_{k=0}^{n} k\mathbb{P}(X = k) = \sum_{k=1}^{n} k\binom{n}{k} p^k (1-p)^{n-k}.$$

Pour calculer cette somme, on introduit la fonction $f : \mathbb{R} \to \mathbb{R}$ définie par

$$f(x) = \sum_{k=0}^{n} \binom{n}{k} x^k = (1+x)^n.$$

En dérivant, on obtient :

$$\forall x \in \mathbb{R}, \quad f'(x) = \sum_{k=1}^{n} k\binom{n}{k} x^{k-1} = n(1+x)^{n-1},$$

donc en multipliant par x :

$$\sum_{k=1}^{n} k\binom{n}{k} x^k = nx(1+x)^{n-1}.$$

On en déduit en évaluant en $x = \frac{p}{1-p}$ que pour $p \neq 1$,

$$E(X) = (1-p)^n \sum_{k=0}^{n} k \binom{n}{k} \left(\frac{p}{1-p}\right)^k = (1-p)^n n \left(\frac{p}{1-p}\right) \left(1 + \frac{p}{1-p}\right)^{n-1} = np.$$

Enfin, si $p = 1$, la formule reste vraie puisque $E(X) = \sum_{k=0}^{n} k \binom{n}{k} 1^k 0^{n-k} = n = np$.

$\square$

Exemple (Tirages avec remise)
Une urne contient 2 boules blanches et 8 boules noires.
Un joueur tire successivement, avec remise, 5 boules dans cette urne. On note X la variable aléatoire représentant le nombre de boules blanches tirées. Que vaut $E(X)$?

On a une succession de 5 épreuves indépendantes avec probabilité de succès $p = \frac{2}{10} = \frac{1}{5}$.
La VA X suit donc une loi binomiale : $X \sim \mathcal{B}\left(5, \frac{1}{5}\right)$, d'où $E(X) = 5 \times \frac{1}{5} = 1$.
Interprétation : on tire 1 boule blanche en moyenne.

Proposition 24 (Espérance de la loi géométrique).
Soit $p \in]0,1[$ *et* $X \sim \mathcal{G}(p)$*. Alors* $E(X) = \frac{1}{p}$.

Preuve : Sous réserve de convergence absolue, l'espérance de X est $E(X) = \sum_{k=1}^{+\infty} k(1-p)^{k-1}p$.
On utilise la série géométrique (dont le rayon de convergence est $R = 1$) :

$$\forall x \in]-1;1[, \qquad \sum_{k=0}^{+\infty} x^k = \frac{1}{1-x}.$$

En dérivant terme à terme ce développement en série entière, on obtient :

$$\forall x \in]-1;1[, \qquad \sum_{k=1}^{+\infty} kx^{k-1} = \frac{1}{(1-x)^2}$$

(avec convergence absolue). En évaluant ceci en $x = 1-p \in]0;1[$, on obtient :

$$\forall p \in]0;1[, \qquad \sum_{k=1}^{+\infty} k(1-p)^{k-1} = \frac{1}{p^2}.$$

Ceci montre que X est d'espérance finie, et $E(X) = p \times \frac{1}{p^2} = \frac{1}{p}$. $\square$

Proposition 25 (Espérance de la loi de Poisson).
Soit $\lambda > 0$ *et* $X \sim \mathcal{P}(\lambda)$*. Alors* $E(X) = \lambda$.

Preuve : Sous réserve de convergence absolue, l'espérance de X est

$$E(X) = \sum_{k=0}^{+\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=1}^{+\infty} e^{-\lambda}\frac{\lambda^k}{(k-1)!} = e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\lambda^{k+1}}{k!} = \lambda e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!}.$$

On reconnaît alors le DSE de e^{λ}, qui converge absolument pour tout $\lambda \in \mathbb{R}$.
Ceci montre que X admet une espérance finie, et que $E(X) = \lambda e^{-\lambda}e^{\lambda} = \lambda$.

$\square$

Exemple (Probabilité de désintégration d'un atome)
On suppose que dans une certaine quantité de matière, il y a un nombre n (très grand) d'atomes radioactifs. La probabilité de désintégration de chaque noyau atomique est très faible... Mais on observe qu'il y a en moyenne $3,2$ noyaux se désintégrant chaque seconde. Quelle est la probabilité qu'au plus 2 noyaux se désintègrent au cours d'une seconde ?
On peut modéliser le nombre X de désintégrations par seconde par une variable X suivant une loi de Poisson de paramètre $\lambda > 0$ (car les désintégrations sont des événements rares).
Le paramètre λ vaut l'espérance de X, c'est-à-dire $3,2$ d'après l'énoncé.
Ensuite ce qu'on veut est

$$\mathbb{P}(X \leq 2) = \mathbb{P}(X=0)+\mathbb{P}(X=1)+\mathbb{P}(X=2) = e^{-3,2}+3,2e^{-3,2}+\frac{(3,2)^2}{2}e^{-3,2} \approx 0,3799.$$

5) Théorème de transfert

Nous admettrons les théorèmes suivants, essentiels pour déterminer l'espérance d'une variable aléatoire composée :

Théorème 26 (Théorème de transfert pour les VA à image finie).
Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire **à image finie** *notée $X(\Omega) = \{x_1, \cdots, x_n\}$ et soit $f : X(\Omega) \to \mathbb{R}$. Alors* $E(f(X)) = \sum_{k=1}^{n} f(x_k)\mathbb{P}(X = x_k)$.

Théorème 27 (Théorème de transfert pour les VA à image dénombrable).
Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire à image dénombrable $X(\Omega) = \{x_n,\ n \in \mathbb{N}\}$ et soit $f : X(\Omega) \to \mathbb{R}$.
Alors $f(X)$ est d'espérance finie ssi $\sum_{n\geq 0} f(x_n)\mathbb{P}(X = x_n)$ *converge absolument* , *et dans ce cas, on a* $E(f(X)) = \sum_{n=0}^{+\infty} f(x_n)\mathbb{P}(X = x_n)$.

Remarque (Intérêt du théorème de transfert)
Ainsi l'espérance de la variable composée $f(X)$ est déterminée **seulement par la loi de** X. Il n'y a donc pas besoin de connaître la loi de $f(X)$ pour calculer son espérance.

Exemple
Soit $X : \Omega \to \mathbb{R}$ une VA suivant une loi géométrique de paramètre $p \in]0;1[$.
Déterminer $E(X^2)$.

On a $X(\Omega) = \mathbb{N}^*$ et $\mathbb{P}(X = k) = (1-p)^{k-1}p$ pour tout $k \geq 1$, donc d'après le théorème de transfert :

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

(sous réserve de convergence absolue). Pour calculer cette somme, on utilise encore les séries géométriques : en dérivant on a, pour tout $x \in]-1;1[$:

$$\sum_{k=0}^{+\infty} x^k = \frac{1}{1-x}, \qquad \sum_{k=1}^{+\infty} kx^{k-1} = \frac{1}{(1-x)^2}.$$

En multipliant par x et en redérivant, on obtient :

$$\forall x \in]-1;1[, \qquad \sum_{k=1}^{+\infty} kx^k = \frac{x}{(1-x)^2}, \qquad \sum_{k=1}^{+\infty} k^2x^{k-1} = \frac{(1-x)^2 + 2x(1-x)}{(1-x)^4} = \frac{1+x}{(1-x)^3}$$

(avec convergence absolue).

En évaluant en $x = p-1$ (qui est bien dans $]-1;1[$), on obtient l'existence de $E(X^2)$ et

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

III Variance d'une variable aléatoire discrète

1) Définition

Lemme 28 (Un moment d'ordre 2 entraîne une espérance finie).
Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle discrète.
Si X^2 *est d'espérance finie, alors X est d'espérance finie.*

Preuve : Pour tout réel x, on a $2|x| \leq 1 + x^2$ donc $|x| \leq \frac{1+x^2}{2}$.
En effet, $1 + x^2 - 2|x| = (1-|x|)^2 \geq 0$.
Donc par comparaison de séries à termes positifs, la convergence absolue de $\sum x_k^2 \mathbb{P}(X = x_k)$ et de $\sum \mathbb{P}(X = x_k)$ entraînent celle de $\sum x_k \mathbb{P}(X = x_k)$

$\square$

Vocabulaire
Lorsqu'elle existe, $E(X^2)$ est appelée le moment d'ordre 2 de X.

Proposition 29 (Moment d'ordre 2 entraîne moment centré d'ordre 2).
Si $E(X^2)$ *existe, alors* $E\Big(\big(X - E(X)\big)^2\Big)$ *existe.*

Preuve : En notant $m = E(X)$ (qui existe d'après la proposition précédente), on a

$$(X - E(X))^2 = (X - m)^2 = X^2 - 2mX + m^2.$$

La VA constante m^2 admet une espérance (égale à elle-même), ainsi que la variable $-2mX$, ainsi que la variable X^2 (par hypothèse). D'où, par somme, $(X - E(X))^2$ admet une espérance.

□

Vocabulaire
Lorsqu'elle existe, $E((X - E(X))^2)$ est appelée le moment centré d'ordre 2 de X.

La proposition précédente donne un sens à la définition suivante :

Définition 30 (Variance d'une variable aléatoire réelle).
Soit X *une variable aléatoire réelle discrète telle que* $E(X^2)$ *existe.*
On appelle **variance de X** *le réel* $V(X) = E\Big(\big(X - E(X)\big)^2\Big)$.

Définition 31 (Variable aléatoire réduite).
On dit que X *est* **réduite** *lorsque* $V(X) = 1$.

Remarque
- On a toujours $V(X) \geq 0$, car $\big(X - E(X)\big)^2 \geq 0$ donc son espérance est positive.
- La variance est une façon de mesurer l'écart des valeurs d'une variable aléatoire par rapport à son espérance (qui est sa valeur moyenne pondérée). On aurait pu trouver plus simple d'utiliser la quantité $E\big(|X - E(X)|\big)$ pour mesurer cet écart, mais la variance a de meilleures propriétés. D'ailleurs l'emploi d'un carré fait penser à la norme euclidienne.
- Notons qu'on a $V(X) = 0 \iff X = E(X)$ presque sûrement, c'est-à-dire ssi $\mathbb{P}\big(X = E(X)\big) = 1$.

Définition 32 (Écart-type).
Soit X une variable aléatoire réelle discrète telle que $E(X^2)$ existe.
On appelle écart-type de X le réel $\sigma(X) = \sqrt{V(X)}$.

Remarque (Lien avec l'écart-type statistique)
L'écart-type mesure **une dispersion des valeurs** de X par rapport à la moyenne pondérée (i.e. l'espérance) de ses valeurs.
Ainsi deux classes peuvent avoir la même moyenne (i.e. espérance correspondant à une variable aléatoire uniforme) sur un DS mais présenter une dispersion des notes très différentes. Deux cas extrêmes pour une moyenne à 10/20 :

- si tout le monde a 10/20, alors l'écart-type est de 0.
- si la moitié des élèves a 20/20 et l'autre moitié a 0/20, alors l'écart-type est de 10.

L'écart-type est donc une sorte **d'écart moyen** entre les notes et la moyenne de la classe.

2) Propriétés de la variance

On a les propriétés :

Proposition 33 (Propriétés de la variance).
Soit X une variable aléatoire réelle discrète telle que $E(X^2)$ existe. Alors :

(i) $V(X) = E(X^2) - E(X)^2$ *(formule de Huygens).*

(ii) $\forall (a, b) \in \mathbb{R}^2,\ V(aX + b) = a^2 V(X)$.

(iii) Si $\sigma(X) > 0$, alors la variable aléatoire $\dfrac{X - E(X)}{\sigma(X)}$ *est centrée et réduite* .

Remarque (Utilité de la formule de Huygens)
En général, on calcule la variance à l'aide de la formule de Huygens : il suffit de calculer $E(X)$ et $E(X^2)$ pour obtenir la variance.

ATTENTION !
La variance **n'est pas** linéaire : si X et Y sont deux variables discrètes, il n'y a pas de formule générale pour calculer $V(X + Y)$. Mais **dans certains cas**, on aura $V(X + Y) = V(X) + V(Y)$ (voir chapitre sur les couples de variables aléatoires).
En revanche, on a **toujours** $V(aX) = a^2 V(X)$ pour $a \in \mathbb{R}$ (et non pas $aV(X)$!)

Preuve : Notons $m = E(X) \in \mathbb{R}$ (qui existe car X possède un moment d'ordre 2).

(i) Par linéarité de l'espérance, on a

$$V(X) = E\left(\left(X - m\right)^2\right) = E\left(X^2 - 2mX + m^2\right) = E(X^2) - 2mE(X) + m^2 = E(X^2) - m^2,$$

ce qui montre la formule de Huygens : $V(X) = E(X^2) - E(X)^2$.

(ii) Soit $(a, b) \in \mathbb{R}^2$. On a

$$\begin{aligned} V(aX+b) &= E\left(\left(aX + b - E(aX+b)\right)^2\right) = E\left(\left(aX + b - aE(X) - b\right)^2\right) \\ &= E\left(\left(aX - aE(X)\right)^2\right) = a^2 E\left(\left(X - E(X)\right)^2\right) = a^2 V(X) \end{aligned}.$$

(iii) Notons $Y = \dfrac{X - E(X)}{\sigma(X)}$.

Par linéarité de l'espérance, on a

$$E(Y) = \frac{1}{\sigma(X)} E(X - E(X)) = \frac{1}{\sigma(X)} \left(E(X) - E(X)\right) = 0,$$

et d'après (ii) :

$$V(Y) = V\left(\frac{1}{\sigma(X)} X - \frac{E(X)}{\sigma(X)}\right) = \frac{1}{\sigma(X)^2} V(X) = \frac{V(X)}{V(X)} = 1.$$

$\square$

3) Variances des lois usuelles

Proposition 34 (Variances des lois usuelles).

(i) Si $X \sim \mathcal{B}(p)$*, alors* $V(X) = p(1-p)$.

(ii) Si $X \sim \mathcal{B}(n,p)$*, alors* $V(X) = np(1-p)$.

(iii) Si $X \sim \mathcal{G}(p)$*, alors* $V(X) = \frac{1-p}{p^2}$.

(iv) Si $X \sim \mathcal{P}(\lambda)$*, alors* $V(X) = \lambda$.

Preuve :

(i) Si $X \sim \mathcal{B}(p)$, alors X est à valeurs dans $\{0,1\}$, donc $X = X^2$.
Ainsi $V(X) = E(X^2) - E(X)^2 = E(X) - E(X)^2 = p - p^2 = p(1-p)$.

(ii) Si $X \sim \mathcal{B}(n,p)$, alors on procède comme pour le calcul de $E(X) = np$ (cf. prop 23). On veut calculer

$$E(X^2) = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = (1-p)^n \sum_{k=0}^{n} k^2 \left(\frac{p}{1-p}\right)^k$$

(pour $p \neq 1$). On introduit le polynôme suivant :

$$\forall x \in \mathbb{R}, \quad \sum_{k=0}^{n} \binom{n}{k} x^k = (1+x)^n.$$

En dérivant et en multipliant par x, on obtient :

$$\forall x \in \mathbb{R}, \quad \sum_{k=0}^{n} k \binom{n}{k} x^k = \sum_{k=1}^{n} k \binom{n}{k} x^k = nx(1+x)^{n-1},$$

puis en répétant la même opération (dérivation et multiplication par x) :

$$\forall x \in \mathbb{R}, \quad \sum_{k=0}^{n} k^2 \binom{n}{k} x^k = \sum_{k=1}^{n} k^2 \binom{n}{k} x^k = x\frac{d}{dx}\left(nx(1+x)^{n-1}\right) = nx(1+nx)(1+x)^{n-2}.$$

(tous ces calculs nécessitent *a priori* que $n \geq 2$ mais restent vrais pour $n = 0$ et $n = 1$).
En évaluant en $x = \frac{p}{1-p}$ si $p \neq 1$, on obtient

$$E(X^2) = (1-p)^n n \left(\frac{p}{1-p}\right)\left(1 + \frac{np}{1-p}\right)\left(1 + \frac{p}{1-p}\right)^{n-2},$$

c'est-à-dire

$$E(X^2) = np(1-p+np)(1-p+p) = np(1-p+np).$$

Finalement, on a donc pour $p \neq 1$:

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

et cela reste vrai pour $p = 1$ car dans ce cas, X est presque sûrement égale à n donc $V(X) = 0$.

(iii) Si $X \sim \mathcal{G}(p)$, on a déjà calculé $E(X) = \frac{1}{p}$ (voir p.43) et $E(X^2) = \frac{2}{p^2} - \frac{1}{p}$ (voir p.47), en dérivant la série géométrique $\sum_{k=0}^{+\infty} x^k = \frac{1}{1-x}$ pour $x \in]-1;1[$.
On en déduit :

$$V(X) = E(X^2) - E(X)^2 = \frac{2}{p^2} - \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2}.$$

(iv) Si $X \sim \mathcal{P}(\lambda)$, alors on a déjà calculé $E(X) = \lambda$ (voir p.44). Ensuite, sous réserve de convergence absolue :

$$E(X^2) = \sum_{k=0}^{+\infty} k^2 \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{+\infty} k \frac{e^{-\lambda}\lambda^k}{(k-1)!} = \sum_{k=0}^{+\infty} (k+1) \frac{e^{-\lambda}\lambda^{k+1}}{k!},$$

c'est-à-dire

$$E(X^2) = \lambda \left(\sum_{k=0}^{+\infty} k \frac{e^{-\lambda}\lambda^k}{k!} + \sum_{k=0}^{+\infty} \frac{e^{-\lambda}\lambda^k}{k!} \right).$$

On reconnaît des séries absolument convergentes :

$$\sum_{k=0}^{+\infty} k \frac{e^{-\lambda}\lambda^k}{k!} = E(X) = \lambda, \qquad \sum_{k=0}^{+\infty} \frac{e^{-\lambda}\lambda^k}{k!} = 1,$$

donc $E(X^2)$ est finie et $E(X^2) = \lambda^2 + \lambda$, donc $V(X) = E(X^2) - E(X)^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$.

$\square$

Remarque
Pour la loi uniforme sur $\{x_1, \cdots, x_n\}$, il n'y a pas de formule simple de la variance. En revanche, si les x_k forment une suite arithmétique, oui.

Exemple (Variance de $\mathcal{U}\{1, \cdots, n\}$)
Si X suit la loi uniforme sur $\{1, \cdots, n\}$, alors $\mathbb{P}(X = k) = \frac{1}{n}$ pour tout $k \in \{1, \cdots, n\}$, donc sa variance est (d'après la formule de Huygens) :

$$\begin{aligned} V(X) &= E(X^2) - E(X)^2 = \left(\sum_{k=1}^{n} \frac{1}{n}k^2\right) - \left(\sum_{k=1}^{n} \frac{1}{n}k\right)^2 \\ &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12} \end{aligned}$$

4) Inégalité de Bienaymé-Tchebychev

Lemme 35 (Inégalité de Markov).
Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire discrète **à valeurs positives** *et admettant une espérance finie. Alors* pour tout réel $a > 0$, $\mathbb{P}(X \geq a) \leq \dfrac{E(X)}{a}$.

Preuve : Pour tout $a > 0$, on décompose $E(X)$ en deux sommes selon les valeurs de X qui sont supérieures ou inférieures à a :

$$E(X) = \sum_{x \in X(\Omega)} x\mathbb{P}(X = x) = \sum_{x \in X(\Omega),\ x < a} x\mathbb{P}(X = x) + \sum_{x \in X(\Omega),\ x \geq a} x\mathbb{P}(X = x).$$

Puisque ce sont des sommes de nombres positifs, on peut minorer :

$$E(X) \geq \sum_{x \in X(\Omega),\ x \geq a} x\mathbb{P}(X = x) \geq a \sum_{x \in X(\Omega),\ x \geq a} \mathbb{P}(X = x) = a\mathbb{P}(X \geq a),$$

ce qui amène l'inégalité voulue en divisant par $a > 0$. $\square$

Comme corollaire de l'inégalité de Markov, on a :

Proposition 36 (Inégalité de Bienaymé-Tchebychev).
Si la variable aléatoire discrète $X : \Omega \to \mathbb{R}$ admet un moment d'ordre 2 alors on a :

$$\forall \varepsilon > 0, \quad \mathbb{P}\big(|X - E(X)| \geq \varepsilon\big) \leq \frac{V(X)}{\varepsilon^2}.$$

. . . .

Preuve : On a l'égalité d'événements $\big(|X - E(X)| \geq \varepsilon\big) = \Big(\big(X - E(X)\big)^2 \geq \varepsilon^2\Big)$ et par l'inégalité de Markov appliquée à la variable $Y = \big(X - E(X)\big)^2$, qui est bien positive et admet une espérance vu que X admet un moment d'ordre 2 :

$$V(X) = E(Y) \geq \varepsilon^2 \mathbb{P}(Y \geq \varepsilon^2) = \varepsilon^2 \mathbb{P}\big(|X - E(X)| \geq \varepsilon\big),$$

d'où le résultat en divisant par $\varepsilon^2 > 0$. $\square$

Exemple (Fréquence d'apparition d'un caractère dans une population)
On sait que 40% des individus d'une population possèdent un certain caractère C (par ex. les yeux bleus). On considère un échantillon de 200 personnes de cette population. Peut-on affirmer, à plus de 80%, que dans cet échantillon il y a entre 30 et 50% de personnes ayant le caractère C ?

Interroger au hasard 200 personnes et voir si elles présentent le caractère C revient à jouer 200 fois d'affilée à *Pile* ou *Face* (avec une pièce truquée telle que la probabilité d'obtenir *Pile* est $p = 0,4$) et compter le nombre de fois où l'on a obtenu *Pile*.

Notons donc X la variable donnant le nombre d'individus de l'échantillon possédant le caractère C. On a $X(\Omega) = \{0, 1, \cdots, 200\}$ et X suit une loi binomiale de paramètres $(n; p) = (200; 0,4)$.
On a donc $E(X) = np = 200 \times 0,4 = 80$ et $V(X) = np(1-p) = 80 \times 0,6 = 48$.

La fréquence d'apparition du caractère C dans la population est $\frac{X}{200}$.
On veut estimer :

$$\mathbb{P}\left(0,3 \leq \frac{X}{200} \leq 0,5\right) = \mathbb{P}(60 \leq X \leq 100) = \mathbb{P}\big(|X - E(X)| \leq 20\big) = 1 - \mathbb{P}\big(|X - E(X)| \geq 21\big)$$

(car $|X - E(X)| = |X - 80|$ est à valeurs entières positives).

Or, l'inégalité de Bienaymé-Tchebychev donne la majoration

$$\mathbb{P}\big(|X - E(X)| \geq 21\big) \leq \frac{V(X)}{21^2} = \frac{48}{441}$$

et donc

$$\mathbb{P}\left(0,3 \leq \frac{X}{200} \leq 0,5\right) \geq 1 - \frac{48}{441} \approx 0,89$$

donc on est sûr à plus de 80% (presque 90) qu'il y a entre 30 et 50% de personnes ayant le caractère C.